

# 빅데이터 기법을 활용한 항해 중 준해양사고 발생원인 분석에 관한 연구

강석용\* · 김기선\*\* · 김홍범\*\*\* · 노범석\*\*\*\*

\* 한국해양수산연수원 시험관리팀, \*\* 한국해양수산연수원 해사안전팀,  
\*\*\* 한국해양수산연수원 오션폴리텍팀, \*\*\*\* 한국해양수산연수원 교육운영팀

## An Analysis of Causes of Marine Incidents at sea Using Big Data Technique

Suk-Young Kang\* · Ki-Sun Kim\*\* · Hong-Beom Kim\*\*\* · Beom-Seok Rho\*\*\*\*

\* Examination Administration Team, Korea Institute of Maritime and Fisheries Technology, Busan 49111, Korea

\*\* Marine Safety Team, Korea Institute of Maritime and Fisheries Technology, Busan 49111, Korea

\*\*\* Ocean Polytech Team, Korea Institute of Maritime and Fisheries Technology, Busan 49111, Korea

\*\*\*\* Education & Operation Team, Korea Institute of Maritime and Fisheries Technology, Busan 49111, Korea

**요 약 :** 해양사고 감소를 위해 다양한 연구들이 수행되어 왔다. 그에 비해 준해양사고에 대한 연구는 미미한 수준에 그치고 있다. 준해양사고는 건수가 많은 대신 내용이 정성적이기 때문에 분석하기에는 현실적인 어려움이 있었다. 하지만 해양사고 감소를 위해서는 준해양사고의 정량적인 분석이 필요하다. 이번 논문의 목적은 준해양사고 경향을 예측하고 해양사고를 감소시키기 위해 빅데이터 기법을 적용하여 준해양사고 데이터를 정량적으로 분석하는 것이다. 이를 위해 10,000여건의 준해양사고 보고서를 전처리 작업을 통해 통일된 양식으로 정리하였다. 전처리된 데이터에 대해서 1차적으로, 텍스트마이닝 기법을 적용하여 항해 중 준해양사고 발생원인에 대한 주요 키워드를 도출하였다. 주요 키워드에 대해 2차로 시계열 및 클러스터 분석을 통해 발생할 수 있는 준해양사고 상황에 대한 경향 예측을 도출하였다. 이번 연구에서는 정성적 자료인 준해양사고 보고서를 빅데이터 기법을 활용하여 정량화된 데이터로 전환할 수 있고, 이를 통해 통계적 분석이 가능함을 확인하였다. 또한 빅데이터 기법을 통해 차 후 발생할 수 있는 준해양사고에 대한 객관적인 경향을 파악함으로써 예방 대책에 대한 정보 제공이 가능함을 확인할 수 있었다.

**핵심용어 :** 준해양사고, 정량 분석, 빅데이터, 텍스트마이닝, 시계열 및 클러스터 분석

**Abstract :** Various studies have been conducted to reduce marine accidents. However, research on marine incidents is only marginal. There are many reports of marine incidents, but the main content of existing studies has been qualitative, which makes quantitative analysis difficult. However, quantitative analysis of marine accidents is necessary to reduce marine incidents. The purpose of this paper is to analyze marine incident data quantitatively by applying big data techniques to predict marine incident trends and reduce marine accident. To accomplish this, about 10,000 marine incident reports were prepared in a unified format through pre-processing. Using this preprocessed data, we first derived major keywords for the Marine incidents at sea using text mining techniques. Secondly, time series and cluster analysis were applied to major keywords. Trends for possible marine incidents were predicted. The results confirmed that it is possible to use quantified data and statistical analysis to address this topic. Also, we have confirmed that it is possible to provide information on preventive measures by grasping objective tendencies for marine incidents that may occur in the future through big data techniques.

**Key Words :** Marine incident, Quantitative analysis, Big data, Text mining, Time series & Cluster analysis

\* First Author : sykang53@seaman.or.kr, 051-620-5802

† Corresponding Author : bsro@seaman.or.kr, 051-620-5779

## 1. 서론

정부는 일자리 창출의 일환으로 공공데이터의 개방과 활용을 제시하고 있으며, 동시에 대국민과 기업의 안전 증진 및 편의 향상을 위하여 해운분야의 빅데이터 구축 및 분석의 중요성을 강조하고 있다.

빅데이터의 개념은 2001년 META Group(현 Gartner Inc., 이하 Gartner Inc.)의 보고서에서 처음 언급되었으며, 이에 대한 다양한 정의가 존재한다. Gartner Inc.의 애널리스트 더그레이니(Doug Laney)는 2001년 그의 연구보고서와 관련 강의에서 데이터의 급성장에 따른 이슈와 기회를 3V 즉, 데이터의 양(Volume), 데이터의 입출력의 속도(Velocity), 데이터 종류의 다양성(Variety)으로 정의한다(Doug, 2011). 또한 2012년 가트너 그룹은 ‘빅데이터는 큰 용량, 빠른 속도, 그리고 높은 다양성을 갖는 정보 자산으로서 이를 통해 의사결정 및 통찰발견, 프로세스 최적화를 향상시키기 위해서는 새로운 형태의 처리방식이 필요하다’라고 정의하였으며(The Gartner Group, 2012), Kim and Yun(2013)은 빅데이터의 가치를 광범위 대용량의 데이터 분석 및 트렌드 감지를 통한 미래예측 가능성, SNS 등을 통한 국민의 목소리 반영 및 선제적 대응력, 저장·처리·분석 기술의 발달로 다양한 단편적 정보의 활용 가능성 증대 등을 제시하고 있다.

빅데이터의 활용은 전 산업분야에서 활발하게 진행되어, 연구와 분석 및 실제 업무 적용 등에 대한 다양한 시도가 이루어지고 있으나 안타깝게도 해운 분야에서는 현재까지 빅데이터에 대한 활용 사례가 많지 않다. 하지만 빅데이터 활용에 대한 시대적 흐름을 반영하여, 향후 안전 분야를 중심으로 그 활용사례가 증가할 것으로 예상된다(Ryu et al., 2014).

본 연구에서는 준해양사고 데이터의 특징을 분석하고, 이를 고려하여 텍스트 마이닝이라는 빅데이터 분석기법을 적용하였다. 이 분석법을 활용하여 수집된 준해양사고 데이터를 분석하였고, 그 결과를 바탕으로 준해양사고 데이터를 정량적으로 분석할 수 있는 표준화된 모델링을 개발하였다. 또한 이를 적용하여 항해 중 준해양사고가 발생하는 원인을 분석하였다.

## 2. 빅데이터 분석 과정 및 기법

### 2.1 준해양사고 데이터

2010년 1월 국제해사기구(International Maritime Organization)가 채택한 국제해양사고조사코드(Code for the Investigation of Marine Casualties)가 발효되면서 해양사고 조사 및 예방을 위

한 국제적 지침이 제공되었고(Lim, 2010), 더불어 해양사고의 감소를 위해 계약국에게도 준해양사고제도의 관리를 권고하고 있다. 이에 따라 중앙해양안전심판원(이하 중해심)에서도 같은 해, 준해양사고제도를 신설하여 매년 해양에서 발생하는 사고를 저감하기 위해, 원인 규명을 통한 유사사고 방지에 만전을 기하고 있다.

해양사고의 조사 및 심판에 관한 법률 제2조제1의2호에 따르면 준해양사고란 선박의 구조·설비 또는 운용과 관련하여 시정 또는 개선되지 아니하면 선박과 사람의 안전 및 해양환경 등에 위해를 끼칠 수 있는 사태로서 해양수산부령으로 정하는 사고로 정의된다.

준해양사고 데이터는 현재 해운선사(이하 선사)에서 선박을 통하여 수집되며, 수집된 데이터는 자체적으로 분석하여 해양사고 예방 및 교육 등에 활용되고 있다. 중해심은 선사로부터 통보된 준해양사고를 재분석하여 이 중 선박이나 사람의 안전 및 해양환경 등에 위해를 끼칠 수 있는 내용을 공표하도록 하고 있다(NAIC, 2018).

하지만 기존의 중해심에서 공표된 내용은, 보고된 내용에 대하여 전문가 집단의 의견이 포함된 단순한 내용 검토로써, 교훈과 이에 따르는 예방 대책 제시가 전부였다. 통계 분석을 실시할 수 있는 정형화된 내용이 부족하여 중요 사항에 대한 객관성이 미흡하고, 실제 현장에 적용하기에는 한계성이 있었다. 이 때문에 준해양사고에 대한 통계분석을 포함하여 정량적인 데이터를 제공하지 못하고 있는 실정이다.

현재 중해심 홈페이지에는 해양사고에 관하여는 재검서를 포함하여 각종 통계자료가 제공되고 있는 반면, 준해양사고에 대하여는 공표자료와 이를 삽화집으로 제작한 ‘나최고 선장의 안전운항 필살기(현 노방심선장의 안전운항 레시피)’가 전부이다.

준해양사고 데이터의 특징은 날짜, 장소, 사고유형 등의 정형화된 내용은 적은 반면, 사고개요, 사고의 원인, 예방대책 등이 비정형화된 텍스트 형식으로 제공되고 있다. 이 때문에 기존의 통계분석을 이용하여서는 정량적 데이터를 제공하기 힘들고, 동시에 객관성을 확보하기가 어렵다. 반면 윌리엄 하인리히(William Heinrich)가 1931년 그의 저서 산업 재해 예방 과학적 접근(Industrial Accident Prevention: A Scientific Approach)에서 언급하였듯이 한 번의 대형사고가 발생하기 전에 그와 유사한 사고가 여러 번 일어나고, 그 사고 이전에 사소한 징후가 일어나므로(Kim, 2014a), 준해양사고 데이터는 선사에서 자발적인 통보만 잘 이루어진다면 해양사고에 비하여 발생건수가 많기 때문에 많은 양의 데이터를 수집할 수 있고, 빅데이터 분석기법을 적용하기에 적합하다고 사료된다.

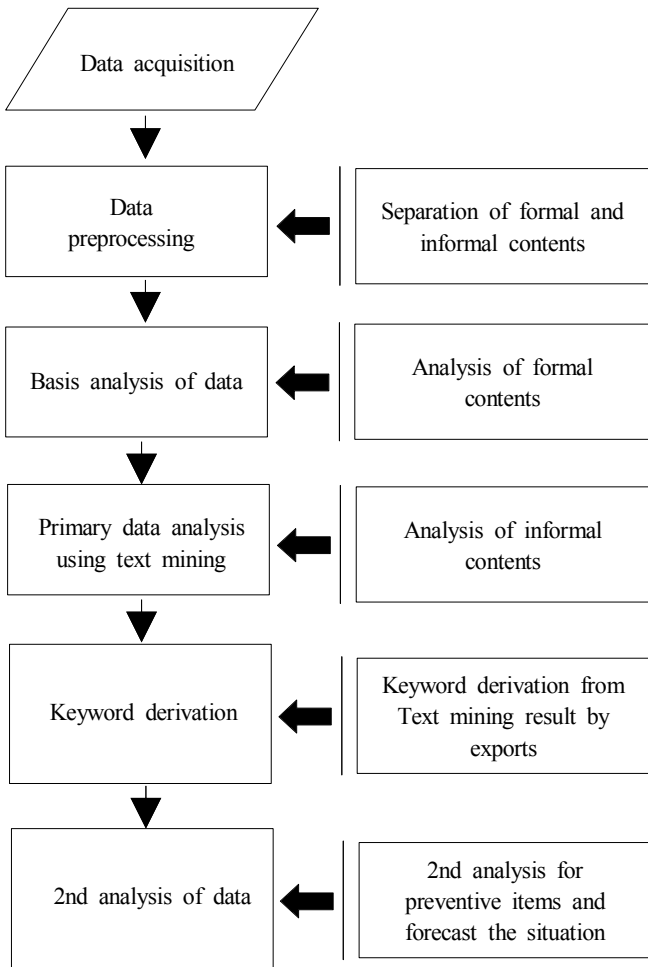


Fig. 1. Data analysis process.

## 2.2 분석 과정

준해양사고제도발전협의회에 소속되어 있는 회원에게 준해양사고 자료를 요청하였으며 그 중 10개 선사로부터 2015년에서 2017년 2분기까지의 준해양사고 데이터 10,000여건을 회신하여 Fig. 1과 같은 과정으로 분석하였다.

분석을 위하여 1차적으로 선행된 작업은 수집된 많은 양의 준해양사고 자료를 분석이 가능하도록 코딩하는 작업이다. 선사로부터 수집된 자료는 엑셀, PDF, 한글서식 등 다양한 형태의 문서로 작성되어 이를 전환하는 작업이 반드시 필요하다. 코딩작업을 통해 전체 자료를 내용은 변경하지 않고 Table 1과 같이 동일한 양식에 정리하였고, Table 2와 같이 선종, 사고종류 등의 정형화된 데이터와 일시, 사고원인 등의 비정형화된 데이터로 구분하였다. 본 연구에서는 비정형화된 데이터에 대하여 텍스트 마이닝이라는 기법을 활용하여 1차 분석하였다.

Table 2. Classification of the Marine incident

| Category  |
|---|
| 5 types of vessels :<br>Container ship, Tanker, Bulk carrier, Liquefied gas carrier, Car carrier                                      |
| 8 types of Marine incidents :<br>Fire, Collision, Standing, Engine Marine pollution, Explosion, Personal death or injuries, Equipment |
| Data and place of occurrence  |
| Detail of event, Cause of accident, Prevention measures   |

Table 1. Data preprocessing

| NO.   | Ship' Type       | Data (Local) | Place     | Marine Incident Type | Details of Event         | Cause of Incident      | Preventive Measures      |
|-------|------------------|--------------|-----------|----------------------|--------------------------|------------------------|--------------------------|
| 1     | Car Carrier      | 2015-10-28   | Anchoring | Equipment            | While on ship sailing... | Failure to follow rule | valve operation for...   |
| 2     | Container Ship   | 2017-01-19   | At sea    | Engine               | During monthly...        | Returned fireman's     | Returned fireman's...    |
| 3     | Bulk Carrier     | 2016-01-25   | At sea    | Fire                 | This ship was...         | Failure to use PPE     | C/E educated all...      |
| ⋮     | ⋮                | ⋮            | ⋮         | ⋮                    | ⋮                        | ⋮                      | ⋮                        |
| 9999  | Liq. Gas Carrier | 2015-12-26   | In port   | Standing             | Before arrival Dalian... | Using defective quip.  | Changed with new..       |
| 10000 | Liq. Gas Carrier | 2017-01-30   | At sea    | Explosion            | After navigation...      | Fire and explosion     | AB switched off...       |
| .     | Tanker           | 2017-03-24   | In port   | Equipment            | Around 1315 LT...        | Failure to follow      | Give education to all... |
| .     | .                | .            | .         | .                    | .                        | .                      | .                        |

1차 분석을 통해 도출된 키워드를 전문가 검증을 통해 의미 없는 데이터를 제거하여, 발생 빈도 상위 10건에 대하여 2차 분석을 진행하였다. 2차 분석은 시계열 및 클러스터 분석을 활용하였고, 특정 기간별 중요 키워드 파악, 단어간의 군집관계 파악 등을 통해 전문가가 사고예방을 위한 보조수단으로 활용할 수 있도록 하였다.

### 2.3 분석 방법

#### (1) 1차 분석-텍스트 마이닝

진술하였듯이 준해양사고 보고서내의 데이터 구조는 정형화된 데이터와 비정형된 데이터로 분류된다. 날짜, 시간, 장소, 위치 등은 정형화된 데이터이고, 사고 개요, 발생 원인, 예방 대책 등은 비정형화된 데이터이다. 이 중 사고예방에 대한 핵심 내용은 비정형화된 데이터이므로, 이에 대하여 빅데이터 분석 기법으로 텍스트 마이닝을 적용하였다.

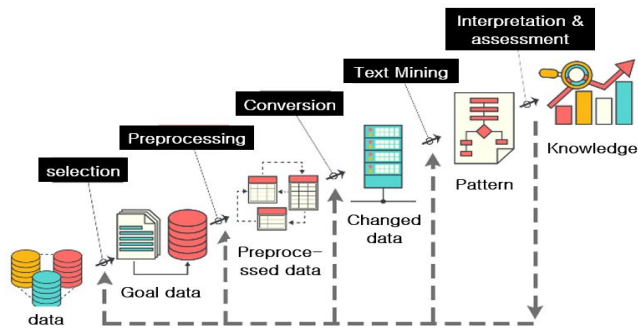


Fig. 2. Text mining process.

(Source: <https://insightcampus.co.kr/shop/pre-processing>)

Fig. 2와 같이 텍스트 마이닝은 비-반정형 텍스트 데이터에서 자연어처리 기술에 기반하여 유용한 정보를 추출 및 가공하는 것을 목적으로 하는 기술이다. 방대한 텍스트 문치에서 의미 있는 정보를 추출해 내고 다른 정보와의 연계성을 파악하여 텍스트가 가진 카테고리를 찾거나 단순한 정보 검색 이상의 결과를 얻어낼 수 있다. 컴퓨터가 텍스트 내용을 분석하고 그 안에 숨겨진 정보를 발굴해 내기 위해 대용량 언어 자원과 통계적, 규칙적 알고리즘이 사용된다(Jo, 2011).

#### (2) 2차 분석-시계열 및 클러스터 분석

텍스트 마이닝을 통해 분석한 결과는 단순히 문장속에 담겨진 중요 키워드 도출이다. 이 결과에 대해 준해양사고의 향후 경향과 예측 분석을 위해서는 2차 분석 작업이 필요하다. 2차 분석을 위한 방법으로는 시계열 및 클러스터(군집) 분석 기법을 적용하였다.

시계열 분석이란 동일한 시간 간격으로 측정된 과거의 주요 값들이 존재할 경우, 시간의 흐름에 따라 나타난 시계열의 패턴을 파악하여 그 패턴이 미래에도 계속 적용된다는 가정하에 그 상황을 예측하는 방법이다(Kim, 2014b).

클러스터 분석은 비슷한 특성을 가진 개체를 합쳐가면서 최종적으로 유사 특성의 그룹을 발견하는 방법이다(Jo, 2011).

Fig. 3은 역대 미국 대통령의 취임 연설문의 단어를 텍스트 마이닝 이후에 클러스터 분석을 실시한 것이다. 유사한 의미의 단어들을 군집 분석하여 정책의 방향성 등을 예측할 수 있다.

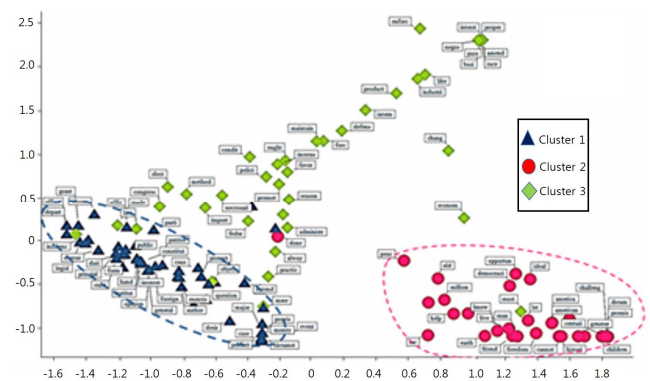


Fig. 3. Cluster analysis diagram of us presidential inaugural addresses.

(Source: Cho et al., 2015)

## 3. 분석 결과 및 고찰

### 3.1 1차 분석 결과

Fig. 4는 2015년부터 2017년 6월까지 분기별로 항해 중 준해양사고가 발생한 원인을 텍스트 마이닝을 적용하여 분석한 결과이다(Rho et al., 2017).

결과에서 확인할 수 있듯이 각 분기별로 가장 빈도가 높은 단어가 크게 표시되며, 빈도가 낮을수록 단어는 작게 표시된다. 하지만 이는 단순히 빈도별 단어 나열에 불과하여 바로 적용은 어렵고, 전문가 집단의 검증작업이 추가되어야 한다. 즉 실제 항해 중 준해양사고 원인과 관계없는 조사나 형용사 등도 포함되어 있으므로 이는 제거되어야 한다. 예를 들어 2017년 2분기에서는 가장 빈도가 높은 단어는 'Inadequate'이나 이 단어는 형용사로 사고원인과는 무관한 단어이므로 이런 단어는 제외하였고, 모든 과정에 대하여 같은 방법을 적용하였다.

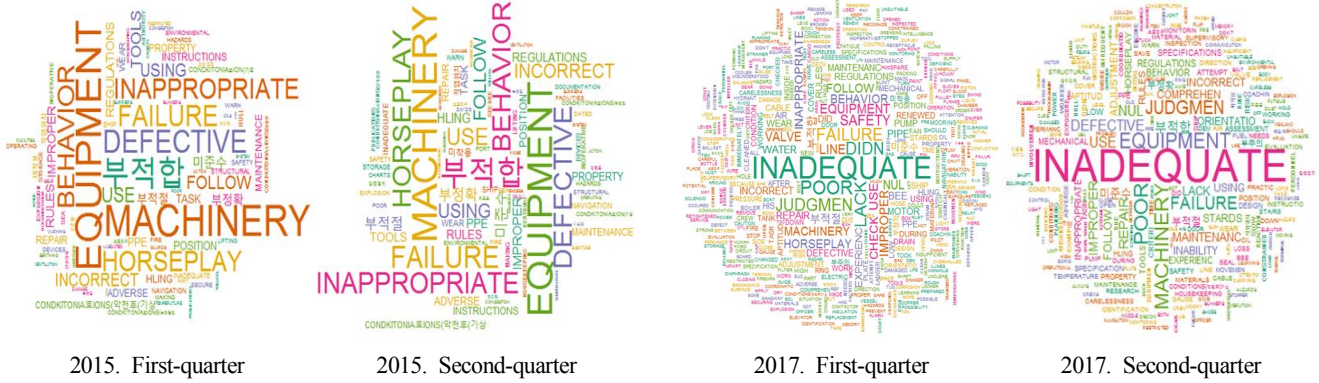
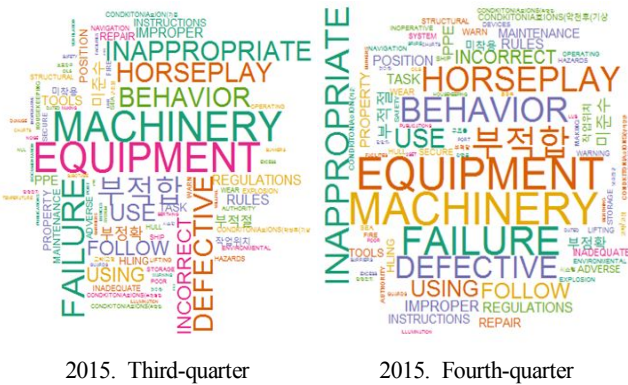
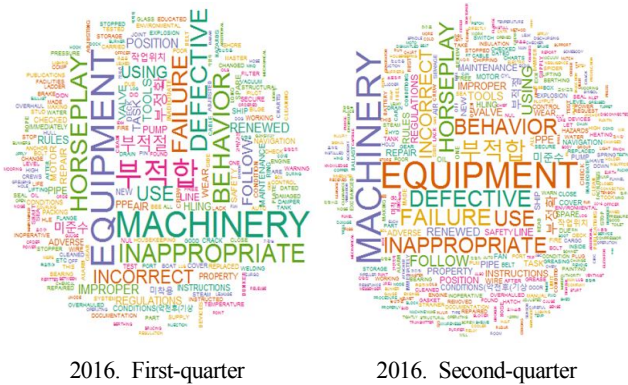


Fig. 4. Text mining on cause of Marine incident at sea (2015 ~ 2017 2<sup>th</sup>).



3.2 2차 분석 결과

1차 텍스트 마이닝과 전문가에 의해 최종 선정된 상위 빈도 10개의 단어를 Fig. 5와 같이 통계 처리하였다. 색상의 차이는 발생률에 따른 것으로 색상이 진해질수록 발생률이 적어지는 형태의 히트맵 그래프이다.



|            | 15-01 | 15-02 | 15-03 | 15-04 | 16-01 | 16-02 | 16-03 | 16-04 | 17-01 | 17-02 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| BEHAVIOR   | 16%   | 17%   | 16%   | 19%   | 17%   | 16%   | 16%   | 16%   | 31%   | 0%    |
| DEFECTIVE  | 18%   | 18%   | 17%   | 17%   | 16%   | 17%   | 17%   | 12%   | 0%    | 31%   |
| EQUIPMENT  | 26%   | 25%   | 26%   | 26%   | 25%   | 28%   | 26%   | 23%   | 38%   | 46%   |
| HORSEPLAY  | 16%   | 17%   | 16%   | 19%   | 17%   | 16%   | 16%   | 16%   | 31%   | 0%    |
| MAINTENANC | 0%    | 0%    | 0%    | 0%    | 0%    | 0%    | 0%    | 0%    | 0%    | 23%   |
| POSITION   | 5%    | 5%    | 5%    | 6%    | 7%    | 0%    | 0%    | 7%    | 0%    | 0%    |
| PPE        | 5%    | 6%    | 6%    | 7%    | 0%    | 5%    | 8%    | 9%    | 0%    | 0%    |
| RENEWED    | 0%    | 0%    | 0%    | 0%    | 7%    | 6%    | 8%    | 9%    | 0%    | 0%    |
| RULES      | 6%    | 6%    | 7%    | 6%    | 5%    | 5%    | 10%   | 7%    | 0%    | 0%    |
| TOOLS      | 8%    | 6%    | 6%    | 0%    | 7%    | 6%    | 0%    | 0%    | 0%    | 0%    |

Fig. 5. Quarterly main keyword rate on cause of Marine incident at sea (2015 ~ 2017 2<sup>th</sup>).

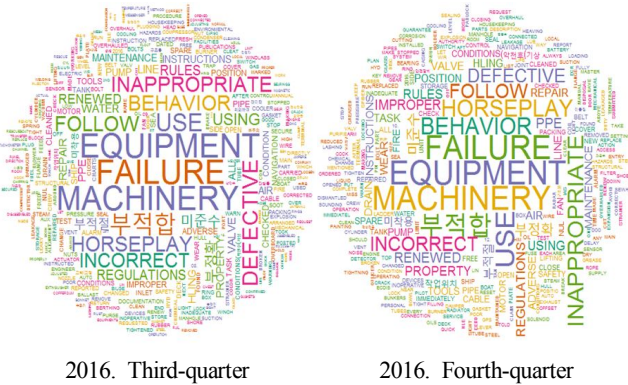


Table 3은 최종 선정된 상위 키워드 10개에 대하여 전문가들이 의미를 부여한 표이다. 여기서도 관련분야를 잘 이해하는 전문가들의 수작업이 추가되어야 한다. 예를 들어 도출된 키워드 중 'DEFECTIVE'라는 단어가 나왔을 때 항행 중 준해양사고 발생원인이 선박의 각종 장비 결함에 의한 고장으로 기인할 수 있다거나 'RENEWED'라는 단어는 기부속, 선용품, 유류 교체 과정에서 준해양사고가 발생할 수 있다는 의미부여는 전문가의 몫이다. 이렇게 부여된 의미를 바탕으로 예방책을 강구하여 해양사고 예방에 이용한다.

빅데이터 기법을 활용한 항해 중 준해양사고 발생원인 분석에 관한 연구

Table 3. Selected keyword analysis

|             |   |
|-------------|---|
| BEHAVIOR    | Marine incident by activities of crew at sea  |
| DEFECTIVE   | Marine incident by defective of ship equipment  |
| EQUIPMENT   | Marine incident by ship's equip. or mach.   |
| HORSEPLAY   | Marine incident by carelessness of crew at sea  |
| MAINTENANCE | Marine incident occurred during the process of maintenance & repair of ship's mach. & equip.                |
| POSITION    | Marine incident due th missed missions at position of port entry & departure or work                        |
| PPE         | Marine incident related to personal protective equipment  |
| RENEWED     | Marine incident occurred during the process of renewing spare parts, articles for ship, oil & L.O           |
| RULES       | Marine incident occurred by non-compliance with company regulations, international and domestic regulations |
| TOOLS       | Marine incident occurred during the process using tools, implement, equipment,                              |

마지막으로 선정된 10개 키워드에 대해서 Fig. 6과 같이 시계열 및 클러스터 분석을 실시하였다.

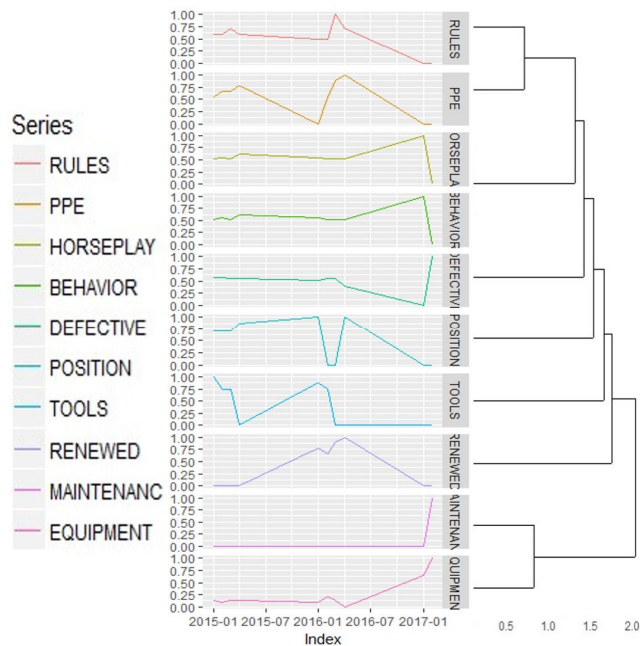


Fig. 6. Time series analysis & cluster analysis on keyword that cause of Marine incident at sea.

분석을 통해 도출된 결과를 바탕으로 항해 중 해양사고에 대한 예방책을 강구하면 다음과 같이 정리할 수 있다.

(1) EQUIPMENT 키워드의 경우 2015년~2016년 사이에 비슷한 분포율을 보이다가 2016년 후반 및 2017년 초반 들어 증가세에 있으므로 2017년도 이후로 항해 중 해양사고 예방을 위하여 장비의 추가적인 정비나 확인이 필요하다.

(2) POSITION, PPE, RULES 키워드의 경우 2015년~2016년에는 비슷한 경향을 보이다가 2017년에 들어 줄어드는 추세를 보이고 있으므로 관련 원인으로 인한 해양사고는 줄어드는 것임을 예측 가능하다.

(3) TOOLS 키워드의 경우 2016년 초반까지 증가세를 보이다 이후 급감하면서 2016년 중반 이후부터는 거의 나타나지 않으므로 향후 비슷한 경향을 보일 수 있다고 예측 가능하다.

(4) RENEWED 키워드의 경우 2015년은 거의 없다가 2016년 상승했다가 2017년에는 다시 줄어드는 경향을 보이므로 2016년에 어떤 사유에 의해 발생했는지 추적해보면 예방 방법의 도출이 가능하다.

(5) BEHAVIOR, DEFECTIVE, MAINTENANCE, HORSEPLAY 키워드의 경우 2015년~2016년까지 비슷한 분포율을 보이다가 2017년에 들어 급증과 급감을 보이고 있으므로 추가적인 분석이 필요하다.

키워드간의 상관관계를 나타내는 클러스터 분석은 분석 프로그램의 특성상 각 키워드에 대해 끝까지 병합하는 과정이 발생하는데 이 중 2단계까지의 관계를 의미하는 것으로 평가하며, 이에 따라 결과를 분석하면 다음과 같다.

(1) 1단계까지의 분석 예를 보면 RULLS이라는 키워드가 나오면 PPE라는 키워드가 많이 나오고 반대의 경우도 그러하므로 키워드간의 관계가 형성되는 것을 알 수 있으며, 이를 통해 예측 분석을 해보면 규정에 의한 개인 보호 장구와 관련하여 준해양사고가 발생할 수 있음을 보여준다. 같은 개념으로 MAINTENANCE와 EQUIPMENT라는 키워드도 관계가 형성되는 것을 알 수 있으며 이를 통해 기기나 장비의 유지·보수 과정에서 준해양사고가 발생할 수 있음을 알 수 있다.

(2) 특히 RULL과 PPE 키워드의 경우, 1단계 관계 형성 이후에 HORSEPLAY라는 키워드와 2단계 관계 형성을 하게 된다.

(3) 상기 두 가지 조합은 1단계와 2단계 과정을 통해 규정에 의한 개인 보호 장구 착용 과정에서 주의력 부족으로 인한 준해양사고와 연관되어 발생할 수 있음을 보여준다.

#### 4. 결 론

이번 연구를 통해 비정형화 및 텍스트 위주로 되어 있는 준해양사고 데이터를 활용하여 기존과 같이 단순한 전문가들의 의견이 반영된 교훈도출을 넘어선 정량적인 값을 도출하고자 노력하였다. 이를 위하여 현재 전 산업 분야에서 많이 이용되는 빅데이터 분석기법을 활용하였고, 항해 중 준해양사고의 발생원인에 대하여 분석해 보았다.

결과적으로 준해양사고 자료에 대해 빅데이터 분석기술을 활용하여 정량화된 데이터로 전환할 수 있고, 이를 통해 통계적 분석이 가능함을 확인할 수 있었다. 나아가 차 후 발생할 수 있는 준해양사고의 객관적인 경향을 파악함으로써 예방 대책에 대한 정보 제공이 가능함을 확인할 수 있었다. 더불어 준해양사고 데이터를 대상으로 새로운 분석기술을 적용한 시도가 처음이라는 점과 이를 통해 일부 사고 예방에 대한 결과를 도출한 점은 의미 있는 것으로 판단된다.

아쉬운 점은 연구에 사용된 준해양사고 데이터가 2년 6개월의 짧은 기간에 10개 선사로부터 수집된 한정된 데이터라는 점과 이로 인한 데이터의 객관성에 대한 한계이다. 또한 빅데이터라 부르기에 데이터의 양이 충분하지 못했고, 현재의 빅데이터 분석기술의 제약으로 분석의 과정 사이사이에 관련내용을 잘 이해하는 전문가 집단의 수동적인 개입이 불가피하다는 점이다. 이러한 부족한 부분은 장기간의 걸쳐 빅데이터를 구성하고, 분석과정 중에 나타나는 오류와 사람의 개입에 대한 부분을 축적된 데이터를 바탕으로 자동화한다면 충분히 개선 가능하리라 사료된다.

마지막으로 이번 연구는 항해 중 준해양사고 발생원인으로 연구의 범위를 제한하였으나 좀 더 많은 자료를 수집하여 정박중, 묘박중, 사건 개요, 예방 대책 등 다양한 관점에서 빅데이터 분석 기법을 적용할 필요가 있으며, 더불어 분석된 준해양사고와 실제 해양사고와의 관계를 비교 검증할 수 있다면 더욱 의미 있는 연구가 되리라 사료된다.

#### 후 기

본 논문은 준해양사고제도의 실효성 확보방안 마련 연구용역(2017년)에 의해 연구되었음.

#### References

[1] Cho, S. G., J. H. Cho and S. B. Kim(2015), Discovering Meaningful Trends in the Inaugural Addresses of United States Presidents Via Text Mining, Journal of the Korean

Institute of Industrial Engineers, Vol. 41, No. 5, pp. 453-460.  
 [2] Doug, L.(2001), 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group Application Delivery Strategies, File. 949, pp. 1-3.  
 [3] Jo, S. W.(2011), Technology of big data age, KT Advanced Institute of Technology, pp. 4-5.  
 [4] Kim, D. J. and S. Y. Yun(2013), Big Data Utilization for Monitoring Territorial Policy Responses and Predicting Policy Demand, Korea Research Institute for Human Settlements, pp. 28-29.  
 [5] Kim, M. J.(2014a), Heinrich's Law, Miraebok Publishing Co, p. 16.  
 [6] Kim, M. G.(2014b), A Study on Analysis and Prediction of Demand of Electric cars using Time-series Analysis, Graduate School Kumoh National Institute of Technology, p. 23.  
 [7] Lim, C. H.(2010), A Study on the Introduction of IMO Casualty Investigation Code and Marine Safety Investigation System in Korea, The Journal of Korean Society on Marine Environment & Safety, Vol. 16, No. 1, pp. 57-63.  
 [8] NAIC(2018), Act on the Investigation and Inquiry into Marine Accidents, Article 31-2.  
 [9] Rho, B. S., M. K. Jang, S. W. Jeon and S. Y. Kang(2017), A study on Development of Near Accident Big Data Construction and Analysis Method, The Journal of the Korean Society of Marine Environment & Safety, Presented at 2017 Autumn Conference of KOSOMES, p. 122.  
 [10] Ryu, J. G., T. I. Kim and T. Y. Ha(2014), Ocean & Fisheries for big data utilization plan, Korea Maritime Institute, pp. 37-38.  
 [11] The Gartner Group(2012), Gartner's 2012 Hype Cycle for Emerging Technologies, <https://www.gartner.com/newsroom/id/2124315>.

Received : 2018. 04. 18.

Revised : 2018. 06. 13.

Accepted : 2018. 06. 27.