

텍스트 마이닝 기법을 활용한 설문 문항 개선에 관한 연구

백연지* · 정창현**†

* 목포해양대학교 대학원 박사과정, ** 목포해양대학교 교수

A Study on Questionnaire Improvement using Text Mining

Yun-Ji Paek* · Chang-Hyun Jung**†

* Ph.D. Candidate, Graduate School, Mokpo National Maritime University, Mokpo 58628, Republic of Korea

** Professor, Mokpo National Maritime University, Mokpo 58628, Republic of Korea

요 약 : 국민의 해양안전문화 수준을 객관적으로 측정하고 해양안전문화 확산을 위한 자료로 활용하고자 2018년에 해양안전문화지수를 개발하였다. 안전문화지수를 산출하는 방법은 안전문화에 영향을 줄 만한 이슈를 포함해야 하고 현 실태를 측정할 수 있는 문항으로 구성되어야 한다. 또한, 사회적·경제적 변화에 따라 지속적인 검증과 보완이 요구된다. 해양 전문가에 의해 설계된 설문 문항이 국민의 관심사와 요구를 잘 반영하고 있는지 확인하기 위해 915명의 해양안전 관련 제안 내용을 분석하였다. 비정형 데이터인 해양안전 제안 내용을 분석하기 위해 텍스트 마이닝 기법을 활용하였으며, 네트워크 분석과 토픽 모델링을 수행하였다. 해양안전 제안을 분석한 결과 ‘교육’, ‘홍보’, ‘안전수칙’, ‘의식’, ‘전문 인력’, ‘시스템’에 관한 내용이 주를 이루었다. 해양안전 제안 사항이 2019년 설문 문항에 반영되도록 18개의 문항을 수정·보완하였고, 설문 문항의 신뢰도를 분석한 결과 내적 일관성은 0.895로 높게 평가되었으며 전년 대비 향상되었다. 해양 관련 전문가뿐만 아니라 국민의 요구사항까지 반영한 개선된 설문 문항으로 해양안전문화지수를 도출함으로써 해양안전문화 확산을 위한 정책 수립에 더 기여할 것으로 기대된다.

핵심어 : 해양안전문화지수, 텍스트 마이닝, 네트워크 분석, 토픽 모델링, 설문 문항

Abstract : The Marine Safety Culture Index (MSCI) was developed in the year 2018 for objectively assessing the public safety culture levels and for incorporating it as data to spread knowledge regarding the marine safety culture. The method for calculating the safety culture index should include issues that may affect the safety culture and should consist of appropriate attributes for estimating the current status. In addition, continuous verification and supplementation are required for addressing social and economic changes. In this study, to determine whether the questionnaire designed by marine experts reflects the people's interests and needs, we analyzed 915 marine safety proposals. Text mining was employed for analyzing the unstructured data of the marine safety proposals, and network analysis and topic modeling were subsequently performed. Analysis of the marine safety proposals was centered on attributes such as education, public relations, safety rules, awareness, skilled workers, and systems. Eighteen questions were modified and supplemented for reflecting the marine safety proposals, and reliability of the revised questions was analyzed. Furthermore, compared to the previous year, the questionnaire's internal consistency was improved upon and was rated at a high value of 0.895. It is expected that by employing the derived marine safety culture index and incorporating the improved questionnaire that reflects the requirements of marine experts and the people, the improved questionnaire will contribute to the establishment of policies for spreading knowledge regarding the marine safety culture.

Key Words : Marine Safety Culture Index (MSCI), Text mining, Network analysis, Topic modeling, Questionnaire

* First Author : vlotusv@naver.com

† Corresponding Author : hyon@mmu.ac.kr, 061-240-7182

※ 이 논문은 “해양안전 관련 제안 내용분석을 통한 해양안전문화지수 개선방안”이란 제목으로 2019년도 해양환경안전학회 공동학술대회 (목포해양대학교, 2019.04.25-26., p.167)에 발표되었음.

1. 서론

1986년 소련 체르노빌 원자력 누출사고에 따른 국제원자력 안전자문단(INSAG)의 보고서에 처음 사용된 ‘안전문화’는 조직과 개인의 자세와 품성이 결집한 것으로서 모든 개인의 헌신과 책임이 요구되는 것이라고 정의되었다. 하지만 근래의 ‘안전문화’는 안전을 실천하는 의식, 안전을 유도하는 제도, 안전을 가능하게 하는 인프라가 결합해 만들어 내는 사회적, 문화적 산물로 정의되고 있다(KOSHA, 2019).

이에 현재의 안전문화 수준을 객관적으로 평가하여 안전문화 선진화를 위한 노력을 유도하고자 문화지수를 개발하여 활용하고 있다. 국내에서는 지자체 교통문화 수준을 객관적으로 평가하여 교통안전에 대한 자율적 경쟁을 도모하고 교통안전 정책의 기초 자료로 활용하고자 교통문화지수를 개발하였으며, 현장 조사와 사고 통계를 기반으로 2008년부터 매년 발표하고 있다(MOLIT, 2019). 미국에서는 비영리 단체인 미국자동차협회(AAA)에서 도로상에서의 안전과 관련된 문화를 이해하고 강화하는 목적으로 교통안전문화지수를 개발하였으며, 교통안전에 영향을 줄 만한 사회적 이슈와 운전자 행동으로 구분한 설문조사를 기반으로 2008년부터 매년 발표하고 있다(AAA, 2019). 그리고 영국에서는 비영리단체인 IAM(Institute of Advanced Motorists)에서 AAA의 교통안전문화지수 개념을 바탕으로 한 자체 설문조사를 개발하여 시행 중이다(IAM RoadSmart, 2016). 해양 분야에서는 한국 선원의 인식조사를 토대로 안전문화를 제고시키는 방안을 제시하였으며, 해양안전문화 수준을 객관적으로 측정하고 해양안전문화 확산을 위한 자료로 활용하고자 2018년에 해양안전문화지수(MSCI, Marine Safety Culture Index)를 개발하였다(Kim, 2013; Paek et al., 2018).

안전문화는 제도와 인프라, 의식이 결합해 만들어 내는 사회적·문화적 산물이므로 안전문화지수를 산출하는 방법은 안전문화에 영향을 줄 만한 이슈를 포함해야 하고 현 실태를 측정할 수 있는 문항으로 구성되어야 한다. 또한, 사회적·경제적 변화에 따라 지속적인 검증과 보완이 요구된다.

본 연구에서는 해양 전문가에 의해 설계된 해양안전문화지수 도출을 위한 설문 문항이 국민의 관심사와 요구를 포함하고 있는지 확인하기 위해 2018년 실시된 설문조사의 해양안전 관련 제안 내용을 분석하였다. 해양안전 제안 내용은 자유 기술한 비정형 데이터이므로 텍스트 마이닝(Text Mining) 기법을 활용하여 분석하였으며, 그 결과를 향후 해양안전문화지수 도출을 위한 설문 문항 개선에 활용하고자 한다.

2. 연구 방법

2.1 분석 대상

2018년 해양안전문화지수를 도출하기 위한 설문조사에서 해양안전에 대한 제안을 자유롭게 기술하도록 한 바 있다. 설문조사는 성별, 연령별, 지역별 표본 할당한 일반 국민 1,000명과 해운선사, 관계기관 등 해양종사자 561명을 대상으로 하였다. 전체 응답자 1,561명 중 ‘안전이 중요하다.’, ‘잘 모름’과 같이 해양안전에 대한 제안과 관련이 적은 내용을 제외한 915명의 의견을 분석 대상으로 하였다.

2.2 분석 방법

설문조사로 수집한 해양안전에 대한 제안 내용은 텍스트 형태의 비정형 데이터이므로 Fig. 1과 같이 텍스트 마이닝 기법을 활용하여 분석하였다. 텍스트 마이닝이란 Feldman과 Dagan에 의해 텍스트 데이터베이스 기반의 지식 발견이라는 개념으로 처음 언급되었으며, 자연어 처리 기술을 바탕으로 기존에 알려지지 않은 유용한 패턴과 지식을 발견하기 위해 기계적 알고리즘으로 텍스트로부터 정보를 자동으로 추출하는 프로세스를 의미한다(Feldman and Dagan, 1995; Hotho et al., 2005). 해양 분야에서는 텍스트 마이닝 기법을 활용하여 ECDIS 사고 보고서와 준해양 사고 보고서를 분석한 바 있다(Kang et al., 2018; Lee et al., 2019).

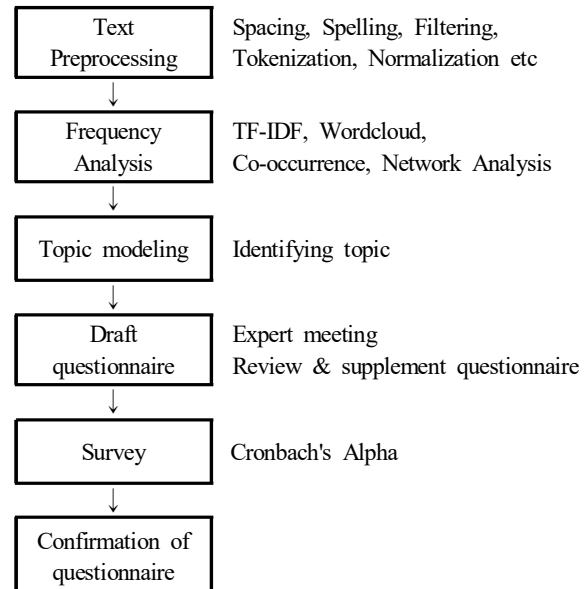


Fig. 1. Procedure for Improving MSCI Questionnaire.

Table 1. Keyword by Uni-gram

No.	Keyword	Frequency	TF-IDF
1	education	390	92.7
2	public relations	134	67.0
3	thorough	65	43.2
4	consciousness	126	39.2
5	observe	46	35.3
6	important	71	35.2
7	management	61	35.2
8	public	118	33.9
9	safety rules	42	32.5
10	rule	19	28.5
11	positive	35	27.5
12	reinforcement	42	25.4
13	interest	40	24.6
14	man	46	23.6
15	legislation	35	22.9
16	facilities	39	22.8
17	insufficiency	35	21.8
18	student	41	21.6
19	prevention	27	20.7
20	execution	32	20.3

3.3 동시출현행렬(Co-occurrence Matrix)과 네트워크 분석

하나의 단어만으로는 전체적인 의미 파악이 쉽지 않고 단어의 순서에 따라 의미가 변화할 수 있다. 이를 구분하기 위해 TF-IDF 가중치를 적용한 TDM(Term-Document Matrix)을 만들고 상위 50개 키워드로 동시출현행렬을 생성하였다.

Fig. 4는 동시출현 키워드의 네트워크를 시각화한 것이다. 네트워크 분석은 각 노드(Node) 간의 상호 작용 및 관계로 만들어진 네트워크의 구조를 계량적으로 분석하는 방법이며 사회과학, 경영학, 응용과학 등 다양한 분야에서 활용되고 있다(Seo, 2016). 단어 간 연관성이 높을수록 서로 가까이에 위치하고, 연결선의 굵기로 어떤 단어와 주로 사용되었는지를 알 수 있으며, 원의 크기는 동시 출현 빈도를 의미한다.

네트워크 분석 결과 ‘교육’과 ‘홍보’는 서로 연관성이 높고 주로 함께 사용되었다. ‘교육’에 관해서는 ‘학생’, ‘국민’을 대상으로 하고, ‘예방’을 목적으로 ‘사전’에 해야 하며, ‘체험’, ‘수영’을 포함하여 ‘관리’, ‘강화’, ‘의무’, ‘지속’해야 한다고 제안하였다. ‘홍보’는 ‘부족’, ‘모르다’라는 반응과 함께 ‘적극’, ‘실시’하고 ‘지속’해야 한다고 제안하였다. 그리고 ‘의식’에 대해서는 ‘국민’의 의식이 ‘부족’하므로 ‘홍보’와 ‘교육’을 통해 의식을 ‘강화’해야 한다고 제안하였다.

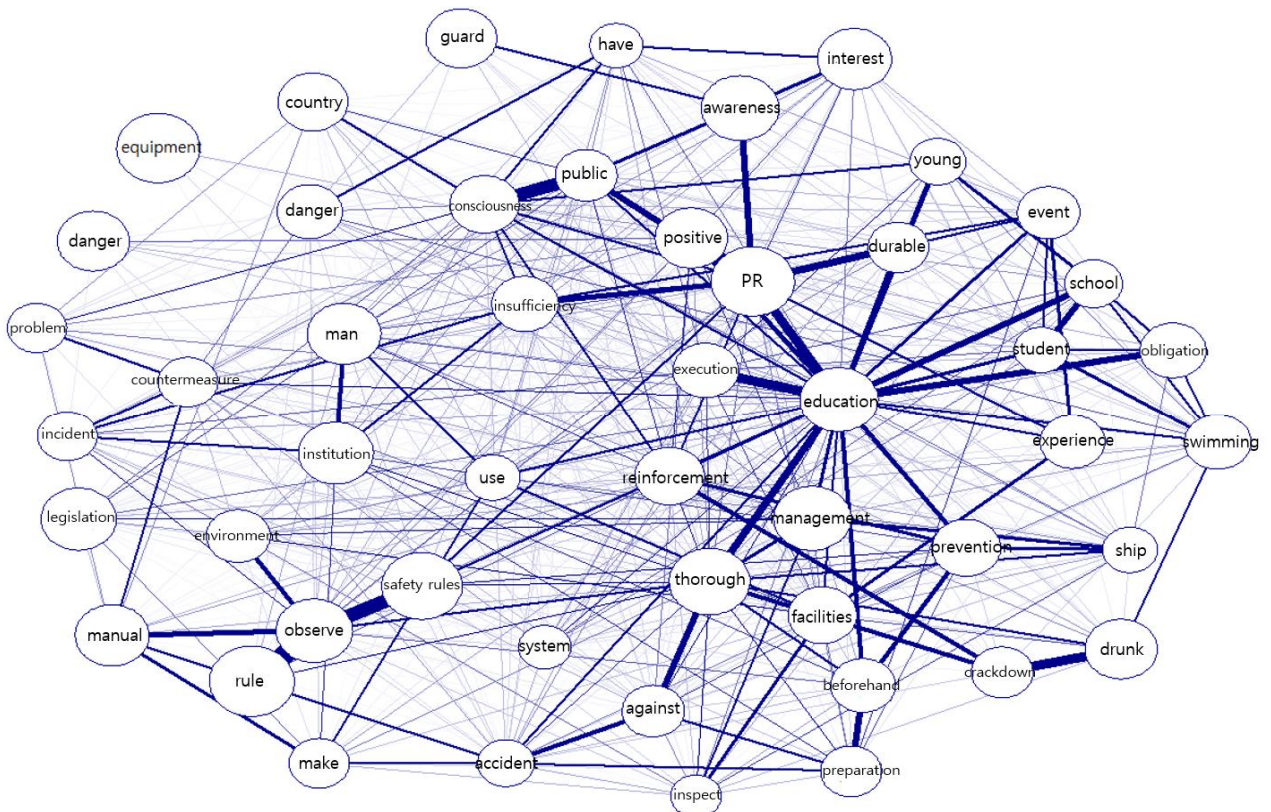


Fig. 4. Network by co-occurrence matrix.

Table 2. Topic models with top ranked keywords of the marine safety proposal

Topic	Keyword									
Topic 1 12.5%	school 19.0%	young 15.2%	experience center 4.4%	have 4.4%	program 3.8%	system 3.8%	teach 3.2%	show 3.2%	kindergarten 3.2%	increase 2.5%
Topic 2 11.7%	observe 18.1%	safety rules 16.8%	passenger ship 7.4%	airplane 3.4%	reach 3.4%	casualty 3.4%	almost 2.7%	violation 2.7%	come out 2.0%	form 2.0%
Topic 3 11.1%	leisure 14.2%	guard 8.5%	garbage 5.0%	sports 4.3%	beach 4.3%	rescue team 3.5%	with 2.8%	play in water 2.8%	advanced country 2.8%	user 2.8%
Topic 4 10.7%	campaign 7.4%	safety jacket 6.6%	beach 6.6%	mass media 5.1%	difficult 5.1%	however 4.4%	safety rules 4.4%	passenger ship 4.4%	not 2.9%	fast 2.9%
Topic 5 10.5%	mass media 9.8%	raise 8.3%	meet 8.3%	have 7.5%	notice 6.8%	public advertisement 4.5%	consider 4.5%	program 4.5%	weak 3.0%	public broadcast 2.3%
Topic 6 9.7%	leisure 17.9%	user 8.9%	awareness 6.5%	not 3.3%	subject 2.4%	afraid 2.4%	infrastructure 2.4%	local government 2.4%	penalties 1.6%	levy 1.6%
Topic 7 9.7%	insensibility 9.8%	no 8.1%	awareness 6.5%	force 5.7%	importance 4.9%	system 4.1%	penalties 3.3%	feel 2.4%	facility 2.4%	tourism resource 1.6%
Topic 8 8.1%	make 16.5%	manual 15.5%	observe 14.6%	personal 4.9%	supervision 2.9%	worker 2.9%	self 2.9%	countermeasure 1.9%	guideline 1.9%	handle 1.9%
Topic 9 8.0%	expert 7.8%	become 7.8%	safety rules 5.9%	observe 3.9%	stop 2.9%	reckless 2.9%	establish 2.9%	many 2.9%	no 2.9%	difficult 2.9%
Topic 10 8.0%	have 12.9%	increase 10.9%	coast guard 10.9%	pay 5.9%	event 5.0%	alertness 4.0%	such as 2.0%	backing up 2.0%	leave 2.0%	beach 2.0%

3.4 토픽 모델링

토픽 모델링은 문서를 이루고 있는 키워드를 바탕으로 주제를 찾아내기 위해 사용되는 방법론으로 대량의 문서 집합에 적용되며 다양한 종류의 데이터에 적용 가능하다(Blei, 2012). 토픽 모델링의 가장 대표적인 방법론은 LDA(Latent Dirichlet allocation)이다. LDA 알고리즘은 생성모델로써 문헌 내의 숨겨져 있는 주제들을 찾아내는 알고리즘이다. 생성모델은 실제 문헌을 작성하는 과정으로 보고 문헌을 작성하기 위해 각 문헌에 어떤 주제들을 포함할 것인지, 또 그에 따라 어떤 단어들을 어떤 주제에서 선택하여 배치할 것인지를 각각의 파라미터로 모델링한다(Park and Song, 2013). 따라서 문헌, 단어 등 관찰된 변수를 통해 문헌의 구조와 같은 보이지 않는 변수를 추론하는 것을 목적으로 하며 결과적으로 전체 문서 집합의 주제들과 문서별 주제 비율, 각 주제에 포함될 단어들의 분포를 알아낼 수 있다(Blei et al., 2003).

전처리 과정을 거친 해양안전에 대한 제안을 LDA 알고리즘으로 토픽 모델링을 수행하였다. LDA 모델의 파라미터 추정을 위해 Gibbs Sampling 방식을 사용하였으며, 모델을 평가하는 지표로 Perplexity 값을 기반으로 최적의 토픽 수를 결

정하였다(Grun and Hornik, 2011; Griffiths and Steyvers, 2004).

Table 2는 샘플링을 5,000회 반복하여 도출한 10개 주제의 출현 확률과 주제별 상위 10개 키워드 분포를 나타낸 것이다. 토픽 1과 6은 ‘교육’, 토픽 4와 5는 ‘홍보’, 토픽 2, 8, 9는 ‘안전수칙’, 토픽 7, 8은 ‘의식’, 토픽 3, 9, 10은 ‘전문 인력’, 토픽 6, 7은 ‘시스템’에 관한 내용이며, 내용을 정리하면 다음과 같다.

(1) 교육

초중고, 유치원과 같이 어릴 때 교육이 필요하고, 보여주 기식 교육이 아닌 체험관과 프로그램 등 교육 체계를 갖추어야 한다. 또한, 수상레저 등 해양 이용자의 교육을 강화해야 한다.

(2) 홍보

공익광고, 대중매체를 활용하여 구명조끼 착용 등과 같은 안전수칙 캠페인을 펼쳐야 한다.

(3) 안전수칙

안전수칙을 모르거나 안전 불감증과 안전수칙을 어기는 것으로 인해 인명피해가 발생하므로 안전수칙을 잘 지켜야 한다.

(4) 의식

안전 불감증을 없애야 하고, 개개인 스스로 안전수칙을 잘 지키는 등 해양안전 의식을 고취해야 한다.

(5) 전문 인력

해수욕장, 수상레저, 물놀이 장소에 구조대, 안전요원, 해양경찰 등 전문 인력을 더 늘려야 한다.

(6) 시스템

해양안전 관련 인프라를 확충하고 해양안전 저해 시 과태료를 부과하는 등 체계를 갖춰야 한다.

4. 설문 문항 개선

해양안전문화지수를 도출하기 위한 설문 문항은 크게 5개 영역으로 구분되어 있다. Part 1은 해양 관련 경험 여부를 ‘예/아니오’로 조사하고, Part 2~5까지는 해양 안전수칙 준수, 해양안전 의식, 해양 안전문화 실태, 해양에 대한 관심 정도를 5점 척도로 조사한다.

4.1 설문 문항 수정

해양안전문화지수는 국민의 인식과 경험을 객관적으로 측정하기 위한 것이므로 변화하는 사회적 관심사와 요구를 반영하여야 한다. 2018년에 설계한 설문 문항은 해양 전문가의 시각만 반영된 것이므로 해양안전에 대한 다양한 시각과 요구사항이 반영될 수 있도록 텍스트 마이닝 기법을 이용해 분석한 해양안전 제안 내용 결과를 반영하여 Table 3과 같이 설문 문항을 수정·보완하였다.

Table 3. Composition of survey questionnaire

Part	Type	Quantity	
		2018	2019
1 Experience in the marine	yes/no	9	10
2 Compliance with marine safety rules	Likert	5	6
3 Marine safety awareness	Likert	11	8
4 Status of marine safety culture	Likert	9	12
5 Interest in the marine	Likert	5	3
Total		39	39

(1) 교육

해양안전 관련 교육 대상자를 확대하고 교육을 지속해야 한다는 제안에 따라 교육 대상자를 성인, 미성년자 및 선박 또는 수상레저기구 이용자로 구분하였다. 또한 ‘해양안전 관련 행사 또는 교육에 참여한 경험’을 묻는 문항을 행사와 교육으로 각각 분리하고, 각 경험을 최근 5년 이내로 한정하였다.

(2) 안전수칙

안전수칙을 잘 모르거나 지키지 않아 문제가 되기 때문에 안전수칙을 잘 지킬 수 있게 유도하는 교육, 홍보, 제재와 같은 방법이 제안되었다. 따라서 해양 안전수칙과 관련하여 어느 분야가 취약한지 확인할 수 있도록 선박 및 수상레저기구, 수영으로 구분하여 안전수칙 준수 여부를 조사하던 기존 문항에 낱시를 추가하였다.

(3) 홍보

해양안전과 관련하여 모른다는 의견이 다수이고, 해양안전문화를 확산하기 위해 적극적인 홍보가 필요하다는 제안에 따라 ‘해양안전 관련 공익광고, 캠페인과 같은 홍보’와 ‘해양안전 관련 행사’가 충분히 제공되고 있는지를 조사하는 항목을 추가하였다.

(4) 인프라

안전요원, 해양경찰, 구조대 등 해양 관련 전문 인력과 함께 해양안전 인프라를 확충해야 한다는 제안에 따라 ‘신속하고 효율적인 구조 시스템 보유 여부’와 ‘해양안전문화 확산을 위한 인력, 시설 등 인프라의 적정성’에 대한 문항을 Part 4(해양안전 문화 실태)에 추가하였다.

(5) 기타

Part 3(해양안전 의식 정도)의 문항 중 ‘개발과 투자’, ‘환경보호’에 관한 문항은 관련성이 적어 삭제하고, 교육 훈련의 법제화를 묻는 문항은 Part 4(해양안전 문화 실태)로 이동하였다. 그리고 Part 5(해양에 대한 관심 정도)의 문항 중 ‘해양에 대한 관심’, ‘해양 관련 뉴스나 기사에 대한 관심’ 문항은 하나의 문항으로 통합하였다.

4.2 설문 문항 신뢰도 분석

측정하고자 하는 사항이 응답자로부터 일관되고 정확한 답변을 얻을 수 있는지, 즉 측정 도구로써 유의미한 가치가 있는지 확인하기 위해 신뢰도 분석을 하였다. 신뢰도 측정 시 가장 많이 사용하는 방법인 내적 일관성에 의한 척도의 신뢰성 평가를 위해 크론바흐(Cronbach)의 알파 계수를 확인하였다. 항목 간의 상관관계가 높을수록 내적 일관성이 높다고 평가하는 것으로 크론바흐의 알파 계수 값이 1에 가까울수록 신뢰도가 높다. 일반적으로 0.8~0.9 이상이면 바람직하고 0.7~0.8이면 수용할 만한 것으로 평가한다. 0.6보다 작으면 내적 일관성이 결여된 것으로 평가한다(Cronbach, 1951).

2019년 설문조사는 일반인의 경우 1,000명을 대상으로 성별, 연령별, 지역별로 표본 할당하여 온라인으로 진행하였으며, 해양종사자의 경우 601명을 대상으로 해운선사, 해양경찰, 해양대학 등 해양 관련 관계기관의 협조를 받아 진행하였다.

Table 4는 2018년 시행된 설문조사와 텍스트 마이닝 기법을 이용해 설문 문항을 수정하여 실시한 2019년 해양안전문화지수 설문조사의 영역별 내적 일치도를 나타내는 크론바흐 알파 계수를 비교한 것이다.

Table 4. Cronbach's Alpha & corrections

	2018	2019	Corrections
Part 2	0.795	0.819	merge 1, add 2
Part 3	0.862	0.862	transfer to part 4, remove 2
Part 4	0.858	0.907	revise 3, separate 2, add 2
Part 5	0.901	0.868	merge 2
Total	0.888	0.895	

수정·보완한 설문 문항의 내적 일관성은 Part 5를 제외하고 모두 향상된 것으로 평가되었다. Part 5는 수정 전보다 크론바흐 알파 계수가 낮아졌지만 0.8 이상으로 여전히 내적 일관성이 높은 것으로 나타났다. 설문 문항 Part 1은 해양 관련 경험을 조사하는 항목으로써 Part 2~Part 5와 척도 기준이 다르므로 신뢰도 측정 항목에서 제외하였다.

5. 결론

본 연구에서는 해양안전문화지수 도출을 위한 설문조사의 설문 문항이 국민의 해양안전 관련 경험과 인식을 측정하는데 적절한지 확인하기 위해 2018년 설문조사를 통해 수집한 915명의 해양안전 제안 내용을 텍스트 마이닝 기법을 활용하여 분석하였다. TF-IDF 가중치를 적용한 키워드의 동시 출현 네트워크를 분석하고, 토픽 모델링으로 주제를 분류하였다.

‘교육’, ‘홍보’, ‘의식’ 등 텍스트 마이닝 기법을 활용해 분석한 해양안전 제안 내용을 18개의 설문 문항에 반영시켰으며, 수정·보완한 설문 문항은 2019년 해양안전문화지수를 도출하는 데 활용되었다. 그 결과 설문 문항의 내적 일관성을 나타내는 척도인 크론바흐의 알파 계수는 2019년 설문조사에서 0.895로 2018년 설문조사 0.888보다 0.007만큼 향상된 값으로 나타나 설문조사를 통해 도출하고 있는 해양안전문화지수의 신뢰성 향상에도 기여할 것으로 판단된다.

해양안전문화지수는 해양안전에 대한 국민의 경험과 인식을 객관적으로 측정하기 위한 지표로써 해양안전문화 확산을 위한 정책 자료로 활용될 수 있을 것으로 판단된다. 해양안전문화를 단기간에 확산시키기는 쉽지 않겠지만 해양안전문화지수의 대상자별, 영역별 비교를 통해 취약점을 확

인하고 이를 개선해 나간다면 해양안전문화를 정착해 나가는 데 도움이 될 것으로 기대한다.

References

- [1] AAA foundation for Traffic Safety(2019), 2018 Traffic Safety Culture Index, <http://www.aaafoundation.org/> (Accessed Nov. 2019).
- [2] Blei, D. M.(2012), Probabilistic Topic Models, Communications of the ACM, Vol. 55, pp. 77-84.
- [3] Blei, D. M., A. Y. Ng, and M. I. Jordan(2003), Latent Dirichlet Allocation, The Journal of machine Learning research, Vol. 3, pp. 993-1022.
- [4] Cronbach, L. J.(1951), Coefficient alpha and the internal structure of tests, Psychometrika, Vol. 16, No. 3, pp. 297-334.
- [5] Feldman, R. and I. Dagan(1995), Knowledge Discovery in Textual Databases(KDT), Knowledge Discovery, Vol. KDD-95, pp. 112-117.
- [6] Griffiths, T. L. and M. Steyvers(2004), Finding Scientific Topics, Proceedings of the National Academy of Sciences, Vol. 101, pp. 5228-5235.
- [7] Grun, B. and K. Hornik(2011), Topic models: An R Package for Fitting Topic Models, Journal of Statistical Software, Vol. 40, No. 13, pp. 1-30.
- [8] Hotho, A., A. Nürnberger, and G. Paaß(2005), A Brief Survey of Text Mining, LDV Forum - GLDV Journal for Computational Linguistics and Language Technology, Vol. 20, pp. 19-62.
- [9] IAM RoadSmart(2016), Measuring attitudes to driving safety & behaviour_The IAM RoadSmart Safety Culture Index, IAM Safety Culture Report, pp. 2-3.
- [10] Kang, S. Y., K. S. Kim, and B. S. Rho(2018), An Analysis of Causes of Marine Incidents at sea Using Big Data Thchnique, Journal of the Korean Society of Marine Environment & safety, Vol. 24, No. 4, pp. 408-414.
- [11] Kim, S. H., Y. J. Lee, J. Y. Shin, and K. Y. Park(2010), Text mining for Economic Analysis, Bank of Korea, Vol. 2019-18, p. 10.
- [12] Kim, Y. M.(2013), Study on Improving Safety Cultures by Analysing Behavior Characteristics of Korean Seafarers, Journal of the Korean Society of Marine Environment & safety, Vol. 19, No. 5, pp. 503-510.
- [13] KOSHA(2019), Korea Occupational Safety & Health Agency, Concept of safety culture, <http://www.kosha.or.kr/> (Accessed

Dec. 2019).

- [14] Lee, J. S., B. K. Lee, and I. S. Cho(2019), Text Mining Analysis Technique on ECDIS Accident Report, Journal of the Korean Society of Marine Environment & safety, Vol. 25, No. 4, pp. 405-412.
- [15] MOLIT(2019), Ministry of Land, Infrastructure and Transport, Transport Culture Index, <http://www.molit.go.kr/> (Accessed Dec. 2019).
- [16] Paek, Y. J., C. H. Jung, and D. G. Yoon(2018), The Development of Marine Safety Culture Index, Journal of Korean Maritime Police Science, Vol. 8. No. 4, pp. 79-106.
- [17] Park, J. H. and M. Song(2013), A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling, Journal of the Korean Society for Information Management, Vol. 30, No. 1, pp. 7-32.
- [18] Salton, G. and M. J. McGill(1983), Introduction to Modern Information Retrieval, McGraw-Hill, New York.
- [19] Seo, S. H.(2016), Fintech Trend Analysis using Topic Modeling of BM Patents, Graduate School of Seoul National University of Science and Technology.

Received : 2020. 01. 08.

Revised : 2020. 03. 09. (1st)

: 2020. 03. 30. (2nd)

Accepted : 2020. 04. 27.