

# 토픽 모델링을 이용한 스마트 항만물류 시스템 기술동향분석 연구

전형모  
한국해양수산개발원  
e-mail:hmjeon@kmi.re.kr

## A Study on Technology Trends in Smart Port Logistics System using Topic Modelling Analysis

Hyongmo Jeon  
Korea Maritime Institute, Busan, Korea

### 요약

21세기에 들어서면서 정보통신기술의 급속한 발전과 함께 전 세계 다양한 산업분야에 디지털 전환이 빠르게 진행되고 있다. 이러한 경향은 항만 물류에서도 예외가 아니다. 전통적으로 항만물류 분야는 신기술 도입에 보수적이었지만, '스마트항만'으로 대표되는 지능형·자동화 항만에 대한 관심이 높아지고 있다. 다만 현재 '스마트 항만'의 정의나 기준이 불분명하고, 국가별로 집중된 연구개발과 응용기술의 방향도 조금씩 다르다. 본 연구는 우리나라를 비롯하여 미국, 중국, 유럽 등 주요국의 '스마트 항만'에 대한 기술개발 동향분석을 실시하여 각국의 연구개발 방향을 파악하고 핵심기술을 발굴하여, '스마트 항만'에 대한 주요주제들을 제시하였다. 분석자료는 미국, EU, 중국, 한국 등에서 지난 20년간 '스마트 항만' 관련 기술특허를 수집하였다. LDA 기반 토픽 모델링 분석방법이 방대한 양의 특허 정보에서 기술 주제를 효율적으로 도출해 내기 위해 사용되었다. 또한 각 주제 중 연구개발이 증가하는 주제와 점차 쇠퇴하는 주제를 파악하기 위해서 해당 토픽에 포함된 주요 키워드들이 연도별로 얼마만큼 발생했는지를 추적하였다.

※ 이 연구는 2022년 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구임(스마트컨테이너 실용화 기술개발)

### 1. 서론

2010년대 이후 다양한 산업 전반에 걸쳐 '4차산업혁명', '디지털화', '스마트' 등의 용어에 대한 언급과 인식이 확산되고 있다. 특히 사물인터넷(IoT), 인공지능(AI), 빅데이터분석(Big Data Analysis), 블록체인(Block Chain), 로보틱스(Robotics) 등의 기술의 발달은 기존의 산업을 보다 효율적이고 생산적으로 만드는 것을 넘어서서 산업의 형태를 변화시키거나 아예 다른 산업으로 바꾸는 경우도 발생하고 있다. 특히 이러한 기술의 영향은 다른 산업보다 물류 산업에 더 크게 작용하고 있다. 이는 위에서 언급한 기술들이 연결성의 강화와 가시성의 확보, 그리고 예측능력의 증대에 초점을 맞추고 있고, 이러한 '연결성', '가시성', '예측'은 결국 물류 산업의 핵심이기 때문이다.

이러한 경향은 항만물류에서도 예외가 아니다. 전통적으로 해운 및 항만 분야는 새로운 기술의 수용에 대해 보수적인 편이다. 이미 20세기 중반부터 생산분야에서는 다양한 자동화 및 최적화 기술들이 개발 및 적용되기 시작하였으나, 항만의 자동화는 다양한 관련기술들의 검증이 어느 정도 이뤄진 20

세기 후반부터 유럽항만을 중심으로 이뤄지기 시작하였다. 그러나 2000년대 들어와서는 '스마트 항만'으로 대표되는 지능형, 자동화 항만 및 관련 기술들에 대한 관심이 크게 높아지고 있다.

다만 아직 '스마트 항만'에 대한 개념이 명확하지 않으며, 어떤 기술들이 개발되고 적용되고 있는지도 제대로 분석이 되어 있지 못한 것이 사실이다. 따라서 본 연구에서는 우리나라를 비롯하여 미국, EU, 중국 등의 스마트항만 관련 기술개발 추이를 분석하고 주요 연구주제를 도출하여 향후 국내 스마트항만 관련기술 개발 방향설정에도 도움을 주고자 한다.

### 2. 분석방법론

'스마트항만'에 대해 우리나라 및 주요 국가의 특허를 분석하여 연구·개발 동향을 분석한다. 동향분석을 위해서는 텍스트마이닝 기법 중 하나인 토픽모델링을 적용한다. 토픽모델링의 결과 각 국가별로 과거 20년간 출원된 특허의 주요 주제를 도출할 수 있으며, 그에 따라 각 국가가 주요하게 추진해 온 분야별 핵심 기술개발의 동향을 파악할 수 있다.

## 2.1 데이터 수집 및 전처리

국내외 특허데이터는 다양한 방식으로 수집이 가능하며, 본 연구에서는 키워트(www.keywert.com)를 통해 특허를 수집하였다. 본 연구에서는 스마트항만에 대하여 한국, 중국, 미국, EU 등의 국가에서 과거 20년간 출원한 특허 약 8만5천여 건을 수집하였다.

[표 1] 국가별 특허데이터 개요

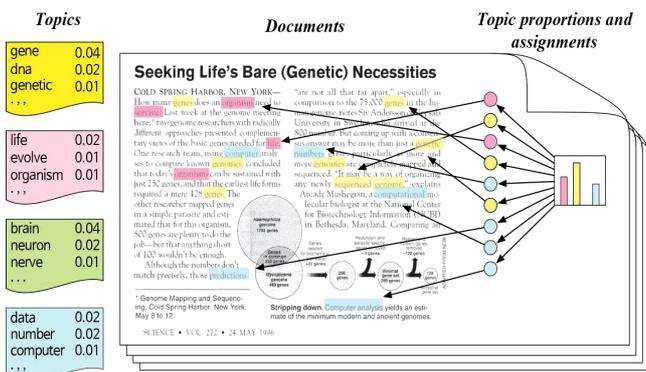
국가	특허출원 기간	조사특허 수
한국	2000. 1 ~ 2020. 4	8,319
중국	2000. 1 ~ 2020. 4	52,100
미국	2000. 1 ~ 2020. 4	17,743
EU	2000. 1 ~ 2020. 4	6,866
소계		85,028

자료: 저자 작성

데이터를 불러온 후에는 분석에 필요 없는 단어와 기호, 빈 공간, 쉼표 등을 제거하기 위해서 전처리 작업을 수행한다. R의 텍스트마이닝 패키지인 'tm'에서 제공하는 'SMART(System for the Mechanical Analysis and Retrieval of Text)'를 활용하여 불용어를 제거하고, 그 외 특허문서에 빈번하게 사용되는 용어를 추가적으로 제거하여 각 특허의 초록을 문장이 아닌 키워드로 구성된 수열로 변환시킨다.

## 2.2 토픽모델링

토픽모델링은 여러 문서의 집합에서 해당 문서를 구성하고 있는 주요 키워드의 발생 확률을 기반으로 해당 문서집합에 내재하고 있는 토픽들을 도출하는 기계학습 기법이 하나이다. 본 연구에서는 잠재 디리클레 할당(LDA)모형을 사용하였다.



자료: R and RStudio, "Topic Model", (검색일 2022. 9. 17)

[그림 1] LDA 토픽모델링 예시

LDA는 전체 문서에 대해 k개의 주제(topic)가 있고, 이 주제에는 n개의 단어(word)가 있다고 가정할 때, 하나의 문서가 완성되기 위해서는 k 주제 중 하나를 정하고, 그 주제에 해당하는 여러 단어 중 하나를 선정하여 문서에 쓰게 되며,

반복적으로 다시 주제를 고르고 해당 주제 안의 단어를 골라 문서를 완성(생성)해 나가게 된다고 가정한다. 하나의 문서가 하나의 주제(topic)만을 포함하지 않기 때문에 하나의 문서 안에는 여러 주제(topic)가 확률적으로 분포하게 되고, 마찬가지로 해당 주제(topic)에 부합하는 단어(word) 역시 확률적으로 문서 안에 분포한다고 LDA는 가정한다. 상기 과정을 역으로 생각하면 문서를 이루는 단어와 문장을 이용하여 토픽의 분포와 단어의 분포를 추정할 수 있고, 이러한 추정을 수행하는 것이 LDA 토픽모델링이다.

LDA를 수행하는데 있어 중요한 결정사항은 토픽의 숫자(K)를 정하는 것이다. 본 연구에서는 Perplexity(복잡도)를 최소화하는 K를 방법을 사용한다. Perplexity를 최소화한다는 것은 토픽 내 문장의 확률을 최대화하는 것과 같다. 일반적으로 R을 활용한 LDA 알고리즘은 김슨샘플링을 반복하여 최적의 토픽구성을 찾아가는 방식인데(Grun and Hornik(2011)), 샘플링이 반복됨에 따라 Perplexity가 감소하다가 일정 지점부터 더 이상 감소하지 않게 되는데 해당 지점이 최적의 토픽수를 판단할 수 있다.

## 2.3 결과의 해석

특허초록에 대한 토픽모델링의 결과로는 키워드로 구성된 정해진 수만큼의 토픽이 생성되며, 나열된 키워드의 조합을 통해 해당 문서가 포함하고 있는 토픽의 의미를 정해야 한다. 그러나 단어의 조각들만으로 해당 토픽이 의미하는 주제를 명확하고 객관적으로 파악하는 것은 해당분야 전문가들에게도 쉽지 않은 일이며, 검증도 어렵다.

본 연구에서는 Newman et al.(2010)의 방법을 활용, 구글 검색을 바탕으로 토픽 내의 상위 N개의 키워드들을 입력했을 때 검색결과로 나오는 문서의 제목과 키워드 간의 일치성을 검토하여 토픽이 얼마나 잘 구성되어 있는지 확인하고 이를 기반으로 1차적으로 토픽을 구성, 전문가 검증을 통해 토픽을 확정하였다.

## 2.4 트렌드분석

각 토픽 중 연구개발이 증가하는 토픽과 점차 쇠퇴하는 토픽을 보기 위해 먼저 토픽의 시계열 변화를 추적한다. 토픽의 시계열 변화란 해당 토픽에 포함된 주요 키워드들이 연도별로 얼마만큼 발생했는지를 의미한다. 본 연구에서는 특허정보 내 출원일을 기준으로 하며, 일별이나 월별 변화보다는 연도별 변화를 추적한다. 이를 위해 먼저 각 토픽에 속한 키워드의 문서별 빈도의 행렬을 구축하여 문서별 키워드 수의 합을 구해 행렬을 나눠주면 각 문서에서 토픽별 키워드의 출현 확률이 구해진다. 각 문서는 특허의 시간별로 발생하므로 이를 활용하여 연도별 토픽의 변화를 추적할 수 있다.

### 3. 분석결과

#### 3.1 중국

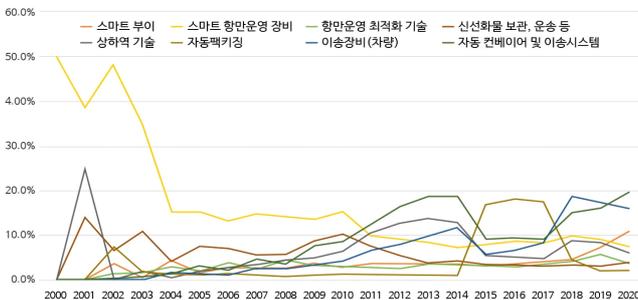
먼저 가장 활발하게 특허를 출원하고 있는 국가인 중국에 대한 모델링 결과를 살펴본다. Perplexity 값을 추적결과 29에서 가장 낮은 값을 보였으나 토픽수 10부터 감소세가 크게 줄어들기 시작하여 감소폭이 작아지는 13을 토픽의 개수로 결정하였다.(김태경·최희련·이홍철(2016))

도출된 중국 항만기술의 13개 토픽의 주제는 항만의 일반적인 기능인 하역 및 이송, 기반시설 관련 배관은 물론, 스마트 기술관련 장비(하드웨어) 및 최적화 시스템(소프트웨어), 재생에너지 기반 스마트 부표, 자동 패키징 및 신선화물 시스템 등에 대한 연구개발이 이뤄져 온 것을 볼 수 있다. 이송장비에도 ‘자동화’, ‘배터리’ 등의 키워드가 포함되어 있어 중국 항만의 자동화에 대한 연구개발이 이뤄져 온 것을 확인할 수 있다.

[표 2] 중국 스마트항만 주요주제

토픽	주제	토픽	주제
토픽1	항만 해상 작업장비, 선박 구조물	토픽8	신선화물 보관, 운송 등
토픽2	스마트 부표	토픽9	상하역 및 접안 기술
토픽3	배관 및 관련시설	토픽10	자동패키징
토픽4	스마트항만 운영/통신 장비	토픽11	이송장비(차량)
토픽5	선체(외부)검사 및 수리	토픽12	자동 컨베이어 및 이송시스템
토픽6	항만운영 최적화 기술	토픽13	벌크(곡물)화물 자동이송
토픽7	액체화물 이동 및 분사		

13개 토픽 중 스마트항만 운영장비 기술은 2000년부터 2004년 집중적으로 연구가 이뤄지다가 점차 줄어들었다. 최근 들어 연구가 늘어나는 기술은 이송장비(차량 및 컨베이어 시스템) 관련 기술이며, 스마트 부표에 대한 연구도 점차 증가하고 있는 것을 볼 수 있다.



[그림 2] 중국 스마트항만 기술 추이

#### 3.2 미국

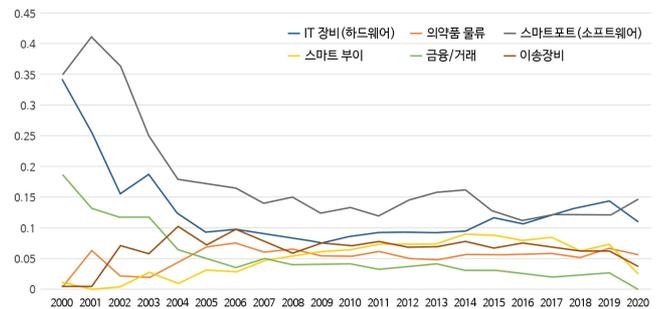
미국 특허데이터의 경우 Perplexity의 감소세가 크게 줄어드는 시점인 12를 토픽의 개수로 결정하였다.

미국은 주로 의약품 물류 기술개발에 집중하고 있는 것을 볼 수 있으며, 식품의 경우에는 패키징에 대한 연구개발을 주로 수행하였다. 그 외에도 액체용기 봉인기술이나 컨테이너 용기기술 등 패키징에 많은 연구를 수행하고 있는 것을 볼 수 있다. 스마트항만을 위한 소프트웨어 분야에서는 미국은 시스템 구성에 대한 연구개발을 수행하여 왔다. 또한 액화가스의 저장 및 이송에 대한 기술 연구가 이뤄졌고, 무엇보다 항만에서의 금융거래 관련 기술 등이 개발되어 온 것을 볼 수 있다.

[표 3] 미국 스마트항만 주요주제

토픽	주제	토픽	주제
토픽1	IT 장비(하드웨어)	토픽7	유류 및 액체화물
토픽2	의약품 물류	토픽8	액체용기 봉인기술
토픽3	스마트포트 시스템	토픽9	금융거래 관련 기술
토픽4	해상구조물	토픽10	신선식품 포장기술
토픽5	스마트 부표	토픽11	이송장비
토픽6	액화가스 저장 및 이송	토픽12	컨테이너 및 용기기술

최근 20년간의 연도별 연구개발 추이를 살펴보면 IT 장비 및 소프트웨어 기술에 대한 연구개발은 줄곧 높은 비중을 차지하고 있으며, 스마트 부표에 대한 연구도 그 중요성이 지속적으로 증가하고 있는 것을 볼 수 있다.



[그림 3] 미국 스마트항만 기술 추이

#### 3.3 EU

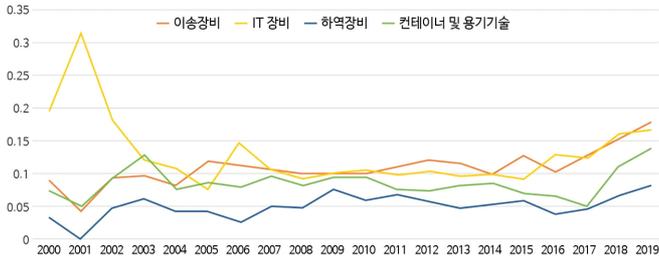
유럽은 EU를 중심으로 특허를 조사하여 분석하였으며, EU 특허자료의 토픽수는 10으로 설정하였으며, 용기 봉인기술, 식품 패키징 기술, 건화물 용기기술, 액체용기기술 등으로 패키징 기술이 세분화가 이뤄졌고, 스마트 관련 기술로 IT 장비 및 운영기술, 이송장비 기술이 하나의 토픽으로 도출되었다.

[표 4] EU 스마트항만 주요주제

토픽	주제	토픽	주제
토픽1	이송장비	토픽6	식품 패키징 기술
토픽2	해상구조물	토픽7	하역장비
토픽3	IT 장비 및 운영	토픽8	컨테이너 및 용기기술
토픽4	용기 봉인기술	토픽9	액체 용기기술
토픽5	액화가스 저장/이송 기술	토픽10	유류 및 액체화물

스마트항만 관련 기술인 이송장비 기술, 하역장비 기술, IT 장비기술, 컨테이너 기술 등의 비중이 최근 들어 지속적으로 상승하고 있으며, 특히 IT 기술은 2000년대 초반부터 지금까지

지 높은 중요도를 유지하고 있는 것을 알 수 있다.



[그림 4] EU 스마트항만 기술 추이

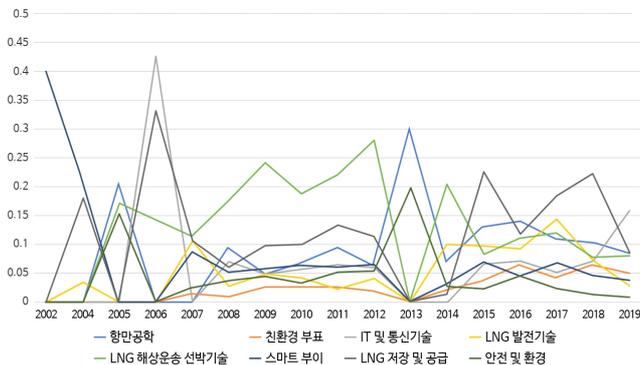
### 3.4 한국

국내특허의 경우 Perplexity의 감소세가 크게 둔화되는 11을 우리나라 특허자료에 대한 토픽모델링 토픽수로 설정하였으며, 토픽모델링 결과 항만공학, 일반 및 스마트 부표, IT 및 통신기술, LNG 관련 기술, 해상 구조물 및 작업선, 환경관련 기술들이 도출되었다. 반면 상하역 기술이나 이송장비 등 스마트항만 관련 주요 기술은 토픽으로 형성되지 못하고 있는 것을 알 수 있다. 이는 국내에서 해당 분야에 대한 연구개발이 저조하다는 것을 의미한다.

[표 5] 한국 스마트항만 주요주제

토픽	주제	토픽	주제
토픽1	항만공학	토픽7	스마트 부표
토픽2	부표	토픽8	해상구조물
토픽3	IT 및 통신기술	토픽9	LNG 저장 및 공급
토픽4	LNG 발전기술	토픽10	해상 작업장비
토픽5	선박 해상수리	토픽11	안전 및 환경
토픽6	LNG 해상운송 기술		

연도별 추이를 살펴보면 꾸준히 기술개발이 이뤄지고 있는 분야는 해상구조물과 LNG 해상운송 기술 등이며, 항만공학 등도 지속적으로 연구개발이 이뤄지고 있으며, 해양환경에 대한 관심도 증가하고 있어 스마트 부표 외에 일반 부표의 성능과 내구성 등에 대한 연구도 그 비중이 점차 증가하고 있음을 볼 수 있다.



[그림 5] 한국 스마트항만 기술 추이

## 4. 결론 및 시사점

각 국가들이 수행해온 연구주제 중 ‘스마트항만’ 관련 연구

명 기술개발 주제는 ‘스마트 부표’, ‘IT 장비’, ‘IT 솔루션(운영 최적화 등)’, ‘상하역장비 및 이송장비 자동화’, ‘스마트 컨테이너’, ‘콜드체인’ 등의 기술로 구분할 수 있다. 토픽모델링 분석 결과에 중국은 최근 20년간 활발한 연구개발을 통해 대부분의 분야에서 활발한 연구개발을 수행하고 있으며, 미국과 유럽 역시 각국의 특성에 맞게 기술을 개발하고 있는 것을 볼 수 있다. 그러나 한국의 스마트항만 관련 기술개발 동향은 경쟁국에 비해 다소 부진하다고 볼 수 있다. 스마트 부표와 IT 및 통신장비 일부에서만 연구개발이 이뤄져 왔고, 스마트항만의 핵심인 상하역장비와 이송장비에 대한 연구개발이 활발히 이뤄지지 못하였으며, 스마트 컨테이너, 운영최적화 역시 실제 특허출원 및 활용까지는 다다르지 못한 것으로 분석되고 있어, 향후 해당분야의 연구 및 기술개발에 대한 노력이 요구된다.

[표 6] 스마트항만 관련 국가별 기술개발 동향

구분	중국	미국	EU	한국
스마트 부표	0	0	-	0
IT 및 통신 장비	0	0	0	0
운영 최적화	0	0	-	-
상하역장비	0	-	0	-
이송 장비	0	0	0	-
기타 자동화 시스템	0	-	-	-
컨테이너 기술	-	0	0	-
콜드체인 관련 기술	0	0	-	-

### 참고문헌

- [1] 김태경·최희련·이홍철, “토픽모델링을 이용한 핀테크 기술 동향 분석”, 한국산학기술학회논문지, 제17권 제11호, 2016.
- [2] Grun, B. and K. Hornik, “Topicmodels: An R package for fitting topic models”, Journal of Statistical Software, vol. 40, no. 13, 2011.
- [3] Ldavis, “A topic model for movie reviews”, <https://ldavis.cpsievert.me>, (검색일: 2022. 6. 14)
- [4] Newman, David., Jey Han Lau, Karl Grieser and Timothy Baldwin, “Automatic Evaluation of Topic Coherence”, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, Los Angeles, California, June 2010.
- [5] R and RStudio, ”Topic Model“, [https://rstudio-pubs-static.s3.amazonaws.com/63854\\_c802d802e2204937a25676d17e896f84.html](https://rstudio-pubs-static.s3.amazonaws.com/63854_c802d802e2204937a25676d17e896f84.html), (검색일 2020. 9. 17)
- [6] REPLET, “[R] LDA를 활용한 토픽트렌트 분석”, <http://textmining.kr/?p=440> (검색일: 2022. 6. 18)
- [7] Wikipedia, 「SMART Information Retrieval System」, [https://en.wikipedia.org/wiki/SMART\\_Information\\_Retrieval\\_System](https://en.wikipedia.org/wiki/SMART_Information_Retrieval_System) (검색일: 2022. 6. 14)