

선박예지정비모델 개발을 위한 LNG 선박 도크 수리 항목의 텍스트 분석 연구

황태민* · 윤익현** · 오정모****

* 목포해양대학교 해상운송시스템학부 석사과정, ** 목포해양대학교 항해정보시스템학부 교수, *** 목포해양대학교 기관시스템공학부 교수

Study on Text Analysis of the Liquefied Natural Gas Carriers Dock Specification for Development of the Ship Predictive Maintenance Model

Taemin Hwang* · Ik-Hyun Youn** · Jungmo Oh****

* Graduate Student, Department of Maritime Transportation System Mokpo National Maritime University, Mokpo 58628, Korea
** Professor, Division of Navigation & Information Systems, Mokpo National Maritime University, Mokpo 58628, Korea
*** Professor, Division of Marine Engineering, Mokpo National Maritime University, Mokpo 58628, Korea

요 약 : 다양한 산업에서 강조되고 있는 정비의 중요성은 각 분야에 다양한 정비전략을 적용하도록 만들었다. 해양산업 역시 그에 따른 정비전략의 변화가 있었으나 타 산업 대비 그 속도가 느려 실제 적용이 되지 않은 채 과거 시행되고 있던 방식을 유지하는 경우가 많다. 특히 선박은 기존에 행해왔던 방식의 정비전략을 사용하고 있는 편이며 해상의 조건에서 선박은 새로운 정비전략의 개발을 필요로 하고있다. 이에 선박예지정비모델은 기기의 정비가 필요한 시점을 예지하여 조치 할 수 있는 정비전략으로서 선박이 항해 중에 처할 수 있는 정비 관련 위험요소들을 줄여 주는 모델이다. 본 연구는 선박예지정비모델의 개발을 위한 연구 중의 하나로서, LNG선박 입거사양서의 텍스트 데이터 분석을 통한 결과를 원문의 내용을 바탕으로 해석해보았다. 공통된 정비항목 조합을 도출하여 선박 내 다른 기기들 사이에 작용하고 있는 상호연관성을 발견하고 이를 앞으로 개발될 선박예지정비모델에 적용하고자 한다.

핵심용어 : 정비전략, 선박예지정비모델, 텍스트 분석, 상호연관성, 토픽 분석

Abstract : The importance of maintenance is leading the application of the maintenance strategy in various industries. The maritime industry is also a part of them, with changes in selecting and applying the maintenance strategy, but rather slowly, by retaining the old strategy. In particular, the ship is maintaining a previously used strategy. In the circumstance of the sea, ship requires a new suggestion for maintenance strategy. A ship predictive maintenance model predicts the breakdown or malfunction of machineries to secure maintenance time with preventive actions and treatments, thereby avoiding maintenance-related dangerous factors. This study focused on applying text analysis to an Liquefied Natural Gas Carriers dock indent document, and the analysis results were interpreted from the original document. The inter-relational patterns observed from the frequency of common maintenance combinations among different parts and equipment in ships will be applied to the development of ship predictive maintenance.

Key Words : Maintenance strategy, Ship predictive maintenance model, Text analysis, Inter-relational pattern, Topic analysis

1. 서 론

정비는 그 자산을 복원하는 행위 또는 원래의 상태를 유지하는 행위의 조합이다(Emovon et al., 2018). 이것은 마치 사

람에게 건강 유지가 중요하듯 기기에게도 초기의 상태를 유지하도록 체계적인 관리가 필요함을 말한다. 어떻게 관리하는지가 기기 수명에 큰 영향을 끼친다는 것이다. 다양한 산업현장에서 기기 정비의 중요성은 안전확보 측면에서나, 효율성, 경제성을 위해 꾸준히 강조되어왔다. 그 중 해양산업 분야는 해상에 노출되는 경우가 많은 환경의 특성상 기기 정비에 취약하다. 특히, 선박에서 정비란 핵심 기술적 측면

* First Author : cjswo21kr@naver.com, 061-240-7281

† Corresponding Author : jmoh@mmu.ac.kr, 061-240-7207

으로, 기기가 요구하는 이용성, 확실성, 효율성을 유지하는 것을 말한다. 상선의 기기체계는 결정적인 역할을 하므로 부적절한 정비는 전손을 포함한 심각한 결과를 초래한다 (Kandemir and Çelik, 2019).

정비작업은 해당 장비가 디자인된 기능을 계속할 수 있게 유지하거나 수리하는 행위들의 조합으로 정의된다. 해양분야의 기기나 관련 장비의 정비는 보통 3가지로 구분되는데, 첫째로 계량 정비(Corrective maintenance) 또는 사후 정비(Reactive maintenance)라고 불리는 방법으로 기기가 먼저 고장 난 후 수리를 행하는 것을 말해, 기기의 중요도가 낮거나 거의 없는 기기들에 사용된다. 둘째로는 예방정비(Preventive maintenance)가 있으며, 기기의 통계적 수명에 따라 계획적으로 시행하는 방법으로, 기기의 사용 수명을 늘리고 치명적인 고장의 위험성을 완화할 수 있다. 예방정비는 작동 중인 장비를 부품의 교체, 검사, 시험, 확인 등의 작업으로 정해진 시간 간격을 두고 정비하는 것을 말하는데, 보통 장비 관리자들은 제조사의 추천과 경험에 따라 정해진 일정에 맞추어 정비하는 것이 그 예다. 예방정비의 가장 큰 장점으로서는 장비의 평균 수명을 늘려주고 치명적인 고장의 위험성을 줄여준다. 그러나 단점으로는 해당 정비가 적시가 아닐 경우 불필요한 수리작업만 일으킨다는 점이 있고 더 나아가 적절한 예방정비의 수준을 정하기 위해서는 수년에 걸친 데이터 수집과 분석이 필요하다는 점이다. 셋째로 상태기준정비(Condition based maintenance)가 있는데, 장비의 작동 상태를 관찰하며 정비를 시행하는 것을 말한다(Emovon et al., 2018).

이렇듯 정비에 관한 연구가 꾸준히 진행됨에 따라 현재는 예측정비(Predictive maintenance)가 원자력발전이나, 우주항공 분야 등 다양한 산업 분야에서 널리 통용되고 있으나 해양 산업 분야는 예외다. 현재 선박에서 가장 많이 행해지고 있는 정비는 예방정비로 분류할 수 있다(Gkerekos et al., 2017).

해상에서 운항 중인 선박의 기기 고장은 단순한 고장 그 이상의 의미를 가지며, 그 예방을 위한 정비전략 선정은 까다로울 수밖에 없다. 게다가 모든 정비가 각각의 기기 구성 요소에 적용 가능하고 효율적인 것은 아니므로 적절한 정비의 전략을 선택하는 것부터가 정비 관리에서 어려운 점이다. 해양업계의 다른 기기/장비들의 정비전략 선정은 그들의 특정 고장 상태 등을 고려해야 하는 복잡한 과정이며 대체로 여러 선정기준을 동반한다. 해양업계의 학술 분야 및 산업 분야에 종사하는 해양기술 전문가들을 대상으로 시행된 개인 인터뷰에 따르면, 해양 관련 기기들의 정비전략 선택에는 22개의 결정기준이 고려되어야 한다고 한다(Emovon et al., 2018).

이처럼 정비전략의 결정기준이 많은 이유는 선박이 기기와 설비의 집합체이기 때문이다. 선박은 그 목적에 맞게 다

양한 기기 구성요소가 있는데, 이를 크게 화물설비, 전기설비, 기관장치 등으로 나눌 수 있다. 화물설비로는 화물격납설비, 압력용기, 압력관 장치, 압력 및 온도 제어장치, 벤트 장치, 동력통풍장치 등이 있고 전기설비로는 회전기계, 배전반, 구 전반, 분전반, 케이블, 변압기 등이 있으며, 기관장치로는 주기관, 보조 기관, 발전기, 추진축 계, 동력전달장치, 프로펠러, 보일러, 가열기, 압력용기, 공기관, 측심장치, 밸브 평형수 장치, 연료유 장치, 조타장치, 계류장치 등이 있다 (KR-Rules & Guidance, 2019).

이러한 선박의 복합적인 구조에서 그 내부의 기기 간의 상호연관성은 선박의 정비에 고려될 필요성이 있다. 예를 들어 주기관과 보조 기관은 서로 간접적으로 연결되어 있어 하나에 고장이 발생하면 다른 곳에도 부정적인 영향을 미친다(Goldsworthy and Galbally, 2011).

본 연구는 해상환경에서 운항되는 선박의 특성을 반영한 선박예지정비모델 개발을 위한 것으로, 기기 간의 상호연관성을 정비항목별 작업리스트의 텍스트 데이터 분석으로 알아내고자 한다. 본 연구에서 사용한 텍스트 데이터는 선박의 수리 도크 입거 시 작성하는 Dock indent로, 구체적으로는 LNG운반선 15척의 실제 도크정비 데이터를 사용하였다. 해당 데이터를 사용한 이유는 도크에서 이뤄지는 작업은 항해 중에는 할 수 없는 정비를 포함한 선박의 전반에 걸친 중요 정비들이 주를 이루기 때문이며, 앞서 언급한 선박의 구성요소의 큰 틀을 안팎으로 기기 간 상호연관성을 찾고자 한다.

2. 텍스트 데이터 분석

2.1 토픽 분석과 선박예지정비모델

잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)를 기반으로 한 토픽모델은 자연어 처리, 텍스트 마이닝, 소셜미디어 분석, 정보 조회 등에 사용되고 있다. 소셜미디어 기관이 온라인 커뮤니티에서의 사람들의 반응, 대화를 분석하는 것을 예로 들 수 있으며, 그러한 상호작용을 통해 유용한 패턴을 찾거나 더 나아가 트위터나 페이스북 같은 소셜미디어에서 어떤 것들이 공유되는지까지도 파악할 수 있다. 토픽 모델은 분리된 별개의 데이터들에도 사용될 수 있는데 이는 방대한 정보 속에서 숨겨진 구조나 의미를 찾아내는 데 사용되기도 한다(Jelodar et al., 2019). 이를 이용한 토픽 분석은 선박의 정비요소 간의 상호연관성 데이터를 도출해 내어, 앞으로 개발될 선박예지정비모델에 적용될 예정이다. 선박예지정비모델은 해당 기기가 정비를 필요로 하는 시점을 예측하여 기기의 오작동 또는 고장을 미리 방지할 수 있는 정비모델이다. 선박예지정비모델이 특정 기기의 정비 필요성

을 예지할 때 고려할 조건 중 하나로 기기 상호연관성을 넣으면 모델의 정확도를 높일 수 있을 것이다.

2.2 텍스트 분석을 위한 전처리 과정

본 연구의 텍스트 분석 과정은 원본 문서의 데이터 전처리, 분석을 위한 데이터 선택, 토픽모델 학습, 키워드의 빈도 분석, 원본 문서를 통한 해당 키워드 검증으로 진행되었다. 전처리 과정을 포함한 연구의 진행 과정은 Fig. 1과 같다.

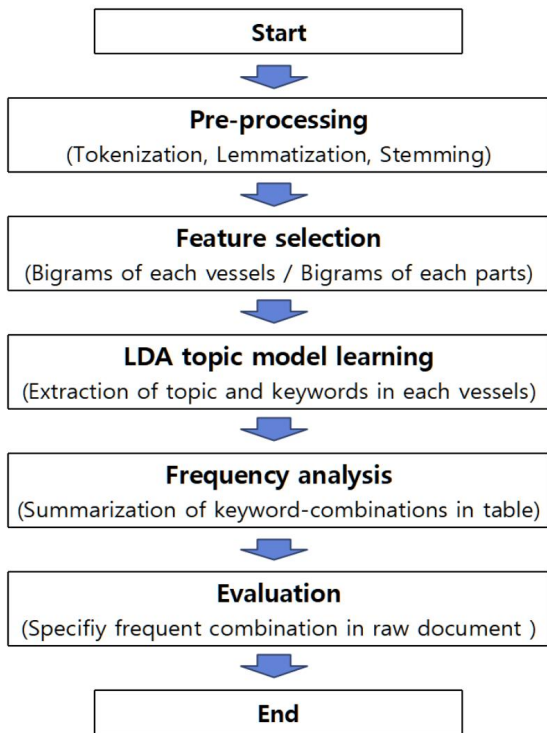


Fig. 1. Text data analysis flowchart.

본 연구의 데이터 분석에 사용된 LNG 선박의 Dock indent 원본문서는 Membrane타입 LNG선박의 데이터를 사용하였다. 선박의 선령별 척수는 입거 당시 선령 10년 미만 2척, 15년 미만 6척, 20년 미만 6척, 20년 이상이 1척으로 총 15척이다. 연구에 있어 기기간의 상호연관성에 집중 하였고 선령으로 인한 차이는 고려하지 않았다.

통상적으로 도크에서 이뤄진 작업에 따라 통상 7개 파트로 나뉘어 있으나, 선박예지정비모델의 개발을 위한 분석이므로 General service, Particular, Paint 파트 등은 제외한 Cargo, Electric, Hull, Machinery의 4개 파트만 사용하였다. 사용된 각 파트 데이터의 내용으로는 Cargo의 Cargo tank, Spray system, Cargo valve, Cargo pump, Heat exchanger 등의 장비에 대한 수리 항목을 포함하고 있으며, Electric파트에는 Generator,

Switch board, Battery charger, Power supply unit, Cable, Board, Controller 등의 전기 시스템의 수리 항목이 있고, Hull 파트에는 선박의 Anchor chain, Bottom, Topside, Rudder, Trunk, Funnel, Sea chest 등의 선박의 구조적 수리 항목이 있으며, Mach파트에는 Engine, Generator, Windlass, Propeller shaft 등의 기관과 갑판 기기들의 수리 항목을 포함하고 있다. 원본문서의 텍스트 데이터로 전처리한 과정은 토큰화(Tokenization), 표제어 추출(Lemmatization), 불용어(Stop words)제거 및 추가로 LDA Topic model의 결과에 큰 의미를 갖지 않는 접사, 판사 제거 작업, 알파벳의 소문자 통일 등을 진행하였다. 먼저 토큰화(Tokenization)는 텍스트로 된 내용의 흐름을 토큰이라는 단어, 용어, 기호 또는 다른 형태의 의미가 있는 어떤 요소들로 끊는 작업을 말한다(Vijayarani and Janani, 2016). 이는 문서가 가지고 있는 문장들을 문법적으로 의미를 갖는 가장 작은 단위로 나누는 과정이고, 표제어추출(Lemmatization)은 같은 형태학적인 패러다임에 속한 단어들을 그룹화하여 표제어라는 정형된 틀의 한 형태에 각각 할당하는 것을 말한다(Gesmundo and Samardžić, 2012). 다시 말해, 단어의 시제, 태 등의 이유로 변화된 형태를 사전적 표제어의 형태로 돌려놓는 과정이다.

이 전처리 과정을 통해 추출된 데이터들은 Bag-Of-Word기법과 Bag-Of-N-gram기법으로 정리될 수 있는데, 텍스트의 통계적 자연어 처리 기법 중 대표적으로 통용되고 있는 Bag-Of-Word(BOW)는 문장 또는 문서의 순서와 상관없이 고정된 값의 길이를 갖는 단어들의 모음을 출현 빈도수로 구분하는 기법이다. BOW는 토픽 도출이나 장문에서의 단어 간 관계 파악에 효과적인 것으로 나타났으며, 이는 문서를 특정 정수값으로 고정된 길이의 단어들로 단축함으로써 자연어 처리 기법 중 벡터 기반의 모델링 기법인, 문서 분류법, 정보 검색 등에 쓰인다. BOW와 달리 N-gram 모델은 자동음성 인식 기술이나 통계기반 기계번역과 같은 단어의 순서와 위치를 고려하는 기술에 널리 사용되며, BOW 모델에서 파생될 수 있는 단어 간의 관계 및 토픽 정보와 N-gram모델의 통합은 언어모델의 적용에 유망한 방법이다(Sethy and Ramabhadran, 2008). 본 연구에서는 N-gram기법 중 Bi-gram을 이용하여 분석하였다.

2.3 파트 구분 데이터

추출된 Bi-gram들이 갖는 의미는 단어의 사전적 의미보다도 그 단어의 출처 문서가 Cargo파트, Electric파트, Hull파트, Machinery파트 중 어디인지가 중요했다. 분석 목적에 맞는 텍스트 데이터들은 Dock indent 내의 원래 출처 파트의 성질을 유지해야 했으므로 토픽모델의 적용에 앞서 추후 토픽모델이 추출한 단어의 해당 파트가 무엇인지를 구분해 줄 파

트 구분 데이터가 필요했다. 파트 구분 데이터는 우선 15척의 모든 선박의 파트별 4개 문서를 각각 통합하여, 모든 선박의 각 파트에서 추출될 수 있는 Bi-gram들을 따로 취합하였고, 이 Bi-gram들 중 서로 다른 파트 간 중복되는 Bi-gram들은 해당 파트를 대표할 수 없기에 제외했다.

2.4 LDA토픽 및 키워드 생성과정

15척의 선박 중 선박 A의 모든 Bi-gram들을 LAD 토픽모델을 이용하여 토픽과 그에 해당하는 키워드 묶음으로 추출하였는데 토픽의 수는 3개로 설정하였고 키워드 수는 한 토픽당 100개로 설정하였다. 여기서 토픽의 수는 3개 이상으로 설정하였을 때 각 토픽 간의 Probability 값 구분이 거의 되지 않을 만큼 낮아 3개가 적당하였고, 토픽당 키워드의 수는 100개 미만으로 설정했을 때, 파트구분 데이터로 걸러 내면 문서의 크기 차이로 상대적으로 텍스트 데이터가 많은 Cargo파트와 Machinery파트의 데이터만 남아 Electric파트와 Hull파트의 조합 데이터가 상대적으로 너무 적어져 토픽 내 키워드들의 Probability값이 다양하게 유지되는 선에서 100개 이상으로 설정하였다. 추출된 Bi-gram들은 테이블의 형태로 Keyword와 Probability 값으로 정리되었는데 여기서 Probability 값은 키워드의 추출 과정에만 고려하였고, 추후 결과 분석에서는 Probability 값보다 선박 간 중복으로 출현하는 빈도수에 더 의미를 두고 분석하였다.

2.5 키워드 파트 구분 및 파트별 조합 생성과정

한 선박의 LAD 토픽모델에서 추출된 Bi-gram 키워드들은 파트 구분 데이터를 바탕으로 구분하여 테이블에 정리하였다. 다른 토픽과 섞이지 않게 한 테이블에 한 토픽 안의 키워드들을 파트별로 구분하였다. 이 한 테이블의 단어들은 LAD 토픽모델이 같은 토픽으로 구분한 단어들로서 그 안의 파트들 간의 조합을 생성하였다. 조합들은 Cargo 파트와 Electric, Hull 파트와 Machinery 파트, Cargo 파트와 Hull 분야, Electric 파트와 Hull 파트, Cargo 파트와 Machinery 파트, Electric 파트와 Machinery 분야 이렇게 2개 파트씩 6개 조합으로 생성하였다. 한 선박의, 한 토픽 안에서 추출된 키워드들의 파트별 조합을 만든 후에 결과 분석 단계에서도 그 조합과 키워드가 어느 선박의, 어느 토픽에서 추출되었는지 알 수 있게 정리하였다. 그리고 위 과정을 토픽마다, 선박마다 같은 방식으로 반복하였다.

2.6 타 선박들로부터의 중복 빈도수 도출

추출된 한 선박의 Bi-gram키워드 분야 조합들을 정리한 테이블의 마지막 열에는, 같은 방식으로 정리된 나머지 14척 선박들의 데이터 테이블에서, 해당 조합이 몇 번이나 출현

하였는지 알 수 있도록 빈도수 값이 추출되도록 하였다. 결과 분석을 위해 해당 테이블은 빈도수 값이 높은 순에서 낮은 순으로 정리되도록 하였다.

2.7 높은 빈도수의 키워드들에 관한 사례 검증

정리된 테이블의 데이터들은 높은 빈도수의 조합 순으로 그 Bi-gram이 원본 문서에서 어떤 의미가 있는지 분석하였다. 해당 Bi-gram조합은 전처리 과정과 분야 구분 데이터를 통해 걸러진 후 추출되었기에 원본문서 내에서 흔하게 등장하는 단어가 아닌 특정 정비작업 내용을 의미한다. 하나의 조합에서 각 Bi-gram이 의미하는 파트의 특정 정비작업들이 높은 빈도수를 보인 이유를 검증하여 분석의 결과를 실제 사례로 해석하였다.

3. 텍스트 데이터 분석 결과 및 검증

3.1 전처리 전과 후의 단어 형태 변화

전처리는 단어들이 실제로는 같은 의미를 갖고 있으나, 원문에서 형태가 달라 다른 단어로 취급되었던 단어들을 통일화 하는게 주 목적이었다. Table 1은 원문의 일부분이 전처리 과정을 통해 변화된 과정을 단계로 정리한 모습으로 각 원문의 한 문장이 전처리 과정을 통해 형태가 바뀌는 모습을 볼 수 있다.

Table 1. Results of pre-processing

Stage	Contents
Raw document	1. All liquid domes & gas domes manhole covers will be opened & restored.
Tokenization	'1', '.', 'All', 'liquid', 'domes', '&', 'gas', 'domes', 'manhole', 'covers', 'will', 'be', 'opened', '&', 'restored', '.'
Lower words	'1', '.', 'all', 'liquid', 'domes', '&', 'gas', 'domes', 'manhole', 'covers', 'will', 'be', 'opened', '&', 'restored', '.'
Lemmatisation	'liquid', 'dome', 'gas', 'dome', 'manhole', 'cover', 'open', 'restore'
Bi-grams	'liquid dome', 'dome gas', 'gas dome', 'dome manhole', 'manhole cover', 'cover open', 'open restore'.

3.2 파트 구분 데이터 구축 결과

파트 구분 데이터는 각 파트를 대표할 수 있는 Bi-gram들이며 각 파트별 내용을 Table 2에 정리하였다.

Table 2. Part separation data

Part	Bi-grams
Cargo	'cargo equipment', 'cover restore', 'lamp cargo', 'protective rubber', 'upper hemisphere', 'cable tank', 'seam pipe', 'inner construction', 'fitting instrument', 'float guage', ...
Electric	'board main', 'contact main', 'lubrication move', 'operational air', 'board panel', 'machinery cabinet', 'anode assembly', 'remote indicator', 'cryogenic power', 'heat lamp', 'torsion sensor'...
Hull	'hull paint', 'high jetwashing', 'abrasion paint', 'coat epoxy', 'engine funnel', 'bulkhead tank', 'anchor chain', 'chain locker', 'mooring winch', 'tank guage', 'line compressor'
Machinery	'main thrust', 'seal ring', 'tube oil', 'wheel generator', 'gasket oring', 'cir pump', 'renewal gasket', 'air reservoir'

3.3 LDA토픽 및 키워드 생성결과

선박A의 경우, 각 토픽의 결과로 나온 Bi-gram들은 아래와 같으며 그 중 일부를 Table 3에 정리하였다.

Table 3. LDA result keywords from vessel 'A'

Topics	Keywords in bi-grams
1	'safety valve', 'power supply', 'gland pack', 'gasket renew', 'tank penetration', 'control valve', 'sea chest', 'main condenser', 'propeller blade', 'electric cable'
2	'bolt nut', 'gasket renew', 'gland pack', 'press guage', 'heat exchanger', 'life boat', 'valve overhaul', 'supply fan', 'ballast tank', 'sea chest', 'fan motor', 'valve seat'
3	'power supply', 'press guage', 'heat exchanger', 'valve seat', 'pump motor', 'sea chest', 'propeller shaft', 'aux pump', 'intermediate shaft', 'stern tube', 'turbine case'

3.4 키워드 파트 구분 및 파트별 조합

선박A의 토픽1에서 도출된 Bi-gram들은 파트 구분 데이터를 통해 걸러져 Table 4와 같이 정리 하였다. 이 Bi-gram들은 파트별 조합의 경우를 모두 고려하여 총 8102개의 조합으로 도출되었으며 그 중 타 선박에서 중복으로 나타나는 경우의 수를 빈도수 값으로 정리하였다.

Table 4. Part separation result of Topic 1 from Vessel 'A'

Part	Bi-grams
Cargo	'disch valve', 'dome cover', 'flame arrester', 'gas sample'...
Electric	'radar limiter', 'auto pilot', 'gyro compass', 'limiter receiver'...
Hull	'moor winch', 'pintle bush', 'wire rope', 'damage crack'...
Machinery	'draft fan', 'governor take', 'guard valve', 'input shaft'...

3.5 타 선박들로부터의 중복 빈도수 확인

결과 빈도수는 아래 Table 5에 정리하였다. 테이블의 내용은 선박 A의 토픽1에서 파트 간 조합의 리스트와 해당 조합이 타 선박에서 등장한 횟수이다. 그 결과로 나타난 선박A의 빈도수 값은 본선을 제외한 나머지 서로 다른 14척 중에 도출된 결과이므로 과반수 이상을 보인 조합을 중점으로 분석하였다. 이후 빈도수가 높게 나타난 조합들에서 기기 상호연관성이 존재하는지에 대한 검증과 해석을 하였는데, 이는 실제 원문에서 해당 Bi-gram이 속한 내용이 어떤 작업인지 찾는 것으로 진행하였다. 여기서 원문의 작성방법에 따라 다소 차이가 있었으나 결과의 해석에 방해가 될 정도는 아니었다.

Table 5. Combination frequency throughtout vessels

Cargo	Electric	Hull	Machinery	Frequency
cargo tank	-	-	seat dye	8
gas sample	-	-	seat dye	8
sample line	-	-	seat dye	8
deformation center	-	-	seat dye	8
center bellow	-	-	seat dye	8
tank penetration	-	-	instruction manual	7
deformation center	-	-	valve seat	7
center bellow	-	-	valve seat	7
-	-	wire rope	seat dye	7
cargo tank	radar limiter	-	-	6

3.6 높은 빈도수 키워드들에 대한 사례 검증결과

Table 5에서 볼 수 있듯, 선박 A가 보인 높은 빈도수의 조합들은 ‘cargo tank-seat dye’, ‘gas sample-seat dye’, ‘sample line-seat dye’, ‘deformation center-seat dye’, ‘center bellow-seat dye’, ‘tank penetration-instruction manual’, ‘deformation center-valve seat’, ‘center bellow-valve seat’ 등의 조합이었다. 각각 ‘cargo tank-seat dye’ 조합이 15척의 선박들 중에서 8번, ‘gas sample-seat dye’ 조합이 8번, ‘sample line-seat dye’ 조합이 8번, ‘deformation center-seat dye’ 조합이 8번, ‘center bellow-seat dye’ 조합이 8번의 빈도수를 나타내었다. 전체 15척 선박 중 8번의 빈도수는 그 상대적 비중이 크지 않다고 사료될 수 있으나, 선령을 포함한 각기 다른 조건의 선박들임을 감안하면 과반수 이상의 결과는 의미가 있다 판단하여 분석을 진행하였다. 해당 조합들이 실제 원문에서 어떤 작업을 의미하는지 알아보기 전에 본 연구에서 사용된 원문의 예시는 Fig 2와 같은 형태이다.

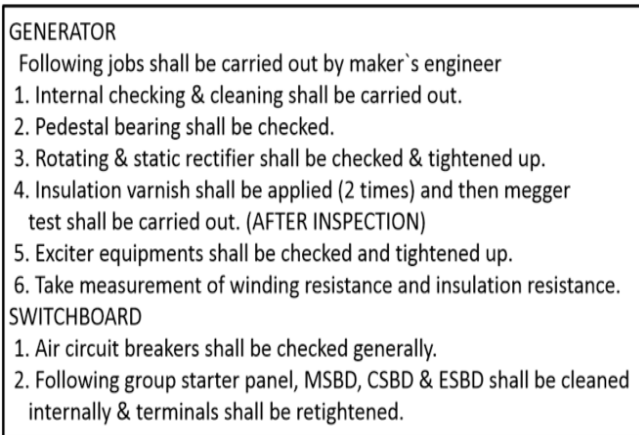


Fig. 2 Example of original dock indent document.

Fig 2는 본 연구에 사용된 원문의 일부로서 2015년 입거 수리를 행한 선박A의 Dock indent 중 Electric파트의 작업내용이다. 원문은 실제 작업의 과정을 구체적으로 기술하고 있어 해당 분석 결과로 도출된 특정 Bi-gram이 속한 작업 내용을 찾아보면 어떤 기기의 어떤 작업인지 알 수 있었다. 먼저 첫번째 조합은 “cargo tank-seat dye”는 Cargo파트와 Machinery파트의 조합이었다. 이는 원문에서 Cargo파트의 ‘Inner construction, fittings and instrument in all cargo tank will be visually inspected by ship's staff and class surveyor.’ 작업과 Machinery파트의 Main feed water pump 작업과 관련된 ‘Governor valve shall be overhauled and valve & seat shall be dye-checked.’ 작업을 의미하였다. 두번째 조합 ‘gas sample-seat

dye’은 Cargo파트의 ‘Tank penetration of gas sampling line’ 작업을 의미하였고, Machinery파트의 ‘Governor valve shall be overhauled and valve & seat shall be dye-checked’ 작업을 의미하였다. 그리고 세번째 조합 ‘deformation center-valve seat’은 Cargo파트의 high duty gas compressor(No.2) 기기와 관련된 ‘Deformation & centering of bellows shall be measured’ 작업을 의미하였고 Machinery파트의 ‘Governor valve shall be overhauled and valve & seat shall be dye-checked’ 작업을 의미하였다.

4. 결론

4.1 테이블에서 빈도가 높은 작업들의 해석

선박 A의 빈도수가 높은 조합의 원문형태의 의미를 해석한 결과, 사실상 각 파트별 큰 하나의 작업의 일부분이었다. Cargo파트의 화물창 점검 작업의 각 밸브당 화물창으로의 침투 여부를 확인 하는 작업과 Machinery파트의 Main feed water pump의 각 밸브의 Overhaul과 Dye체크 작업을 의미하였는데, 이는 선박 내에서 다른 파트에 속해있는 화물창과 Main feed water pump의 상호연관성을 보여주었다. 마찬가지로 Hull파트의 Lift기기의 wire rope교환 작업도 Machinery파트의 Main feed water pump정비 작업과 상호연관성을 보였다. 추가적으로 본 논문에서는 선박A의 한 토크만을 예로 들어 기술하였지만, 선박A의 다른 토크에서도, 선박A가 아닌 다른 선박들에서도 유사한 결과를 보였다. 이는 선박에서 기기가 서로 다른 파트에 포함되어 따로 작동하고 있지만 동시에 상호연관성을 갖을 수 있음을 보여주는 결과였다.

4.2 연구결과의 의미있는 부분

본 연구에서 상호연관성을 해석하는데 의미 있는 부분은, 빈도수가 높은 조합의 단어들이 원문에서 여러 작업에서 동시에 나타나는 혼한 단어가 아니었다는 것이다. 이는 단지 원문에서 혼하게 사용되는 단어여서 빈도수가 높아지는 결과를 우려했던 것과 다르게, 하나의 큰 정비작업에서만 도출되어 본 연구의 목적에 맞는 결과를 보였다. 본 연구는 시작부터 선박예지정비모델의 개발을 위한 목적을 두고 선박의 기기 간 상호연관성을 텍스트 분석으로 진행하였다. 연구 과정에서 부족했던 점으로는 원문의 전처리 과정에서 단어의 손실이 없도록 더욱 세심하게 분석하였다면 더 나은 결과를 보였을텐데 그러지 못한 것이다. 기회가 된다면 다음 연구에서는 더 많은 선박의 데이터로 보다 세심한 분석을 통해 더 뚜렷한 결과를 확인할 수 있기를 바라고 있다.

후 기

본 연구는 산업통상자원부와 한국산업기술진흥원의 광역협력권산업육성사업으로 수행된 결과입니다(과제번호: P0008664).

Received : 2020. 12. 01.

Revised : 2020. 12. 14. (1st)

: 2020. 12. 21. (2nd)

Accepted : 2020. 12. 24.

References

- [1] Emovon, I., R. A. Norman, and A. J. Murphy(2018), Hybrid MCDM based methodology for selecting the optimum maintenance strategy for ship machinery systems, *Journal of intelligent manufacturing*, 29(3), pp. 519-531.
- [2] Gesmundo, A. and T. Samardžić(2012), Lemmatisation as a tagging task, In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Vol. 2: Short Papers)*, pp. 368-372.
- [3] Gkerekos, C., I. Lazakis, and G. Theotokatos(2017), Ship machinery condition monitoring using performance data through supervised learning, *Proceedings of the 2017 Smart Ship Technology Conference*, ISBN 9781909024632, pp. 105-111.
- [4] Goldsworthy, L. and I. E. Galbally(2011), Ship engine exhaust emissions in waters around Australia-an overview, *Air Quality and Climate Change*, 45(4), p. 24.
- [5] Jelodar, H., Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao(2019), Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, *Multimedia Tools and Applications*, 78(11), pp. 15169-15211.
- [6] Kandemir, C. and M. Çelik(2019), A human reliability assessment of marine auxiliary machinery maintenance operations under ship PMS and maintenance 4.0 concepts, *Cognition, Technology & Work*, 22(3), pp. 473-487.
- [7] KR(2019), Korea Register, 2019 KR-Rules & Guidance, www.krs.co.kr (Accessed 11 January 2021).
- [8] Sethy, A. and B. Ramabhadran(2008), Bag-of-word normalized n-gram models, *INTERSPEECH 2008, 9th Annual Conference of the International Speech Communication Association*, Brisbane, Australia, September 22-26, 2008.
- [9] Vijayarani, S. and R. Janani(2016), Text mining: open source tokenization tools-an analysis, *Advanced Computational Intelligence: An International Journal (ACIJ)*, 3(1), pp. 37-47.