

해양사고 인적오류 예방을 위한 핵심 주제어 분석에 관한 고찰

장은진* · 강유미** · † 임정빈

*해양경비안전교육원, **목포해양대학교 대학원, † 목포해양대학교 교수

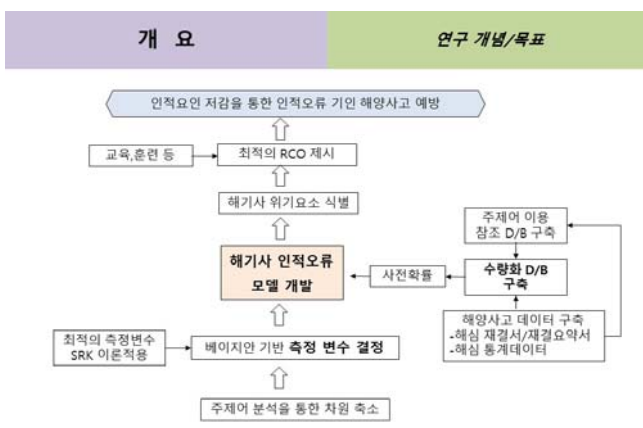
요 약 : 해양사고 원인의 대부분을 차지하는 인적오류 예방은 해양안전에 가장 중요하며 인적오류는 확률기반의 인적 모델을 구축하여 평가할 수 있다. 확률기반 인적 모델을 구축하기 위해 사건의 원인과 결과 사이에 연계성을 갖고 있는 통계 데이터가 필요하다. 이러한 데이터는 정부 공식통계로서 해양안전심판원에서 제공하는 재결서의 내용 분석을 통해 얻고자 하나, 측정변수가 너무 많아 계산량이 방대하다. 본 연구에서는 재결서 분석서의 원인판단에서 기준이 되는 해양안전심판원의 해양사고조사심판정보포털(이하 해심)에서 제공하는 재결서 내용의 핵심적인 내용으로 구성된 '주제어' 데이터를 활용하여 주제어에 포함된 핵심단어 분석절차를 수립하였다. 이들 단어가 구분형태별로 어떻게 분포된 상태인지 알아보고, 선박사고별로 최적으로 설명할 수 있는 단어 객체수를 검토해보고자 한다. 향후 축소된 차원으로도 해양사고 인적과실의 인과관계 설명이 가능하면, 인적모델의 측정변수를 결정하는 경우 쉽게 타당성을 확인 할 수 있어 해양안전을 위한 중요한 자료로 활용할 수 있다.

핵심용어 : 인적오류, 인적모델, 해양안전심판원 재결서, 주제어, 축소된 차원



개 요 **연구 내용**

- ❖ **인적오류 모델 (Human Error Model)** 은 특정 해기사가 해양사고를 유발할 수 있는 가능성을 측정하여 해당 해기사의 위기요소를 식별한 후, 위기요소제거 억제 등을 위해 교육이나 훈련시킬 내용을 제시
- ❖ 이에 확률기반 모델을 위해 개인정보, 선박정보 등 다양한 정보가 포함되어 있고 원인과 결과가 서로 연계되어 있는 데이터 확보가 필요 (수량화 D/B 구축)
- ❖ 또 측정변수를 전달하기 위해 해기사, 선박 등 방대한 정보를 축소하여 대표적인 몇 개의 측정변수로 요약할 위한 이론과 분석이 필요
- ❖ 본 연구에서는 해심에서 제공하는 해양사고조사심판정보 포털에 표시된 재결서 내용의 핵심적인 내용으로 고려되는 '주제어'를 활용 다음 사항을 검토하였다.
 - 수량화 D/B 구축을 위해 주제어 이용한 참조 데이터 구축
 - 측정변수를 S(Skill) R(Rule) K(Knowledge) 이론 적용 가능성 여부
 - 각 요소별 핵심단어 검토 및 차원 축소 가능 검토



참조 D/B 구축 **수집한 주제어 및 적용 주제어**

> 해심 포털 사이트에서 9가지 사고종류에 대한 총 3,012개의 주제어를 수집하여 이 중에서 의미 있는 주제어 총 2,642개 주제어 연구 대상 선정, 사고종류를 9가지 코드로 분류

번호	사건 종류	주제어 건수
1	충돌사건	559
2	집속사건	237
3	좌초사건	408
4	전복사건	237
5	화재/폭발사건	326
6	침몰사건	486
7	기관손상사건	532
8	입멸사상사건(별도)	304
9	기타(별도)	594
계		3683
각종 사상사건별 주제어 현황		
번호	사건 종류	주제어 건수
B01	선원사상사건	201
B02	여객사상사건	26
B03	작업원사상사건	22
B04	도선사상사건	5
B05	행방불명	50
계		304

번호	사건 종류	주제어 건수
901	화물손상사건	16
902	운항저해사건	72
903	혼란사건	203
904	안전저해사건	26
905	해양오염사건	3
906	추진기 손상사건	59
907	키 손상사건	24
908	타기 손상사건	12
909	선수문 탈락사건	4
910	시정물 손상사건	87
911	감위요원 사상사건	4
912	어망 손상사건	35
913	육구 손상사건	43
914	출선신 손상사건	6
계		594

Code	분류	연구에 사용한 주제어 수(개)
1	충돌사건	470
2	집속사건	237
3	좌초사건	374
4	전복사건	193
5	화재폭발사건	289
6	침몰사건	416
7	기관손상사건	503
8	선원사상사건	170
9	여객사상사건	26
	소계	2,642

† 교신저자 : 종신회원, jbyim@mmu.ac.kr
* 정회원, jej0412@korea.com
** 정회원, tntksd183@mmu.ac.kr

참조 D/B 구축

주제어 분류

> 2,642개의 주제어를 임의로 구분하고, 각 분류별 설명변수를 CODE화 하여 수량화 데이터로 변환



주제어 분석

분석 방법

❖ 분석 구분

- 주제어 빈도 분석: 인적요소와 분류요소 사이의 상대적인 분포(%)를 통해 인적요소가 분류요소와 어떠한 관계를 갖는지 양적 분석
- 핵심단어 분석: 인적요소가 분류방법별로 어떠한 내용을 갖는지 질적 분석

❖ 분석 이론

- 유클리디안(Euclidean distance) 거리: 객체의 빈도, 거리, 무게 등 어떠한 단위에 상관없이 3차원 공간에서 발생하는 객체 사이의 거리
- 클러스터(cluster) 분석: 분류형태별로 가장 크게 기여하는 소수 몇 개를 핵심단어 이용 (MATLAB 이용)
- Pareto 분석: 최적의 객체수 검토, 빈도수가 가장 많은 객체부터 적은 객체까지의 순서대로 빈도의 누적분포를 계산, 누적분포가 80% 이상 되는 객체수 결정 (80%는 신뢰수준, CL)

참조 D/B 구축

주제어 분류결과

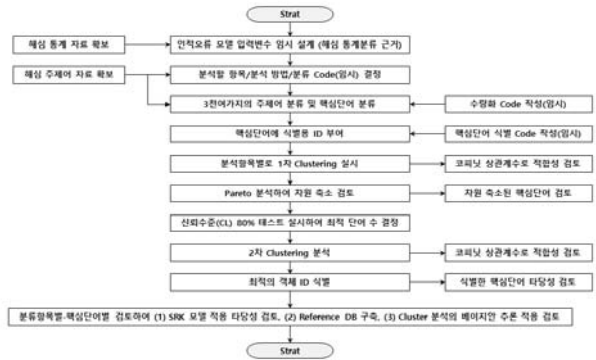
> 주제어를 분류한 결과는 다음과 같다. 우측 그림은 분류한 주제어를 일부 필요한 항목에 대해서만 숫자로 변환시킨 수량화 데이터를 나타낸다.



Maritime Risk Management Lab (MARIM), Abolpo National Maritime University, KOREA

주제어 분석

분석절차



참조 D/B 구축

대표핵심단어 분류

> 2,642개의 주제어 중 빈도가 10회 이상 나타난 핵심단어와 빈도가 10회 이하인 단어는 선정한 대표핵심단어 중 가장 유사한 단어에 포함(단, 10회 이하도 주요사고 대표할 수 있는 단어는 포함)하여 대표 핵심단어로 재분류

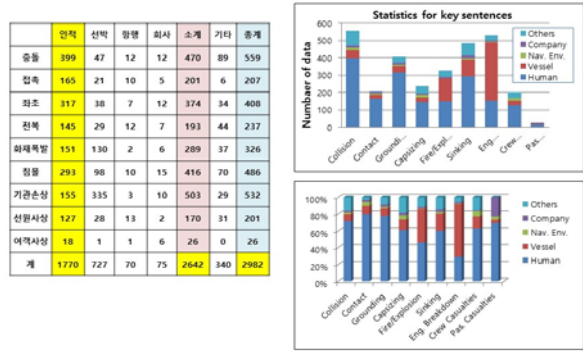
Code	Word	code	Number	Includings
11002	충돌	1	287	
11003	침몰	2	207	
11012	부착물	3	182	과적물
11041	부착물	4	173	
11042	부착물	5	118	
11007	불이행	6	96	
11004	불량	7	92	
21139	충돌	8	81	
11078	양식	9	71	
12001	과장	10	69	
21008	파손	11	62	
11038	미숙	12	52	중동9, 작오
11029	실수	13	47	12
11009	침착	14	44	
12006	노후	15	43	
12004	과다	16	41	
11006	태만	17	38	중동6
13001	기상	18	35	
11062	부족	19	34	
32029	실수	20	30	
11059	미흡	21	29	
11036	차이	22	29	
22009	과중	23	24	
11195	미인식	24	24	
51009	과열	25	23	
51009	과열	25	23	
11046	충돌	25	23	
21149	과열	27	22	
32023	불안	28	22	
11008	과다	29	22	과적물
32027	선박특성	30	21	
11001	불량	31	21	
51010	원래	32	20	
11014	과중	33	17	
31007	불량	34	17	
11020	미표시	35	17	
11106	의행	36	16	
72011	막로	37	15	
42010	과중	38	14	
52028	노후	39	14	
32006	과열	40	13	
12014	실수	41	13	
60019	부족	42	13	포함10
32017	불안	43	13	포함10
11015	태만	44	13	
52025	단단	45	12	
11023	미숙	46	12	
11096	과중	47	12	
11217	과중	48	12	
13004	미인식	49	12	제어방법(1), 불기, 발진, 과적물
32090	과열	50	11	
51027	충동	51	11	
11023	과다	52	11	
11023	과다	52	11	
11054	과중	54	8	
60017	미흡	55	7	
51061	불량	56	7	

자결서 분석 시 새로 항목원인 U-Rule 외부 식별

주제어 분석

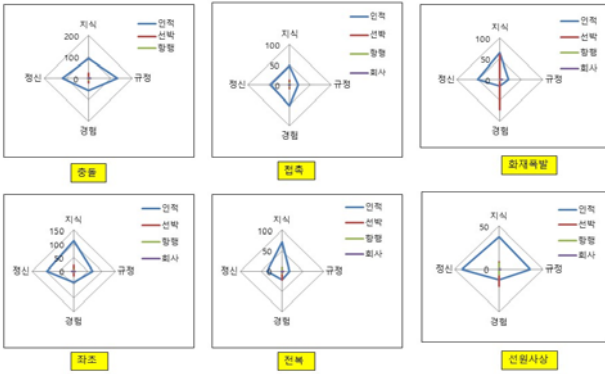
주제어 빈도 분석 결과 - 사고종류 대 분류요소

> 사고종류 vs. 분류요소에 대한 주제어 분류결과 인적요소에 관한 빈도가 가장 큼



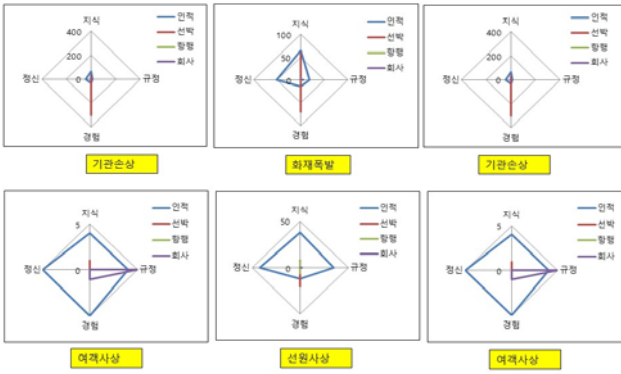
주제어 분석

주제어 빈도 분석 결과 - 분류요소 대 인적요소 분포



주제어 분석

주제어 빈도 분석 결과 - 분류요소 대 인적요소(SRK) 분포



주제어 분석

핵심단어 분석 1. 데이터 분류

사고종류별-분류요소별 핵심단어에 대한 클러스터(Cluster)분석에서 도출된 데이터 중에서 중동사고에서 SRKP의 인적요소에 해당된 56종의 핵심단어 빈도 수명

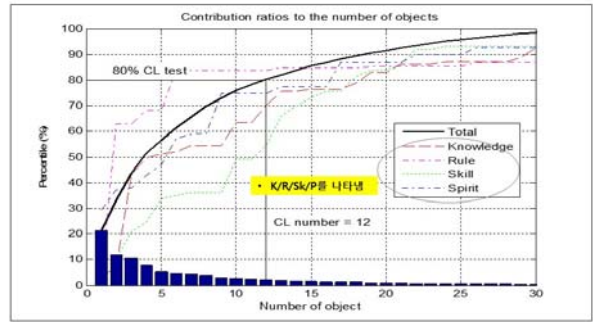
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
K	5	3	45	8	2	1	3	0	0	11	0	8	7	0	1	0	0	3	5	0	4	0	0	1	0
R	2	87	0	8	0	22	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
Sk	8	1	9	3	8	1	1	0	0	11	0	5	10	3	3	2	0	0	5	2	0	2	5	0	1
Sp	35	9	1	6	5	12	2	0	19	0	0	0	3	0	0	0	11	0	0	0	0	0	4	0	0

	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	
0	0	0	0	2	5	0	0	2	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	4	0	0	0	0	
2	0	0	0	0	6	0	0	0	10	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	4	0	0	0	0	
0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3	2	0	0	0	1	0	1	0	0

주제어 분석

핵심단어 분석 2. 최적 단어수 결정

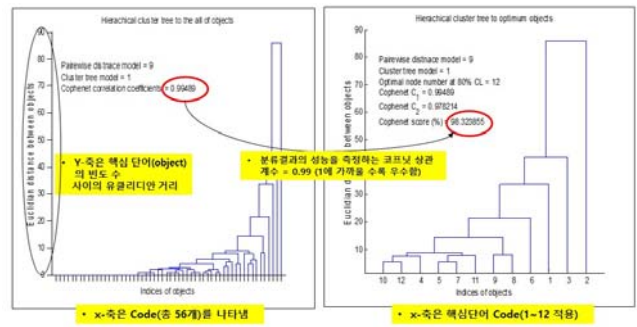
Pareto 분석을 활용, 전체 중동사고 핵심단어를 이용하여 80% CL에 해당하는 단어 수는 총 12개임



주제어 분석

핵심단어 분석 3. 주제어 차원 압축

중동사고의 80% 이상은 12개의 핵심 단어를 이용하면 설명할 수 있으며, 전체 객체를 이용한 경우(왼쪽 그림) 코브릿 계수가 0.99 있는데, 12개로 축소된 경우(오른쪽 그림) 0.98로 되어 약 98% 성능을 나타냄



주제어 분석

검토

- 클러스터 분석으로 57개 단어 모두를 이용하여 분석하는 경우 그 내용이 방대하여 해당사고를 최적으로 설명할 수 있는 단어 수 결정이 필요하며 최적의 객체 수 결정을 위해서 반복적인 계산이 필요
- Pareto 분석을 통해 최적의 객체수를 검토해 주제어 핵심단어 10개의 객체만을 이용하여 80% CL을 만족하나 다양한 객체 사이의 유클리디안 거리를 상호 비교하여 분석하는 클러스터 분석과 달리 누적분포만을 이용하므로 서로 식별한 객체가 다르게 됨
- 본 연구에서는 일단 전체 객체에 대한 누적분포에서 80% CL을 만족하는 객체수를 우선 정하고, 이를 클러스터 분석을 통해 가시화하는 방법을 적용하였으나 추후 사고종류별, 분류요소, 인적요소별 구분방법을 결정하여 최적의 주제어 차원 축소 실시
- 현재 분석한 주제어는 사고건수가 아니기 때문에 통계적인 의미는 없으나 각 사고별 주요요인을 Skill, Rule, Knowledge, Spirit로 분류할 있었으며, 축적점수를 SRK이론 적용이 가능

후기

본 논문은 해양수산부의 '해양안전사고 예방시스템 기반연구(2단계)'과제의 연구결과임을 밝힌다.