

해양사고 온톨로지 구축 및 데이터 관리방안 연구: 서해남부해역 선박사고 상황보고서 분석을 중심으로

이영재* · 강성경** · 구자영***

*, ** 동국대학교 경영정보학과, *** 중부지방해양경찰청

A Study on Marine Accident Ontology Development and Data Management: Based on a Situation Report Analysis of Southwest Coast Marine Accidents in Korea

Young Jai Lee* · Seong Kyung Kang** · Ja-Yeong Gu***

*, ** Department of MIS, Dongguk University, Seoul 04620, Korea

*** Commander of Central Regional Coast Guard, Central Regional Coast Guard, Incheon 22006, Korea

요약 : 매년 해양활동이 증가하며 해양사고 발생빈도가 높아지고 있다. 이에 따라 해양안전을 위한 각종 연구 활동과 정책이 실행되고 있다. 하지만 이러한 노력에도 불구하고 매년 해양사고가 증가하고 있어 이들의 실효성에 대한 문제가 제기되고 있다. 문헌연구 결과에 따르면, 통계연보를 활용한 선행연구는 통계제공항목 간 비교를 통해 두드러지는 항목에 대한 예방책을 제시하고 있다. 2000년대 이후에는 대형 해양사고가 반복적으로 발생하면서 '사고대응'에 대한 사례연구(case study)가 진행되고 있다. 국내 해양안전을 위한 정책수립 연구과정에서 통계연보나 사고사례를 주 자료로 활용하고 있으나 현재 자료는 사후결과 요약정도의 수준이다. 따라서 본 연구에서는 해양사고 사례분석 및 개선방안 관련 문헌연구를 통해 현행 해양관련 연구와 정책의 한계를 탐색하였다. 또한 자료 활용 한계를 개선하기 위한 방안의 일환으로 선박사고 상황보고서 속성분석, 텍스트 마이닝을 통해 해양사고 정보 분류체계인 온톨로지(ontology)를 수정·보완하였다. 해당 항목은 '신고자, 신고수단, 구조세력, 대응·조치사항, 대응취약성, 적재물, 유류유출경위, 피해유형, 사고처리결과'이며, 이 항목들은 분류체계 표준용어를 활용해 향후 지속적으로 수집·활용할 수 있다. 마지막으로 온톨로지를 실질적으로 활용하기 위한 데이터 수집 및 품질확보 방안을 제시했다. 결과적으로 현재 해양안전이 직면한 문제를 명확히 파악하고 '품질이 확보된 충분한 정보'를 활용한다면 보다 다양한 연구와 실효성 있는 정책 실현이 가능할 것이다.

핵심용어 : 해양사고, 사고데이터, 온톨로지, 데이터 품질, 정책수립

Abstract : Along with an increase in marine activities every year, the frequency of marine accidents is on the rise. Accordingly, various research activities and policies for marine safety are being implemented. Despite these efforts, the number of accidents are increasing every year, bringing their effectiveness into question. Preliminary studies relying on annual statistical reports provide precautionary measures for items that stand out significantly, through the comparison of statistical provision items. Since the 2000s, large-scale marine accidents have repeatedly occurred, and case studies have examined the "accident response." Likewise, annual statistics or accident cases are used as core data in policy formulation for domestic maritime safety. However, they are just a summary of post-accident results. In this study, limitations of current marine research and policy are evaluated through a literature review of case studies and analyses of marine accidents. In addition, the ontology of the marine accident information classification system will be revised to improve the current limited usage of the information through an attribute analysis of boating accident status reports and text mining. These aspects consist of the reporter, the report method, the rescue organization, corrective measures, vulnerability of response, payloads, cause of oil spill, damage pattern, and the result of an accident response. These can be used consistently in the future as classified standard terms to collect and utilize information more efficiently. Moreover, the research proposes a data collection and quality assurance method for the practical use of ontology. A clear understanding of the problems presently faced in marine safety will allow "sufficient quality information" to be leveraged for the purpose of conducting various researches and realizing effective policies.

Key Words : Marine accidents, Accident data, Ontology, Data quality, Policy-making

* First Author : yjlee@dongguk.edu, 02-2260-3297

† Corresponding Author : kjy1136@korea.kr, 032-728-8010

1. 서론

삼면이 바다인 우리나라는 높은 어획강도, 활발한 해상교역과 여객수송, 각종 해양·항만시설 등으로 복잡한 해상교통 환경을 가졌다. 2017년 기준 우리나라 전체 선박 등록 수는 95,309척이며 이 중 어선 등록이 66,736척으로 70%를 넘는 수준이며, 레저선박이 19,494척(약 20.4%), 일반선이 9,079척(약 9.5%) 정도로 집계되고 있다(Korean Maritime Safety Tribunal, 2018).

2017년 해양 선박사고 발생현황을 살펴보면 어선·낚시어선이 63.98%, 레저선박 21.45%, 예부선 3.73%, 화물선 3.63%, 여객선 1.01%, 유조선 1.89%, 유도선 0.63%, 기타 3.63%로 나타났으며, 사고의 절반 이상이 어선에서 발생했다. 사고유형은 기관손상이 29.27%로 가장 높은 비중을 차지하며, 충돌(13.48%), 부유물 감김(12.62%) 순으로 사고 발생 빈도가 높다. 사고 발생원인은 정비 불량 43.23%, 운항부주의가 32.97%로 가장 많으며, 특히 정비 불량은 2013년 24.42%에서 2017년 43.23%까지 늘어 사고 주요 원인으로 지적되고 있다.

선박 등록 수 및 이용객이 매년 증가하는 만큼 해양사고의 위험성도 높아졌다. 이에 따라 정부에서는 과거 주요 해양사고 발생사례를 기반으로 시대적 상황에 기인한 법·제·개정 및 안전대책 등을 수립·시행 중에 있다. 1960년대 이후 해양안전 관리 필요성이 대두되면서 해사관련 법·제도 등이 체계적으로 제정되기 시작했으며, 2014년 세월호 사고 이후 기존의 법·제도, 조직, 시스템 등 해상안전관리체계가 전면적으로 재정비 되었다(Korea Maritime Institute, 2017).

문제는 정부의 지속적인 정책정비, 안전대책 추진에도 불구하고 해양사고 예방의 한계점이 드러나고 있다는 것이다. 우이산호(‘14), 세월호(‘14), 오룡호(‘14), 돌고래호(‘15), 스텔라데이지호(‘17) 등 계속되는 선박사고로 해양안전에 대한 국민의 불안감이 높아지고 있다. 이는 곧 반복적으로 발생하는 해양사고에 대한 객관화, 고도화, 전문화된 사고원인 규명과 분석을 통해 사고 재발방지 및 사전예방을 위한 실효성 있는 대책을 실시해야 함을 시사한다.

물론 해상안전정책 정비나 체계적인 예방대책을 실시한다고 해서 모든 사고가 미연에 방지되는 것은 아니다. 사고라는 것은 불확실성과 복잡성이라는 특성을 가지기 때문에 실제로 언제 발생할지, 어떤 원인으로 발생할지 명확하게 예측하기 어렵다. 특히 해상에서 발생하는 사고의 경우 해조류, 시정, 기상 등 환경의 영향을 많이 받고, 육상에 비해 현장접근수단이나 구조능력, 사고통제와 관련한 제약요소가 존재하기 때문에 사고가 발생하면 육지 사고에 비해 구조측면에서 많은 어려움이 따른다(Roh, 2014).

가장 중요한 것은 반복적으로 발생하는 해양사고에 대비·

대응하기 위해 현재 직면한 문제와 이를 해결하기 위해 필요한 정보를 파악하여 실효성 있는 정책을 수립하고 적용하는 것이다. 과거부터 현재까지 해양안전을 위한 사고분석과 정책연구가 진행되어 왔음에도 불구하고 사고는 매년 증가하고 있다. 대부분의 해양안전 연구나 정책수립은 해양사고 경향을 파악할 수 있는 통계연보나 사례집을 활용하고 있고, 결과적으로 유사한 예방·대응 정책이나 개선방안을 제시하고 있다. 이는 우리가 사고예방과 대응을 위해 정책을 수립하는 과정에서 제한적인 자료를 활용하고 있음을 보여주는 대목이기도 하다.

해양사고의 발생이 해양안전 정책변화에 많은 영향을 미치는 것은 분명하다. 하지만 현재 정책수립이나 연구에서 활용하고 있는 자료는 문제를 해결하기 위해 ‘필요한 정보’지만 ‘충분한 정보’는 아니다. 또한 전문적인 사고 조사·분석 시스템이 부재하고, 해양안전 데이터 품질관리를 위한 표준 관리 프로세스도 없기 때문에 실제로 정책수립과 연구에 활용되는 데이터의 품질이 확보되었다고 할 수 없다.

데이터의 품질은 곧 분석결과의 질을 결정하고, 분석결과는 의사결정, 곧 정책수립을 위한 핵심요소라고 해도 과언이 아니다. 특히나 해양사고의 경우 사고 발생 후 초동대응의 적절성과 신속성에 따라 인명, 재산, 환경오염 피해규모가 크게 달라지기 때문에 정확한 데이터에 근거한 의사결정이 중요하다. 육지와 달리 상황 변동성이 큰 해상이라는 특수성을 고려할 때 예방정책 또한 품질이 확보된 데이터를 기반으로 수립하고 시행해야 한다.

결국 해양안전을 위한 실효성 있는 정책수립과 개선책 모색을 위해서는 직면한 문제에 대한 명확한 이해를 토대로 ‘신뢰성이 확보된 충분한 정보’를 활용해야 한다. 지금까지 진행되어온 선행연구에서 제시한 개선책이나 정부 정책에 대한 효과는 이미 매년 ‘증가하고 있는 사고’가 보여주고 있다. 따라서 현재 활용하고 있는 자료의 신뢰성과 충분성에 대해 현 실태를 검토해볼 필요가 있다.

본 연구에서는 해양사고 사례분석 및 개선방안 관련 문헌 연구를 통해 현행 해양관련 연구와 정책의 한계를 모색한다. 또한 이를 개선하기 위한 방안의 일환으로 선박사고 상황보고서 속성분석 및 텍스트 마이닝을 통해 해양사고 정보 분류체계인 온톨로지(ontology)를 재정립한다. 이 과정은 비구조적인 텍스트를 구조화하는 것으로 상황보고서 상의 텍스트를 추출하여 데이터베이스화(DB)한 후 수차례의 전처리를 통해 표준용어 분류체계를 제시하는 것이다.

해양사고 온톨로지 정립을 위해 최근 5년 간 사고 발생이 가장 많았던 서해남부 해역 관할 5개관서(군산, 부안, 목포, 여수, 완도)에서 작성한 ‘15~‘17년 선박사고 상황보고서 6,300여건을 활용하였다.

2. 문헌연구

2.1 해양사고 안전관리 선행연구 동향분석

우리나라는 1970년 이후 해양사고 통계를 집계하기 시작했다. 1970년대 중대사고 건수는 1,533건에서 2000년대 7,942건으로 대폭 상승했다. 대형 해양사고 증가에 따라 시대별로 각종 안전정책 시행되고, 해양안전 연구에 있어서는 사고 원인분석을 통한 예방대책을 제시하는 연구가 주를 이루고 있다(Kim et al., 2017). 분석 자료는 대부분 해양경찰청 통계연보 또는 해양안전심판원의 통계자료, 관련 사례집 등을 활용하고 있다.

Lee et al.(2011)는 해양안전심판원의 해양사고 통계자료를 활용하여 1990년~2009년까지 국내에서 발생한 해양사고 발생 환경과 원인을 조건별로 조사하고 분석하였다. 그 결과 100톤 미만 소형선박의 사고 발생비율이 매우 높게 나타나 이를 해결하기 위한 방안으로 이동통신 단말기를 활용하는 방안을 제시하였다. 국내 해양사고의 대부분이 소형선박에서 발생함에 근거하여 첨단 항해장비 도입을 고려할 경우 공간부족, 설치비용, 장비운영 등 현실적 어려움이 존재하기 때문에 이동통신 단말기인 휴대폰을 활용하여 효과적으로 사고를 예방할 수 있을 것으로 판단했다.

Kang et al.(2013)는 국내에서 발생한 해양사고 중 가장 높은 비율을 차지하는 충돌사고와 기관손상 사고에 대한 원인과 해결방안을 모색하기 위해 어선과 비 어선으로 구분하여 분석을 실시했다. 분석대상 자료는 국토해양부, 농림수산식품부, 중앙해양안전심판원 사례집이며, 2001년부터 2010년 사이 발생한 충돌, 기관손상 사고가 선종, 원인, 인명피해와 어떤 결과를 나타내는지를 설명하고 있다. 결과적으로 우리나라 어선 등록 척수가 전체 선박의 90%를 차지하고 있기 때문에 사고 비중이 높으나 척당 해양사고 발생률은 비어선이 어선보다 5배 높고, 원인에 있어서는 선종에 관계없이 모두 운항과실, 조작 및 작업행위 부적절 등이 압도적으로 높아 인적오류에 대한 교육과 의식개선이 필요함을 강조했다.

Cho et al.(2017)는 어선사고의 다발은 우리나라 해양사고 발생 경향을 좌우하기 때문에 어선에 대한 사고 저감 방안 수립과 시행이 시급하다고 언급했다. 해당 연구에서는 연간 발생하는 사고의 편차를 고려하여 1996년~2015년까지 20년간의 선박사고 자료를 토대로 우리나라 연근해 운항선박의 전체적인 사고 방지 대책을 모색하였다. 지난 20년간 발생한 해양사고를 선종, 사고종류, 사고원인에 따라 발생률을 비교하였으며, 해양사고자의 인적 특성을 고려하여 연령, 승선경력에 따라 사고 빈도를 비교하였다. 또한 사고 당시의 시간대, 기상, 속력, 초인거리 등 어떤 상황에서 사고가 많이 발생하는지 특성을 구분해서 보여주고 있다. 해당 연구는 해양사고 통계자료를 토대로 연도별 빈도분석이 주를 이루고

있으며, 기관손상 사고가 많음에 대비해 철저한 정비가 필요하고, 사고 원인이 모든 선박에 있어 인적요소 부주의에 의해 발생하는 비중이 높기 때문에 운항자에 대한 경계강화 및 교육철저가 필요하다고 저감대책을 제안하였다.

Kim(2018)은 해무로 인한 저시정 시 충돌사고를 예방하기 위한 해양경찰의 역할을 제언하였다. 지난 2017년 해무로 인한 저시정은 150회 이상으로 최대치를 기록했고, 저시정 상황에서 충돌사고도 지속적으로 증가하는 추세이다. 2016년 대비 2017년 저시정 충돌사고는 52% 증가하였으며 그 원인은 인적과실에 의한 운항부주의가 80%에 달한다고 밝혔다. 해상에서의 안개에 대한 레이더 등 첨단 항해 장비를 이용하더라도 원거리 물체를 탐지할 수 있는 기술이 미흡한 현 실정에서 아무리 노련한 항해사라도 시정문제는 항해의 큰 장애요인이 되고 있다. 특히 소형어선, 레저선박 등 소형선박의 경우 발견이 더욱 어렵기 때문에 출항을 적극 통제하고 준법계도를 강화해야 하며, 경비함정은 관제구역 밖 항해 선박에 대해 안전방송을 적극 실시하여 충돌사고를 예방토록 해야 한다고 제언했다.

앞선 연구들의 공통점은 모두 통계연보나 사례집에 기반하여 유사한 예방정책(대안)을 제시하고 있다는 것이다. 대부분의 해양사고 분석 연구에서 활용되는 통계자료는 수치의 편차는 있으나, 분석결과를 살펴보면 대부분 사고가 다발하는 선종, 원인, 시간대, 기상, 피해규모 등이 거의 유사하게 귀결된다. 통계연보나 자료에서 제공되는 데이터 수준에서만 분석을 진행하고 정책 방향을 제시하는 경우 대부분 항목 간 비교를 통해 두드러지게 부각되는 부분을 개선해야 할 사항으로 인식하기 때문이다.

물론 모든 연구가 통계자료를 토대로 예방에 관한 고찰만 진행하고 있는 것은 아니다. 2000년대 이후 대규모 해양사고가 반복적으로 발생하면서 사고대응에 대한 국민적 관심과 불안감이 고조되고 있는 시점에서 해양사고 대응에 대한 개선방안 연구도 활발히 진행되고 있다.

Roh(2014)의 연구에서는 해양사고 통계와 주요 여객선 사고사례 검토를 통해 사고 원인을 분석하고 해양경찰의 역할을 예방과 구조측면에서 제시했다. 특히 해양사고를 대응하는데 있어 해양경찰의 해양사고대응시간(Time of Arrival of Coast Guard)에 대해 설명하며 사고초기의 신속성과 대응역량이 얼마나 중요한지를 언급했다. 한 예로 침몰이 시작되고 피구조자가 신고하는 시간, 상황실에 접수가 되고 해경이 현장에 출동하는 시간까지를 해양사고대응시간으로 표현한다면, 더불어 구조선박이 사고 현장에 도착하기까지 선박의 침몰시간을 지연시키는 것, 구명정탐승, 구명동의 착용, 해경 함정의 현장출동시간을 줄이는 것이 초기대응에 있어 매우 중요하다. 해양사고에서 선박이 물에 잠겼을 경우 사람이 생존 가능한 시간은 5~10분이며, 선박에서 퇴선 후 구명조끼에 의지한 경우 10℃ 해수에서 2~3시간 생존

할 수 있다. 구조인력의 신속한 출동이 중요한 이유이다.

하지만 해상의 경우 육지에 비해 구조 접근이 쉽지 않고, 장애요소가 많기 때문에 해결세력만으로는 모든 사고를 대응하기 어렵다. 자원이 한정적일 뿐만 아니라 사고 해상 환경변화 또한 예상하기 어렵기 때문이다. 특히 대형사고가 발생할 때마다 대응역량에 대한 문제점이 지속적으로 제기되면서 해양사고대응에 대한 민관 협업 거버넌스 필요성도 대두되고 있다(Ju, 2016).

위의 선행연구처럼 해양사고 예방과 관련한 연구에 비해 대응체계에 대한 연구는 통계자료 등의 계량자료를 주 분석자료로 활용하기 보다는 다양한 사례연구(case study)를 통해 대두된 문제점을 공통적으로 도출하고 개선할 사항들을 제안하는 특성을 보인다.

2.2 해양사고 분류체계 관련 문헌연구

국내에서 선행되고 있는 해양사고 안전관리를 위한 대다수의 선행연구들은 해양경찰청, 해양안전심판원 등에서 제공되는 통계자료를 활용하고 있으며, 이 통계자료는 사고경향을 파악하기 위한 분류체계에 따라 매년 공개되고 있다.

분류는 대상을 유사성에 따라 종류별로 나누거나 배열하는 것이며, 해양경찰청에서 발간하는 「해상조난사고 통계연보」나 해양수산부 해양안전심판원에서 제공하는 「해양사고통계」의 분류체계는 통계조사대상의 차이로 완전 동일하게 구성되어 있지는 않으나 사고현황을 파악하기 위한 발생건수, 선박톤수, 시간대, 피해유형, 사고원인 등 일부 분류체계와 세부분류는 유사하게 구성되어 있다.

문제는 현재 제공되고 있는 통계 분류항목으로는 정책수립 활용에 있어 한계가 있다는 것이다. 1970년대 이후 해양사고 통계를 집계한 이후 통계자료가 다양해지고 있다고는 하나 더욱 대형화되고 복잡해지는 해상활동과 이에 따라 증가하는 해양사고 추세를 생각한다면 현재 관리하는 통계 기록물 이상의 정보관리와 활용이 필요하다.

Kim et al.(2004)는 대형화되는 건설 산업에서 통합적으로 정보를 활용할 수 있는 분류체계를 ‘온톨로지(ontology)’ 기반으로 제시했다. 다양화되고 방대해진 정보를 효과적으로 관리하기 위해 개념들 간 상호연관성까지 고려한 것인데, 이러한 온톨로지의 적용을 통해 기존 단순 분류체계로는 정보의 의미론적인(semantic) 부분을 파악하기 어려운 부분을 보완할 수 있는 체계를 제안했다.

온톨로지는 정보를 조직하는 체계적인 방법으로 공유되는 개념에 대한 정형화된 사전이며 특정 도메인에 관련된 단어들이 계층적으로 표현된다. 온톨로지는 상호연관성을 표현하는데 뛰어난 개념으로 종합적인 정보를 표현하는데 사용한다.

Lee(2011)는 인적재난 저감을 위한 전조정보 위험등급 제시를 위해 연구과정 중 인적재난 사례분석과 온톨로지를 활용하였다. 재난정보를 분류하기 위해 사고사례를 분석하고 텍스트마이닝을 통해 최종적으로 재난전조정보 온톨로지 기본요소를 취약성, 자원, 위험(재난), 대응·조치사항, 이해관계자, 재해(피해)로 구분하였다. 여기서 취약성이란 실제 위험(재난)을 발생시킬 수 있는 물적요인의 결함이나 인적요인의 불안정한 행위 등이며, 적절한 대응조치에 따라 경감될 수 있다. 자원은 인적·물적자원으로 구분되며 재난이나 사고관리를 위해 수행해야 할 임무가 있는 것을 의미한다. 위험(재난)은 발생 가능한 모든 사고나 재난을 말하며, 대응·조치사항은 재난 및 사고를 예방하거나 피해를 최소화 하는 수단을 말한다. 이해관계자는 안전관리를 위한 요구사항을 지정하고 관리, 이행하며 내용에 대한 책임을 가진 사람이나 조직이다.

위의 재난안전관리 6개 기본요소에 따라 해양사고 선행연구나 관련 문헌에서 다루지고 있는 분류체계 구성을 파악했다(Table 1). 사고 발생 ‘원인’ 과 관련성이 높은 ‘취약성’에 대한 분류체계는 대부분의 문헌에서 다루고 있었다. 그리고 사고현황과 관련하여 사후에 비교적 쉽게 파악할 수 있는 정보인 사고유형 및 선종(피해자원), 선박피해유형(재해)에

Table 1. Basic factor analysis by literature review

Literature review / Basic factor	Korea Coast Guard (2018)	Korean Maritime Safety Tribunal (2018)	Statistics Korea (2017)	Japan	Kim et al. (2011)	Kim and Kwak (2011)	Kwak (2011)	Kim and Kang (2011)
Stakeholder								
Vulnerability	cause of accident	cause of accident		cause of accident	cause of accident	cause of accident	cause of accident	weather
Resources	vessel type, organization structure	vessel type		vessel type				
Risk (accident)	accident type	accident type	accident type	accident type				
Response·actions taken								
Disaster (results)	vessel damage, human casualties	vessel damage, human casualties						

대한 분류체계 또한 어느 정도 갖춰져 있었다. 반면, 사고발생 시 대응 조치사항이나 조치사항을 이행하는 구조세력, 이해관계자 대한 부분은 다소 연구가 미비한 실정이며, 자원부분에 있어서도 피해대상 관점의 자원인 ‘선종’ 분류체계가 주를 이루고 있다.

3. 상황보고서 속성분석 및 해양사고 온톨로지 정립

국내 해양안전을 위한 정책수립, 연구과정에서는 대부분 통계연보를 활용하고 있으나, 현재 제공되고 있는 통계 분류항목으로는 사고 예방 및 대응을 위한 실효성 있는 정책제안에 한계가 있다. 해양경찰청의 통계연보는 상황보고서를 근간으로 작성되나 현재 상황보고서는 작성자가 직접 텍스트로 서술하는 비구조적 서식이다. 이로 인해 상황보고서에 많은 정보가 기록됨에도 불구하고 실질적으로 수집하여 관리하는 항목은 통계연보 수준으로 제한적이다.

따라서 문헌연구 결과 및 위와 같은 문제발견에 따라 한 정적인 분류체계에서 탈피하여 다양한 데이터를 지속적으로 수집할 수 있도록 하는 해양사고 온톨로지를 구축한다.

3.1 사고 유형별 상황보고서 용어 속성 분석

상황보고서를 작성 시점별로 나눠보면 신고 접수 후 가장 먼저 전파되는 발생보고서(1보), 진행사항 및 확인되는 사항을 작성하는 중간보고서(2보~최종보고서 전), 그리고 발생보고서부터 사고 종료시점까지의 모든 정보를 담은 최종보고서로 구분할 수 있다.

본 연구에서는 상황보고서 상에서 표현되고 있는 사고 데이터의 속성들이 무엇인지 파악하기 위해 보고서 상 작성된 용어를 추출하여 분류하였다. 기존 분류체계(문헌연구) 중

그대로 활용할 수 있는 부분은 수용하고, 분류체계 범주에 속하지 못하는 새로 추출된 용어의 경우 유사 속성(attribute)끼리 묶어 새로운 범주(category)를 구성했다.

용어를 속성별로 분류한 결과(Table 2), 사고유형에 따라 다소 차이는 있으나 대부분 사고구분을 위한 정보(사고 기본정보), 초기 수집되는 사고정보 및 전과정보, 사고 발생장소에 대한 정보, 선박제원(선박정보), 기상/해상특성 정보, 사고유형(연쇄위험/복합사고), 재해(피해결과-인명피해, 선박피해, 해양오염피해), 투입자원, 대응조치사항, 취약성, 사고종료(사고처리결과)와 관련된 텍스트로 작성된다.

3.2 해양사고 분류체계 및 온톨로지 정립

상황보고서는 비구조적인 텍스트로 작성되고 있고, 온톨로지 정립을 위해서는 텍스트를 구조화해야 한다. 이를 위한 과정은 텍스트 마이닝(text mining)을 활용한다. 이 분석기법은 구조화되지 않은 텍스트에서 패턴과 지식을 찾는 것이나 텍스트는 불규칙하고 암시적이기에 이를 구조화하는 과정이 필요하다.

일반적으로 구조화 하는 과정은 전처리를 통해서 진행한다. 전처리는 분석에 적합한 형태로 데이터를 가공하는 것으로, 본 연구에서는 텍스트 추출하여 데이터베이스(DB)화한 후 수차례 전처리 과정(유사어, 중복단어, 혼용어 등 정제)을 통해 표준용어 분류를 제시하였다.

텍스트 구조화를 위한 엑셀(excel) DB 테이블의 속성(attribute)은 상황보고서 속성분석 결과(Table 2)를 토대로 작성했다. 상황보고서에서 사용되는 용어 속성을 엑셀 열(column)에 기입하고, 해당 속성과 관련된 텍스트를 상황 보고서에서 추출하여 DB에 입력한다. 추후 속성 별로 모아진 텍스트를 전처리하여 최종적으로 남은 텍스트 간 특성을 고

Table 2. Status report terminology attribute classification

Attribute classification	Specifics
Basic information about the accident	Serial number, document number, case title, number of reports
Initial accident information and radio wave information	date & time of the accident, time of the report, time of the shipment, means of filing the report, the declarer, the current status of the ship (incident situation),
Accident location information	location of the accident, diameter, address of the accident
Vessel Information (Ship Spec)	vessels, tonnage, vessel type, shipment, unregistered vessel, load type. number of passengers, number of foreigners, etc.
Weather / Sea Characteristics Information	Weather condition, cloudy/clear, wind direction, wind speed, digging, visibility distance, high / low tide (hour)
Accident (risk) information	1st accident, 2nd accident, 3rd accident (chain risk, compound accident)
Damage result information	Rescuer (fair condition), injured person, dead person, missing person, type of ship damage, type of spilled oil, runoff
Incident Response Information / Input resource information	first input force, initial arrival time, dispatch distance, additional input forces, input equipment / resources, response measures, system availability
Cause of the accident	response vulnerability, alcohol intake, unlicensed, incident vulnerability (final cause of accident), oil spill cause
End of incident information	accident end date, accident end time, accident end location, final process results

려해 온톨로지 기본요소를 분류했다.

다음 Table 3은 문헌연구, 상황보고서 속성분석, 텍스트 마이닝을 통해 수정·추가한 해양사고 분류체계이다. 기존에 분류체계는 있으나 수정을 진행한 경우는 기존 분류가 명확하지 않아 전처리 과정에서 서로 중복되거나 혼선, 사각지대가 발생한 용어가 있는 경우이다. ‘신고자, 신고수단, 구조세력, 피해유형’의 경우가 수정된 사례이다. 그 외 새로 추가(보완)한 분류체계는 ‘대응·조치사항, 대응취약성, 적재물, 유류유출경위, 피해유형, 사고처리결과’이다. 충분히 통계항목으로 향후 지속적으로 수집할 수 있음을 보여준다. (본 연구에서 수정·추가한 분류 외에도 분류 항목은 더 존재하며, 추가로 수정·정립해야할 속성들이 있으나 상황보고서 자료 누락 문제로 진행하지 못했다.)

Table 3. Marine Accident Classification Framework

Classification	Subdivision (standard term)
Informant (correction)	Accident party
	Third party (public)
	Third party (nearby vessel)
	Marine Police
	Government agency
	Related agency
Reporting means (correction)	Telephone
	SSB
	VHF
	Report a distress
	Visit
	Marine Police TRS
	Marine Police VTS
Rescue forces (correction)	Marine Police
	Civil marine rescue
	Private fishing boats, civilians
	Fire/Military/Pipeline
	Self-effort
Response measures (addition)	Rescue
	Search
	Maritime work
	Control work
	Fire suppression
	Safety management
	Investigation/Research/Analysis
	Situation verification
	Towing/Lifting
Response vulnerability (addition)	Bad weather
	Marine Environment (Feature)
	Vessel hazard status
	Lack of capacity
	Lack of resources
	Instruction failure
	etc.

Table 3. (continued)

Classification	Subdivision (standard term)	
Load (addition)	Oil	
	Non-oil	
Oil spillage (addition)	Outflow during work	
	Deliberate outflow	
	Old/defective	
	Environmental factors (weather, marine)	
	Accident result	
Damage type (correction, addition)	Unknown cause	
	Physical damage	Loss of utility (sinking)
		Burnt
		Property damage
		Pore
		Amputation
		Modification
		Scratch
		Crack
		Breakdown
		Elimination
		Loss
		Flooding damage
		Partial disappearance
		Others
		No damage
		Casualties
Injury		
Death		
Disappearance		
No spill		
Pollution damage	Oil spill	
	Natural extinction	
	No spill	
Accident handling results (addition)	Rescue	
	Aid (resources)	
	Inability to rescue	
	Future response measures	
	Response measure completed	
	Report cancellation	
	Error (no abnormality)	
etc.		
Other classification framework (Annual statistics)	• outbreak area	
	• outbreak type	
	• outbreak cause	
	• reporting route	
	• first arrival forces (rescue forces)	
	• case reception route	
	• case of the accident vessel	
	• gross tonnage	
	• vessel type	
	• quality of radiation	
	• climate/month/time	
• office		
• emergency patient occurrence place, etc.		

Table 4. Defining basic factors of marine accident ontology

Basic factor	Definition
Stakeholder	<ul style="list-style-type: none"> An individual (or group) who designates and manages requirements (legislation, guidelines, manuals) for disaster / accident management Owners of resources that have been affected by a disaster / accident
Vulnerability	<ul style="list-style-type: none"> Factors to be managed in order to minimize risk (accident) prevention and disaster (damage) The more vulnerable factors, the greater the scale of the disaster (damage) or the vulnerability causes the accident
Resource	<ul style="list-style-type: none"> Resources from responders' point of view: Resources required to implement disaster management responsibilities (rescue forces, rescue input resources, etc.) Resources from victims' point of view: Resources damaged by the accident
Risk (accident)	<ul style="list-style-type: none"> Factors causing actual accidents and disasters (vessel accidents, non-vessel accidents)
Response actions taken	<ul style="list-style-type: none"> Means to prevent disasters / accidents or minimize damage Includes all measures related to a series of processes from prevention to preparation, response and recovery for marine accident management
Disaster (damage results)	<ul style="list-style-type: none"> Result of damage caused by accident (damage of property due to loss of ship, damage of human life, damage of environmental pollution (marine pollution), etc.)

수집되는 비정형 데이터를 포함하여 구조화된 자료는 온톨로지 기본요소 개념에 모두 포함된다. 선행연구에서 제시한 재난·사고 온톨로지 기본요소 개념을 수정하여 해양사고 온톨로지 기본요소를 Table 4와 같이 정의하였다.

앞서 문헌연구에서 언급한 것과 같이 온톨로지는 방대한 정보를 관리하는데 효과적이고 개념들 간 의미는 물론 상호연관성을 파악할 수 있다는 장점이 있다. 새롭게 정의한 해양사고 온톨로지 기본요소에 따라 상호연관성을 도식화하여 살펴보면 요소 간 상호 유기적으로 연계된 것을 알 수 있다(Fig. 1).

위험(선박/비선박사고)은 취약성이 존재할 때 발생하고, 취약성은 사고를 일으키는 원인이 된다. 위험(선박/비선박사고)이 발생하면 선박, 적재물, 승선원 등은 위협을 받게 되며, 이러한 자원이 피해를 입게 되면 재해가 발생했다고 할 수 있다. 구체적인 사고피해의 결과는 인명피해, 물적피해, 환경피해(해양오염 등)로 나눌 수 있다. 취약성, 또는 재해를 줄이기 위해서는 조치사항(대응사항)이 이해관계자 또는 자원에 의해 수행되어야 한다.

기본요소 간 상호연관성이 있다는 것은 단순히 사고결과에 대한 일부 결과(현재 통계제공 항목)만이 해양사고 저감

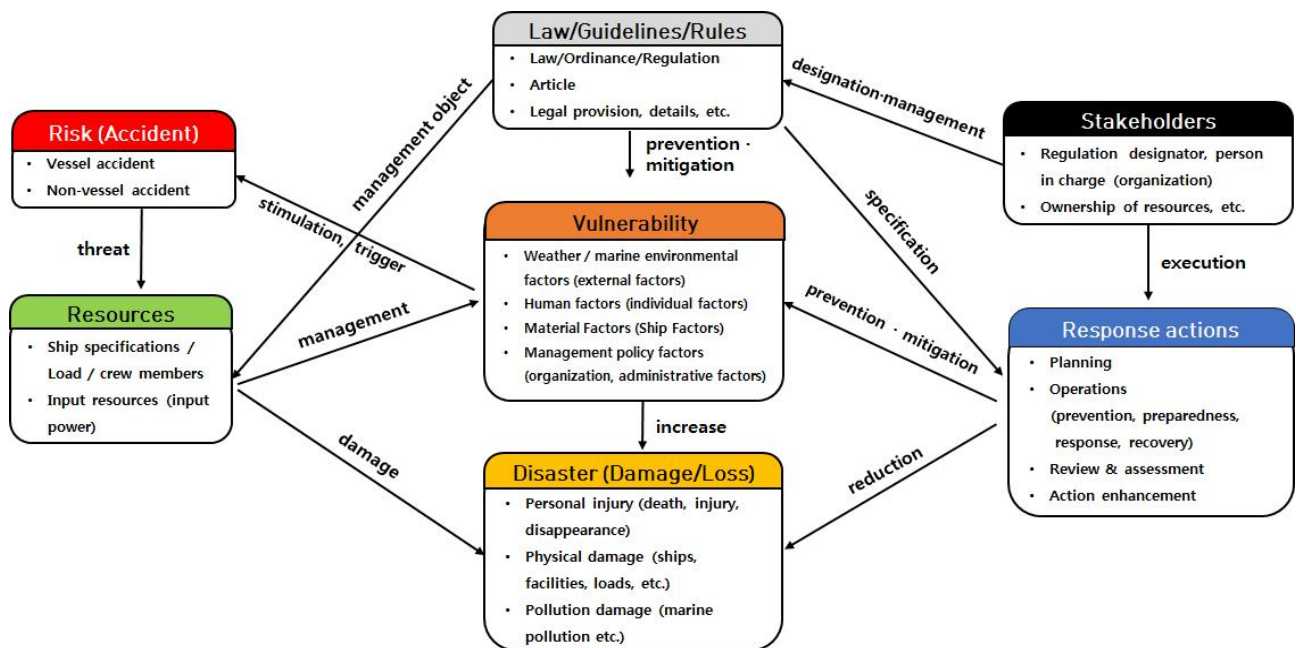


Fig. 1. Correlation between basic factors of marine accident ontology.

이나 예방을 위해 중요한 정보가 아님을 의미한다. 대응조치사항, 자원, 이해관계자 등 다양한 기본요소 내 관련 정보가 수집·저장되고 분석결과를 다양한 연구와 정책수립에 활용되어야 함을 의미한다.

4. 해양사고 데이터수집 및 품질관리 방안

상황보고서는 작성자마다 직접 서술형으로 작성하는 비정형 형태의 보고서이며, 현장에서 확인되는 사항에 따라 작성항목이 달라짐은 물론, 담당자마다 중요하다고 판단되는 사항에 따라서 보고서의 내용이 달라진다는 특성이 있다. 문제는 정형화되지 않은 보고서로 인해 추후 데이터 수집 및 저장은 물론 데이터 활용에 있어 품질에 대한 문제가 야기될 수 있다는 것이다. 데이터 품질이란 목표 달성을 위해 사용되는 데이터들이 업무 목적에 적합하도록 정확도와 유용성, 가치를 유지하는데 필요한 특성을 말한다. 실제로 본 연구에서 해양사고 분류체계 정립을 위해 속성별 용어를 추출하고 텍스트 마이닝을 실시하는 과정에서 불확실한 정보와 누락된 데이터가 많았으며 이러한 문제는 향후 활용에 있어 신뢰성과 정확성에 영향을 미치게 된다는 점에서 매우 민감하다고 할 수 있다.

상황보고서는 해양사고 발생부터 대응, 종료시점까지의 모든 사항을 확인할 수 있는 핵심문서이므로 필요한 정보는 누락하지 않고 작성하여 지속적으로 데이터나 정보가 수집될 수 있도록 해야 한다. 즉, 비정형 보고서상에서도 일부 주요항목에 대해서는 정형화된 양식을 제공하여 데이터가 누락되거나 불확실하게 기재되는 상황을 사전에 예방하는 것이다. 이를 위해서는 사고유형별 상황보고서 특성뿐만 아니라 작성되는 시점별 특성까지도 고려한 정형화된 보고서 서식 개발이 필요하다.

온톨로지 정립을 위한 상황보고서 속성분석 과정에서 이러한 표준 서식 개발을 위한 몇 가지 시사점을 발견할 수 있었다. 첫째, 표준 서식은 반드시 조사되어야 할 항목과 작성 시점별 정보가 일관성 있게 유지되어야 하는 항목을 파악하여 개발되어야 한다. 이는 정보가 누락되거나 보고서가 작성되는 시점별로 내용이 변경되는 내용의 불확실성을 제거할 수 있도록 해준다. 둘째, 최종보고서(종합보고서)에서 통계연보로 직접 연계되는 항목을 고려해서 서식을 개발해야 한다. 최종보고서는 사고 상황에 대한 전반적인 정보를 모두 담고 있으며, 해양경찰청에서 발간하는 통계연보 작성의 근간이기 때문에 통계기입 주요정보는 반드시 작성되도록 해야 한다. 국내의 많은 해양관련 연구가 통계연보를 활용하고 있기 때문에 데이터 품질은 연구결과에 영향을 미치며 이는 곧 신뢰성의 문제로 지적될 수 있다. 통계연보가 단순히 연구용으로만 활용되는 것이 아니라 정책수립을 위한 기

초자료로 활용됨을 생각한다면 데이터 품질확보를 위해 통계연보의 품질, 즉 통계연보의 근간이 되는 상황보고서 상 수집되는 데이터 품질부터 확보해야 한다.

즉, 일관성 있는 데이터 기록을 위한 상황보고서 표준 서식을 활용하여 데이터를 일관성 있게 수집할 수 있으며, 일부 구조화된 표준항목의 경우 데이터의 품질까지 확보할 수 있다. 데이터의 수집과 저장, 품질관리의 중요성은 아무리 강조해도 지나치지 않으나, 가장 중요한 것은 ‘상황보고서’가 작성되는 ‘본질’을 잊어서는 안 된다는 것이다. 상황보고서는 사고발생 후 급박한 순간에서 사고 상황 공유를 위해 작성되는 것이므로 데이터 수집과 품질에 연연하여 너무 정형화되고 고정된 틀을 제공할 경우 신속한 사고대응에 방해될 수 있다. 따라서 표준 서식은 상황보고서 작성 시점별 특성을 고려하여 개발하고 활용해야 한다.

5. 연구결론 및 시사점

국민소득 증대, 주5일 근무정착에 따른 수상레저, 해상관광, 낚시 등이 활발해지고, 높은 어획강도, 활발한 해상교역, 여객수송 등은 해양사고 발생가능성 또한 증가시켰다. 이에 따라 해양사고를 줄이기 위한 각종 정부정책이 등장하고, 학계에서는 해양안전을 위한 연구를 진행해오고 있다. 문제는 이러한 노력에도 불구하고 사고감소폭은 크게 줄지 않고 있다는 점이며, 이는 곧 실제 사고나 피해를 줄일 수 있는 정책이나 대책이 실현되지 못하고 있음이 드러나는 대목이라고 할 수 있다.

정책수립, 다양한 개선대책을 제안하기 위해 행정기관 및 연구자들이 주로 활용하는 자료는 통계연보 및 사고사례이다. 두 자료는 해양사고의 현황과 경향을 파악하기 위해 가장 쉽게 접근할 수 있는 공개 정보이기 때문이다. 본 연구에서 진행한 문헌연구에 따르면 대부분의 선행연구들이 통계연보나 사례집에 근거하여 유사한 개선책을 제시하고 있다. 이는 해양사고 분석연구에 활용되는 통계자료가 수치 편차는 있으나 분석결과에 있어서는 모두 같은 통계자료를 활용하기 때문에 도출되는 결론이 대부분 유사할 수밖에 없다는 한계를 명백히 보여주는 대목이다. 이는 연구자뿐만 아니라 행정기관에서 정책수립과정에서 통계연보나 사례집에 있어서만큼은 제한적인 자료를 활용하고 있음을 보여준다.

본 연구에서는 현행 해양안전 관련 문헌연구를 통해 선행연구의 동향을 살펴보고, 기존 연구의 한계와 이를 개선하기 위한 방안을 ‘해양사고 온톨로지’와 ‘데이터 수집 및 품질확보방안’을 통해 제시하였다. 해양사고 온톨로지의 경우, 기존 통계연보 및 선행연구, 서해남부 해역 관할 해양경찰서에서 작성한 지난 3년 간(‘15~’17)의 상황보고서 중 주요 선박사고 상황보고서 속성을 분석하여 텍스트 마이닝을

통해 제시하였다.

기존 분류체계 대비 수정·추가된 항목은 ‘신고자, 신고수단, 구조세력, 대응·조치사항, 대응취약성, 적재물, 유류유출경위, 피해유형, 사고처리결과’이며, 이 항목들은 분류체계 표준용어를 활용해 향후 지속적으로 수집·활용할 수 있다. 현재 통계연보에서 제공되고 있는 항목은 대부분 사고유형, 피해결과, 원인 등에 초점이 맞추어져 있으며 초기대응이나 현장대응, 투입자원(세력) 등에 관련한 통계항목은 다소 미비한 편이다. 따라서 새로운 분류체계 활용을 통해 지속적으로 데이터를 수집하고 저장할 필요가 있다.

더불어 데이터의 품질확보가 수반되어야 한다. 이를 위해 현재 비정형으로 작성되는 상황보고서 서식을 구조화하여 사용할 것을 제안하였다. 반드시 조사되어야 할 항목과 작성시점별 정보가 일관성 있게 유지되어야 할 항목, 통계연보로 직접 연계되는 항목을 고려한 표준 서식을 활용해야 데이터의 품질이 확보될 수 있다. 상황보고서는 통계연보의 근간이 되고, 통계연보는 정책수립이나 연구에 활용되는 주요 자료이므로 데이터의 품질은 매우 중요하다고 할 수 있다.

매일 발생하고 있는 크고 작은 사건·사고, 이로부터 작성되는 상황보고서에서 정형화할 데이터를 발견하고 구조화하여 지속적으로 수집하고 관리한다면 무심코 지나쳤던 ‘의미’ 들이 비로소 보이기 시작할 것이다. 상황보고서는 해양안전을 위해 우리가 해결해 나가야 할 많은 정보와 시사점을 내포하고 있다고 해도 과언이 아니다. 다만 비정형이다 보니 의미 있는 정보와 패턴, 우리가 나아가야 할 방향을 적시에 찾지 못하고 있을 뿐이다. 지금부터라도 상황보고서에서 중요한 정보를 찾아내어 이를 활용하기 위한 노력이 시작되어야 한다.

가장 중요한 것은 반복적으로 발생하는 해양사고에 대비·대응하기 위해 현재 직면한 문제는 무엇인지, 이를 해결하기 위해 필요한 정보는 무엇이고, ‘신뢰성’ 있는 정보가 ‘충분히’ 확보 되어있는지를 명확히 이해해야 비로소 실효성 있는 정책수립과 연구가 진행될 수 있다는 것이다. 현재 해양사고 저감·예방을 위한 다양한 국가 안전정책이 시행되고 있으나 그 피해가 눈에 띄게 줄지 않는 점을 생각한다면, 현재 활용하고 있는 자료와 연구결과에 대한 유의미성을 다시 한 번 생각해야 한다. 우리가 활용하고 있는 정보가 의사결정에 있어 의미 있는 자료인지, 실제 사고 예방과 대응을 위해 더 많은 신뢰성 있는 정보가 필요하지는 않은지 현실태를 점검해볼 필요가 있다.

향후에는 본 연구에서 제시한 해양사고 온톨로지에 대한 적절성 검증과정이 필요하며, 상황보고서 분석 과정에서 데이터 누락, 정보부족 등의 문제로 진행하지 못한 기존의 다른 분류체계에 대한 검토 또한 진행되어야 할 것이다.

후 기

본 연구는 2018년 서해지방해양경찰청 연구비 지원으로 수행된 ‘해양사고 분석 및 데이터 관리체계 구축연구’의 결과물로 작성된 논문입니다.

References

- [1] Cho, H. K., B. S. Park, D. H. Kang and S. S. Kim(2017), The Main factor and Counterplan for Marine accidents in Korea, Journal of Fisheries and Marine Sciences Education, Vol. 29, No. 3, pp. 746-756.
- [2] Japan(2019), Japan Transport Safety Board, <http://www.mlit.go.jp>.
- [3] Ju, J. K.(2016), A Study on the Possibility of Cooperative Governance System for Marine Accident Response: From the Perspective of Major Marine Accidents in the Korea waters, Korean Police Studies Review, Vol. 15, No. 4, pp. 213-234.
- [4] Kang, I. K., H. S. Kim, J. C. Kim, B. S. Park, S. J. Ham and I. H. Oh(2013), Study on the marine casualties in Korea, Journal of the Korean Society of Fisheries and Ocean Technology, Vol. 49, No. 1, pp. 29-39.
- [5] Kim, D. J. and S. Y. Kwak(2011), Evaluation of Human Factors in Ship Accidents in the Domestic Sea, Journal of the Ergonomics Society of Korea, Vol. 30, No. 1, pp. 89-98.
- [6] Kim, D. S.(2018), A Study on the Prevention of Ship Collision in Low Visibility: Focusing on the Role of Korea Coast Guard, Journal of Korean Maritime Police Science, Vol. 9, No. 3, pp. 71-85.
- [7] Kim, H. R., E. J. Park, H. G. Kim and S. H. Yoon(2004), Building an Ontology based on the Unified Construction Information Classification System, The Journal of Society for e-Business Studies, Vol. 9, No. 3, pp. 95-112.
- [8] Kim, H. T., S. Na and W. H. Ha(2011), A Case Study of Marine Accident Investigation and Analysis with Focus on Human Error, Journal of the Ergonomics Society of Korea, Vol. 30, No. 1, pp. 137-150.
- [9] Kim, H. Y., C. H. J and D. B. Kim(2017), Analysis of the Marine Accidents Prevention Effect due to Implementation of Maritime Safety Policy, Journal of Korean Maritime Police Science, Vol. 7, No. 2, pp. 129-144.
- [10] Kim, S. K. and J. P. Kang(2011), A Study on the Relationships between the Casualties of Fishing Boats and Meteorological Factors, Journal of Fisheries and Marine Sciences Education, Vol. 23, No. 3, pp. 351-360.

- [11] Korea Maritime Institute(2017), Large-scale maritime accidents, preventive measures first, KMI Weekly Report, Vol. 53, No. 53, pp. 1-21.
- [12] Korean Maritime Safety Tribunal(2018), 2017 Maritime Accident Statistics and Accident Cases.
- [13] Korea Coast Guard(2018), Maritime Accident Statistical Yearbook 2017.
- [14] Kwak, S. Y.(2011), Identification and quantitative risk analysis of maritime ship accident in the domestic sea, Pusan National University.
- [15] Lee, S. J., H. S. Kim, Z. J. Long and S. K. Lee(2011), A Study on the Korea Marine Accidents and the Countermeasures, Journal of Navigation and Port Research, Vol. 35, No. 3, pp. 205-211.
- [16] Lee, Y. J.(2011), Developing an Intelligent System for the Analysis of Signs Of Disaster, Journal of The Korean Society of Societal Security, Vol. 4 No. 2, pp. 29-40.
- [17] Roh, H. R.(2014), Cases Analysis of Maritime Accidents and Countermeasure of Korea Coast Guard -Focus on the Passenger Ship Accidents-, Korean Association of Public Safety and Criminal Justice Review, Vol. 23, No. 4, pp. 127-160.
- [18] Statistics Korea(2017), Current Status of maritime accident: 2017 Evaluation of the Quality of Regular Statistics Result Report.

Received : 2019. 05. 17.

Revised : 2019. 06. 25.

Accepted : 2019. 06. 27.