

입출항 프로세스 내 발생하는 비정형 데이터의 데이터 체계 유형화에 관한 연구

박미란, 김낙일, 김원현, 윤승정, 문지호

(주)슈타겐

mrpark@shutagen.com, nikim@shutagen.com, whkim@shutagen.com, sjyoon@shutagen.com,
jhmoon@shutagen.com

Typology of Unstructured Data in Port Entry and Departure Processes

Park Mi Ran, Kim Nak il, Kim Won Hyun, Yoon Seong Jeong, Moon Ji Hyo

Shutagen Co., Ltd.

요약

본 논문은 해상에서 사용하는 항해 통신 장비(이하, 항통장비)에 대한 데이터 형태를 음성에서 문자, 그래프 등 시각화 또는 유형화된 데이터로 만들어야 할 필요성에 대하여 제시하고자 한다. 입출항 시에 선박과 항만 사이에서 발생하는 데이터는 VHF와 같은 항통장비를 사용하여 발생하는 음성 데이터로 저장되거나 관리되지 않는다. 다만, 선박 입출항 전에 세관, 검역 등 출입국에 관련된 정보를 PLSM3.0에 제공하고, 하역, 선물품 공급 등과 같은 항만 서비스와 관련된 정보는 Port-MIS(해운항만물류시스템)에 제공함으로써 정형 데이터를 수집 및 관리하고 있다. 이외에 선박 운영될 때 발생하는 데이터는 일반적으로 음성 데이터로 비정형 구조를 가지고 있기 때문에 통신의 상태에 따라 데이터의 품질이 좌우되는 경우가 빈번하고 입출항 프로세스 내 업무 특성상 여러 사람이 번잡한 통신을 사용해 업무를 수행할 수 밖에 없다. 입출항 프로세스는 각 업무에 따라 수행하는 이해관계자, 방법, 절차 등이 고정되어있는 정형화된 프로세스이다. 예를 들어, 선박의 예상 도착시간을 Agent와 공유하고, 항만 터미널에서 확인하고 항만 스케줄을 확인하여 입출항이 가능한지 반복적으로 확인한다. 이러한 과정은 이메일과 같은 통신 방법으로 수행하는데 이와같이 데이터는 형식은 있지만 관리하는 사람에 의존하여 스케줄을 확인하여 조정해야만 한다. 만약 입출항 과정에서 발생하는 예외상황이 발생할 경우 신속하게 대응할 수 없게 된다. 따라서, 음성뿐만 아니라 문자나, 그래프, 영상을 결합하여 일시적으로 보는 것 뿐만 아니라, 이를 저장하여 대응 시나리오를 생성할 수 있게 한다면 입출항 프로세스의 복잡성과 선박의 안전 운항에 대한 대응책을 확보할 수 있게 될 것이다.

I. 서론

본 논문에서는 디지털 포트 개발에 필요한 데이터의 활용성과 통합 및 관리의 효율성을 높이기 위해 비정형 데이터의 유형과 정형화 해야 하는 이유에 대해서 연구하고자 한다. 가까운 미래의 차세대 항만과 선박으로 많은 연구가 되고 있는 디지털 포트와 자율운항선박(Autonomous Ship)과 친환경 선박(Green Ship)으로 대표할 수 있는 스마트 선박(Smart Ship)은 현재에 비해 훨씬 자동화되고, 데이터 통합이 강화되며, 환경 친화적인 방향으로 발전할 것이다. 이에 따라 차세대 기술을 적용하기 위해서는 현재 업무 프로세스 내에서 발생하는 비정형 데이터에 대해 정형화 해야 할 필요가 있다. 비정형 데이터를 효과적으로 활용하기 위해서 인공지능 및 인지 기술, 머신 러닝 등을 사용하여 이를 분석하고 정형 데이터로 변환함으로써 정형 데이터를 사용하는 한 가지 모델보다 나은 결과를 얻은 연구들이 있다. 데이터 웨어하우징연구소(The Data Warehousing Institute, TDWI)에서는 텍스트 분석을 통해 콜센터에서 권한(Entity), 테마 및 감정을 추출하고 이를 예측 모델 구축에 사용했다. [1]또 다른 연구는 의료 분야에서 발생하는 텍스트, 이미지, 신호 등과 같은 데이터가 약 80% 구조화나 활용되지 않아 이를 관리하기 위해 유형별 정의, 표준화, 추출, 시각화 시스템에 대한 연구였다.[2] 앞선 선행 연구와 같이 데이터의 정형화에 대한 이유는 다음과 같다. 첫 번째, 비정형 데이터를 유형화 하면 데이터를 분석하고 활용하는 효율성이 크게 증가한다. 정형화된 데이터는 구조화된 형태로 저장되어 쉽게 검색, 분석, 그리고 통합이 가능하다. 이는 실시간 의사 결정, 예측 분석 및 보고서 생성 등에 있어서 중요한 역할을 한다. 두 번째로, 정형화된 데이터는 자동화 시스템 및 여러 소프

트웨어 솔루션과 통합이 가능해진다. 예를 들어, 물류 관리 시스템, 선박 운영 관리 시스템, 보안 시스템 등 다양한 시스템 간의 데이터 교환이 원활해질 수 있다. 이를 통해 작업의 자동화 수준을 높이고, 운영의 일관성을 유지할 수 있다. 세 번째, 비정형 데이터의 유형화는 데이터의 품질과 정확성이 증가한다. 데이터의 일관성을 유지하고 오류를 줄일 수 있으며, 중복 데이터를 제거하고 최신 데이터를 유지하는 데 용이하다. 이는 신뢰할 수 있는 데이터 기반 시스템 설계가 가능하다고 할 수 있다. 마지막으로 데이터 접근 권리관리, 감사 로그 생성, 데이터 암호화 등 보안 조치를 체계적으로 적용할 수 있다. 따라서, 디지털 포트로 전환하는데 있어 비정형 데이터를 정형화 하는 것은 데이터를 보다 효과적으로 활용하고, 운영의 효율성을 높이며, 보안 및 규제 요구사항을 준수하는 데 필수적이다. 이를 통해 항만-선박 운영의 복잡성을 줄이고, 데이터 기반의 스마트한 의사결정을 가능하게 한다. 따라서, 본 연구는 항만 서비스 중 입출항 프로세스 범위 내에서 발생하는 비정형 데이터로 범위를 한정하고 이를 정형화 하는 체계를 구축하는 방법에 대해서 논하고자 한다.

II. 본론

비정형 데이터는 그 형태가 텍스트, 이미지, 영상, 오디오, 센서 등과 같이 다양하며 구조화 되지 않았다는 특징을 가진다. 또 높은 해상도의 이미지나 영상 파일 등과 같은 대용량 데이터를 포함하여 일부 데이터는 실시간으로 처리 및 분석 할 필요가 있다. 따라서 본 연구에서 입출항 프로세스 내 발생하는 비정형 데이터의 범위에 대한 정의는 아래와 같다. 입출항 프로세스 내 비정형 데이터를 정형화하는 체계를 구축하는 방법은 먼저,

데이터를 수집하고 전처리 과정을 거친 후, 데이터를 저장한다. 이렇게 저장된 데이터는 통합되어 시각화 등을 통해 활용된다. 먼저, 이메일이나 CCTV 영상, 사진, 센서, 장비 로그 등 다양한 루트로 여러 종류의 데이터를 수집하고, 수집된 데이터들은 오류나 중복 데이터를 제거하는 데이터 정제 과정을 거친다. 그 뒤 자연어 처리(Natural Language Processing,

를 추출하는데 이미지 인식 기술(Image Recognition)이나 정보 추출(Information Extraction)과 같은 전문적인 기술이 필요하다.

종류	정의
텍스트 데이터 (text data)	항만-선박 운영과 관련된 보고서, 이메일, 메시지, 계약서, 선적 서류 등
이미지 및 비디오 데이터 (image and video data)	CCTV 영상, 드론 촬영 영상, 선박 내외부 사진 등
오디오 데이터 (audio or voice data)	선박-선박, 항만-선박, 항만-항만 간 무전기를 통한 대화, 비상 상황 보고 등에서 발생하는 음성 데이터
센서 데이터 (sensor data)	항만 및 선박에 설치된 다양한 센서에서 수집된 데이터, 기계 상태, 연료 사용량, 위치 정보, 기상 조건 등을 포함하는 데이터

표 1. 입출항 프로세스 내 비정형 데이터 유형

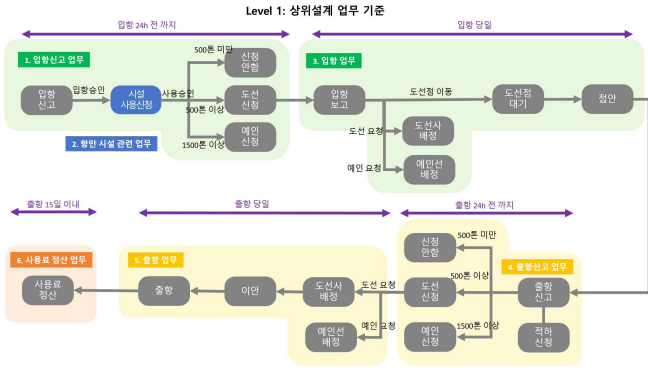


그림 1. 입출항 프로세스 분석 범위

NLP나 이미지 인식 기술(Image Recognition)을 활용하여 처리되는 데이터 변환이 이루어진다. 이러한 과정을 데이터 전처리 과정이라 한다. 이 과정에서 데이터를 항만 또는 선박 산업에 적용되는 국제 표준에 맞추어 변환해야한다. 그렇지 않으면 다양한 출처의 데이터가 서로 다른 형식과 구조를 가지게 되어 데이터 통합이 어렵고 기존과 새로운 시스템 간 호환이 되지 않아 활용성이 떨어지게 된다. 전처리된 데이터들은 SQL 방식의 데이터베이스를 설계하여 저장 및 관리될 수 있으며, 스키마 매핑(Schema Mapping) 등으로 데이터를 통합적으로 활용할 수 있다. 또한 Power BI, NodeXL, Chart.js, Three.js 등을 사용하여 실시간 데이터 시각화 대시보드를 구축하고 분석하여 주기적인 현황 자료를 생성할 수 있다.[3]

현재 항해 및 통신 장비에 대한 기자재 테스트는 문서로 되어 있는 테스트 시나리오를 가지고 정량적인 기준값을 가지고 혹은 정성적인 상황 적합성을 가지고 검증한다. 현재 이러한 테스트 시나리오는 데이터베이스화 되어 있지 않고 인증하는 사람에 의존해서 테스트한다. 이런 경우 비슷한 검증사례가 있으면 반복해서 해당 테스트 시나리오의 문서로 되어 있는 지침으로 기자재 테스트를 할 수밖에 없다. 이런 경우 테스트하는 사람의 주관이 개입되거나 혹은 잘못 판단 할 수 있는 휴먼 에러를 발생시킬 수 있다. 왜냐하면 테스트해야 할 항목이 많은 경우 그러한 경우가 발생하기 때문이다. 따라서 본 연구에서 항해 및 통신 장비 중 ECDIS를 유형화된 데이터베이스로 구현하기 위하여 문서로 되어 있는 것을 시험 단계 구분, 시험 항목, 시험방법, 판정 기준의 필드를 구성하여 데이터베이스화 하였다. 이는 국가법령 기준 “선박용물건 품목별 시험 항목, 시험방법 및 판정기준(제2조 관련)을 기준으로 한 것이다. 두 번째, 데이터 유형화는 선박의 실운항 정보를 데이터 베이스화 하는 것이다. 울산정보산업진흥원에서는 전기추진 선박인 태화호를 운행하여 실제 다양한 15개의 사례를 발생 시간, 위도 경도, 코스 수준, 속도, 깊이, 풍향, 바람 속도, 실행 일자로 테이블을 생성하여 기자재 테스트 시에 운항환경 시나리오를 설정할 때 활용하도록 하였다. 세 번째, Port-MIS의 정보를 자체 데이터베이스화하는 것이 필요하였다. Port-MIS는 입출항 정보를 담고 있는 정보로 공개하고 있다. 항후 선박의 유형과 선적 종류에 따라 재활용할 수 있는 입출항 정보를 얻고자 하는 것이다. 입/출항 시 선박, 에이전트, 터미널 관리사, 도선, 예선사 등의 다양한 이해관계자가 개입되어 있어 커뮤니케이션의 복잡성과 반복성은 피할 수 없다. 하지만 이러한 커뮤니케이션을 유형화한 Port-MIS의 정보를 활용한다면 복잡성과 반복성을 피할 수 있어 이해관계자에게 효율적인 의사전달이 이루어질 것이다. 네 번째, 국제기준 정보인 OSP, CIMAC, IEE, IMO등이 산재 되어 있다. 만약, 항해 및 통신 장비를 테스트한다면 이 중에 적용할 표준이 데이터베이스화되어 일괄적으로 볼 수 있어야 한다. 선박을 건조하거나 선박의 항해 및 통신 장비는 해당 국가의 표준이 우선시 된다. 하지만 해외로 입항하는 선박인 경우 국제 표준을 따르지 않고 개발된 항해 및 통신 장비는 입항 허가가 불허될 수 있다. 예를 들어, 일부 국가에서는 기후적 특성으로 매우 추운 곳으로 항해 및 통신 장비가 최소의 온도를 견뎌야 할 수도 있다. 이런 점에서, 본 연구에서는 국제 표준을 데이터베이스화 하여 항해 및 통신 장비의 테스트에 대응할 수 있게 된다.

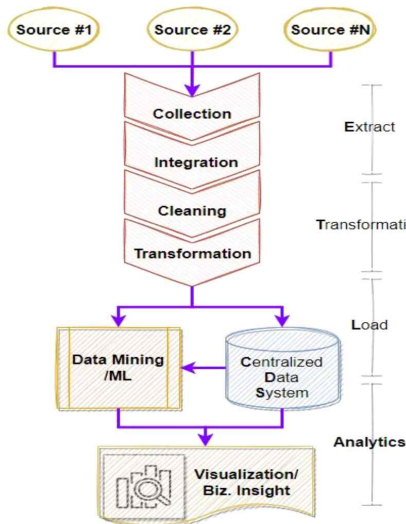


그림 2. 데이터 처리 단계

입출항 프로세스 내에서 발생하는 비정형 데이터의 유형은 아래와 같이 크게 네 가지로 분류할 수 있다. 텍스트 데이터는 항만 또는 선박과 관련해 중요한 정보와 지식을 포함하고 있지만 형식이 일관되지 않아 분석이 어렵고, 이미지 및 비디오 데이터, 오디오 데이터는 수집하고 유용한 정보

III. 결론

본 연구는 입출항 프로세스 내에서 발생하는 비정형 데이터에 대해 유형

화 하는 방법과 필요한 이유에 대해 논하였다. 관련된 비정형 데이터는 텍스트, 이미지 및 비디오, 오디오, 센서 데이터 등으로 유형을 분류하였다. 이중 텍스트로 되어 있는 국가법령 기준 "선박용품건 품목별 시험 항목, 시험방법 및 판정기준(제2조 관련)에 대하여 항해 및 통신 장비 레이더, 전자해도 시스템, 자기컴퍼스, 자이로컴퍼스, 음향측심기, 신속 거리계, 선 회율 지시기, 타각 지시기, 위성항법장치, 선박자동식별장치, 자동조타장치, 항해 기록장치, 선교 항해당직 경보장치 등을 데이터베이스화 하고 있다.

두 번째는 실운항 데이터를 수집하여 데이터베이스화 하고 있지만 향후에는 출발점과 도착점의 위/경도를 입력하여 가상의 운항 데이터를 생성하는 것을 개발할 것이다. 세 번째는 입출항 커뮤니케이션을 최소화(데이터화)할 목적으로 Port-MIS 데이터를 정기적으로 수집하여 인공지능으로 학습하여 최적의 스케줄을 제공하고자 한다.

또한 각 데이터에 맞는 국제 표준 등 글로벌 규정을 적용하여 데이터 관리와 통합할 수 있는 체계를 구축할 것이다.

디지털 포트에 데이터 기반으로 입출항 프로세스를 디지털화하는 과정에서 새로운 시스템의 개발과 운영, 기존의 시스템과 통합 관리 등의 이유로 필수적이다. 하지만 입출항 프로세스 내에서 발생하는 많은 데이터들은 정형화 되지 않은 형태로 생성되고 방치되고 있어 이를 정형데이터와 통합하는 과정이 필요하다. 따라서, 입출항 프로세스 내에서 발생하는 비정형 데이터를 수집하고 분석해 이를 활용하기 위한 연구를 진행하였다. 하지만 국내에는 항만-선박 간 발생하는 비정형 데이터에 대한 관심이 부족해 데이터의 수집조차 쉽지 않았으며, 데이터의 소유, 보안 등에 대한 관리도 이루어지지 않고 있었다. 디지털 포트 실현은 정형 또는 비정형 데이터의 관계, 체계 구축과 함께 운영의 효율성, 안정성 그리고 지속 가능성을 크게 향상 시켜 해운 및 물류 산업 전반에 걸쳐 큰 혁신이 가능하게 할 것이다.

ACKNOWLEDGMENT

This study is part of a paper under the UIPA-led project for verifying maritime logistics communication equipment.

참 고 문 헌

- [1] Fern Halper, (2018). 3 Use Cases for Unstructured Data. TDWI Orlando Leadership Summit. 2018.
- [2] Hyoun-joong Kong. (2019). Managing Unstructured Big Data in Healthcare System. US National Library of Medicine, National Institutes of Health, January, 31, 2019.
- [3] Byung-chang Son, (2023). The meaning of big data and understanding of traditional data processing structures, S-Core Insight Report · Issue & Trend, February, 2023.