

텍스트 마이닝 기반의 미국 국방 표준 동향 분석을 통한 한국 국방 표준의 발전 방안 연구

채수환, 심보현, 염슬기, 홍성돈*
국방기술품질원 표준인증연구부 표준화연구1팀

A Study on the Development of Korean Defense Standards through Text Mining-Based Trend Analysis of United States Defense Standards

Soohwan Chae, Bohyun Shim, Seulki Yeom, Seongdon Hong*
Standardization Research Team 1, Standardization & Certification Research Division,
Defense Agency for Technology and Quality

요약 본 연구는 국방 분야 선진국인 미국의 표준에 대한 제정 동향을 파악하고, 한국 국방 표준에 적용 가능한 방안을 검토하였다. 이를 위해 웹에 등록된 MIL-STD를 비롯하여 다양한 미국 국방 문서에 대해 제목을 중심 데이터를 수집한 후, 텍스트 마이닝을 이용하여 단어 빈도를 분석하고 그 결과를 워드클라우드 형태로 생성하였다. 그 결과, 시대별로 MIL-STD에 등장하는 단어의 동향을 파악할 수 있었다. 문서 자체의 형식으로 인해 많이 등장하는 단어, 전 시대에 걸쳐 많이 등장하는 단어도 있는 반면, 과거에는 자주 쓰이다 현재는 많이 쓰이지 않는 단어나 과거에는 주목을 받지 못하다가 현재에 와서야 많이 등장하는 단어도 파악이 가능하였다. 또한 MIL-STD를 포함한 다양한 국방 문서를 대상으로 생성한 워드클라우드를 통해 그 특징을 도출하였다. 결론적으로 한국 국방 표준도 재료의 안전한 사용 및 다양한 화물 운반 기준을 마련하는 것에 대한 고민이 필요한 것을 확인하였다. 더 나아가 국방 분야 표준 및 규격에 4차 산업혁명 등과 관련된 최신 기술을 반영하여 표준을 선점하고 규격을 마련하는 것이 중요할 것이다. 또한 국방표준 문서체계를 명확하게 정립하고 효율적인 관리에 투자한다면, 국방 분야 표준의 질 향상을 기대할 수 있을 것이다.

Abstract This study examined the trend of standards established in the United States and to find points that can be applied to Korean defense standards. The titles of various United States defense standard documents registered on the web were selected for this research. The wordcloud was created after analyzing the frequency of words appearing in the title using text mining. The trend of words appearing in MIL-STD by era was obtained. This study identified words that appear often due to the format of the document itself, words that appear regularly throughout the era, words that are used frequently in the past but are not used much in the present, and words that did not receive attention in the past but appeared recurrently in the present. In addition, the characteristics of each document were derived through the wordcloud produced for various defense documents. In conclusion, Korean defense standards also require a consideration of safe and efficient management, transport, and load design of hazardous materials. Furthermore, the quality of defense standards can be expected to improve if the defense standard document system can be established, focusing on efficient management.

Keywords : Text Mining, Defense Standard, MIL-STD, Trend Analysis, Wordcloud

*Corresponding Author : Seongdon Hong(Defense Agency for Technology and Quality)

email: schong@dtaq.re.kr

Received December 29, 2020

Accepted March 5, 2021

Revised January 29, 2021

Published March 31, 2021

1. 서론

표준은 시대와 관계없이 강조되고 있으며, 4차 산업혁명 시대에 접어들면서 기술적 우위 및 시장 선점을 위해서도 표준을 활용하고 있다. 이와 함께 최근 유행하는 코로나19 대응을 위한 방역 상황에서 한국의 검사법, Drive Thru 진료 등이 국제적 신뢰를 받으며 유효성이 입증되고 있으며, 이러한 K-방역모델에 대해 ISO 국제 표준화를 추진하면서 자연스럽게 표준의 역할과 중요성이 주목받았다. 대한민국 정부에서도 2020년도 표준 혁신역량 강화에 3,258억 원을 투입하는 등 기업 성장과 삶의 질 제고에서 표준의 중요성을 높게 인식하고 있으며[1], 그 결과 국내에서도 이에 따라 연구 성과평가법 개정으로 표준을 국가의 연구 성과 유형에 추가하였다.

현재 국방규격은 연구개발 완료 단계에서 제정되는데, 규격 작성시점에는 최신 기술을 반영되었으나 운영유지 과정에서 기술 발전이 지속적으로 반영되지 못할 경우 부품단종과 같은 문제가 발생할 수 있다. 그러나 현재의 군수품 표준화 업무체계 속에서 제도과 인력의 한계로 민간의 기술을 신속하게 반영하는 것에는 제약이 따른다. 이를 개선하기 위하여 성능형 규격 작성 확대와 같은 정책 및 제도 발전을 지속 추진하고 있으며[2-4], 최근에는 우수한 민간기술의 국방 분야 적용을 확산하고 선진화된 국방 표준 제정하기 위해 민·군 기술을 협력하는 형태의 사업이 진행되고 있다[5]. 다만 이러한 정책이 대내외 환경 변화가 우선시 되는 경우가 종종 발생하고 체계적인 발전이 필요한 시점이기도 하다. 따라서 이를 개선하기 위해서는 국방 표준 관리 벤치마킹을 통해 중장기적인 로드맵을 수립하고 이를 실행해 나가는 전략이 필요하다.

한편, 연구 분야에 있어서 사회의 트렌드 분석을 통한 미래 과제를 도출하는데 텍스트 마이닝 기법이 자주 활용되는데, 품질경영학회지 전체에서 많이 사용되는 단어의 빈도수를 계산하여 한국 품질의 연구동향을 분석한 것이 그 사례 중 하나라고 할 수 있다[6]. 이 연구에서는 40년간 품질경영학회지 초록에서 자주 등장하는 단어를 추출하여 시대 변화에 따라 중요 단어가 변화하고 있음을 확인하고, 시대적으로 중시되는 내용을 파악하였다. 그 결과 기대와는 달리 최근 4차 산업혁명 단어들(AI, Big data, IoT 등)의 사용이 다소 미약한 것을 발견하였고 이를 바탕으로 한국 품질 연구의 보수성을 꼬집었다. 이 외에도 다양한 분야에서 텍스트 마이닝이 사용되었다. 항공 분야 논문 분석을 통해 미래 유망분야를 전망하였으며[7], 북한의 보도내용을 시계열 분석을 통해 도발과

의 연관성을 연구하기도 하였다[8]. 그리고 경제 분석을 위한 기법으로 텍스트 마이닝 방법론을 소개하였으며[9], 국내의 기사 중 코로나19와 관련된 온라인 교육에 대해 텍스트 마이닝 분석을 진행하였다[10]. 이와 같이 텍스트 마이닝 기법이 빠른 속도로 발전함에 따라 이를 활용한 연구가 주요 학술지에 게재되고 있다.

그러나 텍스트 마이닝을 표준과 연관시켜 진행한 연구는 많지 않았다. 표준 교육 관련 연구 동향을 분석한 연구가 있었지만[11], 이는 표준이나 표준화에 대한 고찰보다는 교육을 중심으로 다루었다. 또한 데이터 마이닝을 활용하여 한국산업표준(KS)의 활용을 분석한 연구도 진행되었다[12]. KS 보유 현황 및 제정 활동을 분석하고 열람실적을 통한 표준의 활용도까지 확인하는 등 다양한 데이터를 활용하여 인상적이었지만, 텍스트 마이닝을 직접적으로 이용하지는 않았다. 특히 국방 표준 분야에서는 텍스트 마이닝 관련 연구가 부재한 실정이다.

따라서 본 연구에서는 국방 표준화 업무에 대한 정책적인 로드맵 수립이나 연구 방향의 설정을 위해 텍스트 마이닝 기법을 활용하고자 하였다. 이를 위해 미국의 국방표준인 MIL-STD의 제목을 연도별로 추출하여 자주 등장하는 단어의 추세를 시계열 기법 등을 통해 분석하고자 한다. 이러한 분석 결과로부터 과거에서 현재까지 표준의 변화양상을 분류하고 앞으로 관심이 필요한 분야를 찾고자 하였다. MIL-STD와 더불어 미군의 표준체계에서 함께 활용되고 있는 MIL-HDBK, MIL-PRF, MIL-DTL에 대해서도 텍스트 마이닝을 활용하여 대한민국 국방 표준의 발전 방향에 대해 제시하고자 한다.

2. 본론

2.1 이론적 배경

2.1.1 텍스트 마이닝 개요

텍스트 마이닝은 비정형 데이터인 텍스트로 이루어진 자료에서 유용한 정보를 추출하는 기법을 말한다[13]. 여기서 비정형 데이터란 텍스트, 그림, 영상과 같이 그 자체로는 컴퓨터가 처리할 수 없는 데이터이다. 텍스트 마이닝은 이러한 데이터를 전처리하여 의미 있는 단어를 식별하고 문서와 단어 사이의 상관관계를 도출하는 알고리즘이다[9].

2.1.2 텍스트 마이닝 방법론

여러 텍스트 마이닝 방법 중 대표적인 네 가지에 대해서 조사하였다. 첫 번째는 BOW(Bag of Words) 기법으로, 단어의 순서는 고려하지 않고 단어의 빈도수에 중점을 두는 표현 방법이다. 주로 어떤 단어가 얼마나 자주 등장하였는지 파악하여 해당 문서의 주제나 중요한 단어 등을 판단할 수 있다[14]. 이러한 BOW는 다루기 쉬운 장점이 있으나, 단어의 빈도수만 확인하기 때문에 문장 내에서의 의미 파악이 제한된다는 단점이 있다.

두 번째는 DTM(Document Term Matrix) 기법으로, BOW가 특정 문서에 등장하는 각 단어들의 빈도를 나타낸 행렬이라면, DTM은 그 문서들의 행렬을 합친 전체 문서와 단어에 대한 행렬이다. 이를 통해 전체 문서의 정보를 파악할 수 있어, 여러 문서를 한 번에 분석할 수 있는 장점이 있다. 그러나 한 번에 여러 문서와 그 문서에 있는 모든 단어를 대상으로 행렬을 만들다 보니 저장 공간을 낭비하게 된다는 단점이 있다. 따라서 행렬의 크기를 줄이기 위해 의미가 있는 단어만 추출하거나 어간 추출, 불용어 처리 등의 전처리 작업을 실시하게 된다.

세 번째는 TF-IDF(Term Frequency-Inverse Document Frequency) 기법으로, 단어의 상대적 중요도에 따라 가중치를 부여하는 방식이다[15,16]. 이는 빈도가 높은 단어라도 모든 문서에서 흔하게 나타나는 단어라면 그 정보의 가치를 낮다고 판단하여 낮은 가중치를 부여하게 된다. 대표적으로 the, a, is 등의 단어와 같이 문서에서 별 의미는 없지만 흔하게 등장하는 단어의 TF-IDF 값은 낮게 계산이 된다[17]. 이는 주로 검색 결과의 중요도 결정이나 문서의 유사도 확인 등에 사용된다.

Table 1. Explanation of Text Mining Methods

Method	Explanation
BOW (Bag of Words)	<ul style="list-style-type: none"> Frequency-based word expression method Numericalize text data using frequency of appearance without considering order
DTM (Document Term Matrix)	<ul style="list-style-type: none"> A combination of BOWs of different documents A matrix of the frequency of words appearing for each document
TF-IDF (Term Frequency-Inverse Document Frequency)	<ul style="list-style-type: none"> Statistical figures indicating how important a word is within a particular document in a multi-documents Low weight, if commonly found in all documents
Wordcloud	<ul style="list-style-type: none"> Visualized in the form of a cloud of word so that you can intuitively understand keywords, topics, concepts, etc. of documents A word that appears frequently is expressed in large words

마지막으로는 워드클라우드(Wordcloud) 기법이다. 워드클라우드는 빈도수에 따라 크기가 다른 단어들을 구름 모양으로 표현한 것으로, 어떤 단어가 주로 사용되었는지 한 눈에 파악할 수 있다[18]. 일반적으로 많이 사용되는 단어일수록 글자 크기가 크고 가운데에 배치된다. 반면 적게 사용되는 단어일수록 글자 크기가 작고 바깥쪽에 배치된다. 지금까지 설명한 텍스트 마이닝 방법들을 Table 1에 정리하였다.

2.2 연구 방법

2.2.1 데이터 추출

미국에서는 상호 운용성을 향상시키기 위해 다양한 군사규격을 개발하여 관리하고 있는데, Table 2에서 보는 바와 같이 국방표준(Defense Standard, MIL-STD), 국방규격(Defense Specification, MIL-SPEC), 국방핸드북(Defense Handbook, MIL-HDBK), 성능형 규격(Performance Specification, MIL-PRF), 상세형 규격(Detail Specification, MIL-DTL) 등이 대표적인 사례이다[19].

Table 2. Classification and Explanation of United States Military Standards

Type	Explanation
Defense Standard (MIL-STD)	A document that establishes uniform engineering and technical requirements for military-unique or substantially modified commercial processes, procedures, practices, and methods.
Defense Specification (MIL-SPEC)	A document that describes the essential technical requirements for military-unique materiel or substantially modified commercial items. Performance specifications are preferred over detail specifications.
Defense Handbook (MIL-HDBK)	A document that provides standard procedural, technical, engineering, or design information about the materiel, processes, practices, and methods.
Performance Specification (MIL-PRF)	States requirements in terms of the required results with criteria for verifying compliance but without stating the methods for achieving the required results. A performance specification defines the functional requirements for the item, the environment in which it must operate, and interface and interchangeability characteristics.
Detail Specification (MIL-DTL)	A specification that states design requirements, such as materials to be used, how a requirement is to be achieved, or how an item is to be fabricated or constructed.

국방표준은 군에서 활용되는 공정, 절차, 방법에 대한 통일된 공학적, 기술적 요구 사항을 설정하는 문서이며, 국방규격은 군수품에 대한 필수적인 기술적 요구 사항을 설명하는 문서이고, 국방핸드북은 장비나 공정, 방법에 대한 표준 절차나 기술, 공학, 설계 정보를 제공하는 문서이다. 여기서 국방규격은 성능형과 상세형 규격으로 나눌 수 있는데, 성능형 규격은 요구 사항의 달성 방법은 명시하지 않고, 결과적인 관점에서 성능을 검증하기 위한 기준과 기능적 요구 사항, 작동 환경, 상호 운용성을 기술한 문서이며, 상세형 규격은 사용할 재료, 요구 사항의 달성 방법, 제작 및 구성 방법과 같은 상세한 요구 사항을 설명한 문서이다.

본 연구에서는 Table 2에서 소개한 규격 문서의 제목을 텍스트 데이터로 사용하였다. 이를 위하여 ASSIST 사이트의 Quick Search (<https://quicksearch.dla.mil>)를 이용하여 제정 날짜, 상태, 제목 등의 정보를 추출하였다. 규격의 본문에 대해서 텍스트 마이닝을 진행하면 보다 정확한 데이터 분석이 가능하지만, 본문 자료 확보가 제한적이고, 자료가 확보되었다고 하더라도 텍스트 데이터의 추출이 어려웠다. 따라서 이번 연구에서는 우선 제목에 한정하여 분석하기로 결정하였다. 이때 사용된 데이터의 양은 Table 3과 같다.

Table 3. Number of MIL-STD, MIL-HDBK, MIL-PRF, and MIL-DTL according to the Issued Year

Year	MIL-STD	MIL-HDBK	MIL-PRF	MIL-DTL
2010s	626	300	2,580	5,243
2000s	223	122	283	447
1990s	1,016	119	79	206
1980s	217	46	5	0
1970s	94	21	1	0
1960s	130	6	0	0
1950s	120	5	0	0
1940s	8	0	0	0
Total	2,434	619	2,948	5,896

2.2.2 연구 과정 설계

앞서 조사한 텍스트 마이닝 방법론 및 선행연구를 바탕으로 본 연구에서는 Fig. 1과 같이 분석 과정을 설계하였다. 먼저 비정형 데이터인 텍스트를 ASSIST 사이트에서 수집한 뒤 제목을 추출하였다. 이를 다시 종류 및 연도별로 분류하여 텍스트 마이닝을 거쳐 단어의 빈도수를 계산하였으며, 이를 바탕으로 각 문서별로 워드클라우드를 생성하였다. 그리고 발생 빈도가 높은 단어를 선별하여 시대별 추세를 분석하였다. 과거와 현재에 쓰이는 단

어의 변화를 통해서 한국군의 표준에 대한 미래 발전 방안을 고민해 보았다.

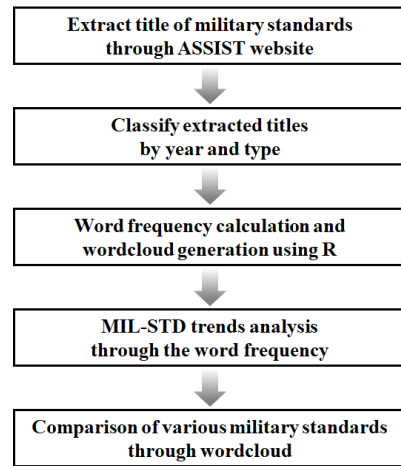


Fig. 1. Schematic Diagram of the Research Process

2.2.3 텍스트 마이닝

텍스트 마이닝을 위해 R(R version 4.0.0) 소프트웨어를 사용하였으며, 최신 R 소프트웨어는 <https://www.r-project.org>에서 다운받을 수 있다. R은 오픈 소스 생태계를 가지고 있어 다양한 분석 툴과 최신 분석 기법을 쉽게 사용할 수 있어서 사용하게 되었다. 또한 크게 복잡하지 않은 분석의 경우에는 일반적인 컴퓨터로도 충분히 수행이 가능하다. 이번 연구에서는 가정용 노트북을 활용하였으며, tm, SnowballC, wordcloud, RColorBrewer 패키지를 사용하였다. 여기서 tm은 영문 텍스트 마이닝을 위하여 설치한 패키지이며, SnowballC는 텍스트 마이닝 과정 중 어근을 추출하는 과정에서 사용하였다[20,21]. 나머지 패키지인 wordcloud와 RColorBrewer은 빈도 분석을 바탕으로 워드클라우드를 생성하는 데에 활용하였다. 실제 활용한 코드는 다음과 같다.

2.3 연구 결과

2.3.1 MIL-STD의 시대에 따른 분석

Table 3의 MIL-STD 제목으로부터 각 연도별 노출되는 빈도수를 계산한 뒤 일정 횟수 이상 등장한 단어를 선별하여 분석하였다. 이때 일반적으로는 각 빈도수 및 TF-IDF 분석으로부터 핵심 단어를 파악하게 되지만 [10], 본 연구는 데이터를 문서 제목으로 한정함에 따라

데이터의 양이 부족하여 빈도수 기준 상위 20개의 단어에 대한 분석을 진행하였다.

```
# Package installation
install.packages("tm")
install.packages("SnowballC")
install.packages("wordcloud")
install.packages("RColorBrewer")

# Load package
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")

# Import text
text <- readLines("file_name.txt")
docs <- Corpus(VectorSource(text))

# Text refining (spaces, numbers, punctuation removed, root
extraction, etc.)
docs <- tm_map(docs, stripWhitespace)
docs <- tm_map(docs, tolower)
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removeWords, stopwords("english"))
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stemDocument)

# Check frequency and print top 100 words
tdm <- TermDocumentMatrix(docs)
m <- as.matrix(tdm)
v <- sort(rowSums(m), decreasing = TRUE)
d <- data.frame(word = names(v), freq = v)
head(d, 100)

# Visualization with wordcloud
wordcloud(words = d$word,
          freq = d$freq,
          min.freq = 20,
          max.words = 100,
          random.order = FALSE,
          rot.per = 0.1,
          colors = brewer.pal(8, "Dark2"))
```

먼저 시대에 관계없이 빈도수를 기준으로 상위 20개 단어를 Table 4에 노출 순서대로 정리하였다. 여기에서 보면 Document가 가장 많이 등장하였으며, Load와 Mod, Unit이 그 뒤를 이었다. 이 단어들은 총 2,434개의 제목 중에서 500번 이상 등장하였다. 이는 과거와 현재에 이르기까지 표준을 문서(Document)로 표현하면서 나타나는 결과로 볼 수 있으며, Load는 탄약 및 폭발물의 안전하고 효율적인 운반을 위한 표준에서 자주 등장한다. Mod는 Modified의 약자로 제정된 표준을 개정하거나, 다른 표준을 활용하면서 발생하는 단어로 볼 수 있다. 또한 Unit은 표준이 하나의 체계로부터 구성품 단위

로 분류되어 추가로 제정되면서, 이를 표현하는 용어로 볼 수 있어 대부분의 상위에 포진된 단어가 표준의 기본 구성요소로 볼 수 있다.

Table 4. MIL-STD Word Frequency Ranking over the Whole Period

Ranking	Word	Frequency
1	Document	859
2	Load	775
3	Mod	551
4	Unit	526
5	Container	328
6	Standard	304
7	System	290
8	Milstd	238
9	Test	234
10	Domestic	230
11	Pallet	195
12	Part	194
13	Equipment	172
14	Materials	169
15	Requirements	165
16	Method	162
17	Information	160
18	Bomb	150
19	Replenishment	147
20	Underway	145

위에서 전 시기를 대상으로 살펴보았다면, 이번에는 10년 단위로 구분하여 그 빈도를 관찰하였다. 중요한 단어를 선별하여 그 주요 단어들을 대상으로 연도별 등장 현황을 그래프로 정리하였다(Fig. 2). 이때 단순히 노출되는 단어 수로 나타내면 시대에 따라 수집한 표준의 수에 차이가 있어 상호 간 비교 시 오류가 있을 수 있다. 따라서 각 단어의 등장 횟수를 해당 연도의 총 표준 수로 나누었으며, 이는 Fig. 2의 세로축에 해당한다. Fig. 2(a)에서 단어의 노출빈도가 시대에 따라 변화되는 것을 볼 수 있는데, 이를 각 유형별로 다시 분류하여 Fig. 2(b)로 표현하였다. Fig. 2(b)의 그래프 개형을 살펴보았을 때, 단어의 노출 유형을 크게 4가지 정도로 구분할 수 있다.

여기서 첫 번째 유형(①)은 과거에 비해 현재에 그 비중이 높지 않은 경우이며, 대표적으로 Test를 들 수 있다. 그러나 실제로는 Test가 포함된 표준 수가 줄어든 것이 아니고, 1980년대 이후 표준 생성이 급격히 증가하면서 전체 표준에서 차지하는 비중이 상대적으로 줄어든 경우이다. 이는 제정 또는 수정되는 표준 수가 증가함에 따른 부수적인 결과이며 그 중요도 자체가 감소한 것은 아니라고 볼 수 있다.

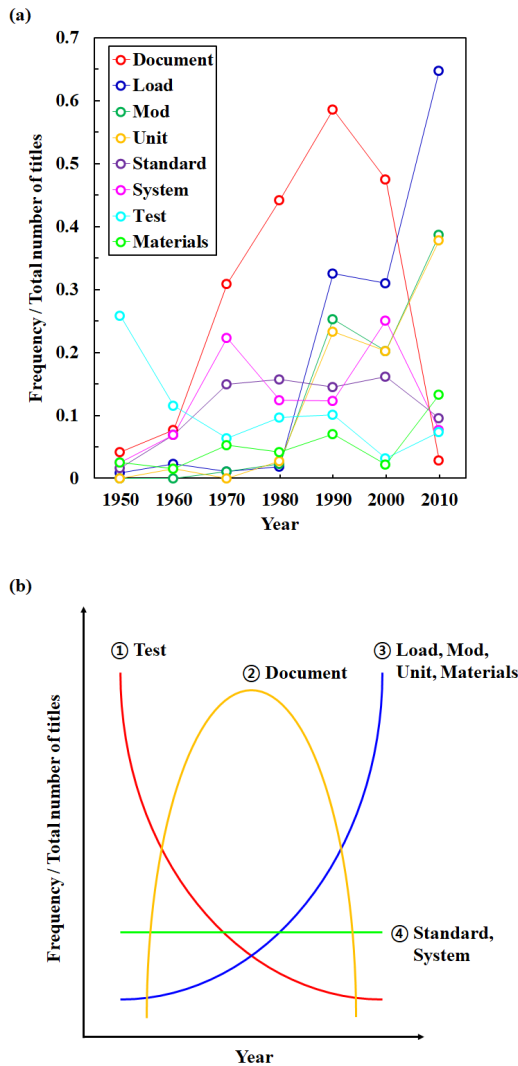


Fig. 2. (a) Graph of Words Frequency Ratio according to the Issued Year, (b) Simplified Graph of Words Frequency Ratio according to the Issued Year and Classification by Four Types

반면 두 번째 유형(②)은 과거와 현재에는 많이 사용되지 않았지만 그 중간 시기에 많이 사용된 단어 유형이다. 대표적인 예로 Document가 있는데, 이는 제목에서 특정 무기 체계와 관련된 문서라는 의미로 사용되는 경우도 있지만, 많은 경우 괄호 안에 No S/S Document 라는 문구로 등장한다. 이는 취소된 문서에 대해서 대체하는 문서가 없다는 의미이다. 만약 대체할 문서가 있다면 제목 뒤에 “S/S by MIL-000” 형태의 문구가 추가된다. 이는 취소된 문서이며 MIL-000로 대체되었다는 의미이

다. 결과적으로 “No S/S Document”를 제목에 포함하는 문서는 이미 취소된 표준이며 이를 대체할 표준이 따로 제정되지 않았다는 의미를 갖는다. 따라서 1970년대부터 2000년대에 제정된 표준은 취소된 비율이 높아 이러한 문구가 많이 쓰이게 된 것이다. 반면 2010년대에는 이 문구가 붙은 표준이 거의 없는데, 이는 최근에 제정된 표준이라 그 대부분이 취소되지 않은 상태이기 때문이다.

세 번째 유형(③)은 과거에는 많이 사용되지 않았지만 최근에 많이 사용되는 단어 유형이며 Load, Unit, Mod, Materials가 대표적인 예다. Load는 Loading, Truckloading 혹은 Carloading 등과 같은 단어로도 사용되거나 Unit과 함께 결합되어 사용되며, 이는 화물 운송과 관련된 단어이다. 이와 관련된 표준인 MIL-STD-1320D, For Designing Unit Loads, Truckloads, Railcar Loads, and Intermodal Loads for Ammunition and Explosives를 통해 탄약 및 폭발물의 안전하고 효율적인 운반을 위해 수많은 표준이 제정되었음을 알 수 있다. 이러한 단어의 비중이 증가하게 된 것은 미군이 현대에 들어서 군수에 대한 중요성을 재인식하였기 때문이다. 따라서 이러한 장비나 물자의 보관 방법 등에 중점을 두고 이동수단에 따른 적재나 포장 등을 표준으로 만들게 된 것이다. Materials는 2010년대부터 Hazardous Materials로 묶인 한 단어로 자주 등장하는데, 이는 모든 품목에 대한 인체 유해성이나 안전성 확보를 위한 것이며 민간 분야의 제품에 대한 요구 사항과도 일치한다. 또한 앞서 살펴보았던 운반과도 결합되어 위험한 물질의 안전한 운반 방법 설계가 강조되었다. 이와 같은 단어들은 한국군의 표준체계에도 시사하는 바가 크다고 할 수 있다. 그리고 Mod는 1990~2010년대까지 활발하게 사용되고 있는데, 장비의 수정 또는 변형을 나타낸다. 각 장비에 대해 번호 순서로 할당되며 초기 버전은 Mod 0으로 지정된다. 이와 관련된 표준인 MIL-STD-1661, Mark and Mod Nomenclature System이 1978년도에 제정된 바 있다. 즉, 1970년대 후반에 Mod에 대한 표준이 제정되어 1990년대 이후에 활발하게 사용되는 것을 확인할 수 있었다. Mod 0부터 시작해서 Mod 1이나 Mod 2가 많이 쓰이며 더러 Mod 7이 등장하기도 한다. 따라서 특별한 의미보다는 표준을 지속적으로 관리하고 있다는 의미로 판단하는 것이 합리적이라 생각된다.

네 번째 유형(④)은 과거와 현재 모두 꾸준히 사용되는 단어 유형이다. 대표적으로 Standard와 System이 있는데, 시대의 변화에 관계없이 지속적으로 사용되고 있다.

텍스트 마이닝을 위한 기본 데이터가 미국의 국방 표준이므로 Standard가 일정한 비중으로 노출되는 것이며, 체계를 의미하는 System은 표준 자체의 시스템을 정의하는 문서에서 사용되기 때문에 자주 등장한다. 두 단어 모두 전 시기에 걸쳐 꾸준히 등장하는 것으로 보았을 때, 과거부터 지금까지 국방표준관련 문서에서 표준과 시스템이 중요한 개념으로 쓰이는 것을 확인할 수 있었다.

2.3.2 국방 표준 문서 종류별 비교

앞서 MIL-STD를 대상으로 시대별 분석결과 4가지 유형을 알 수 있었으며, 표준 영역에 있어서 주요 관심사를 도출할 수 있었다. 이에 대하여 표준의 형태별로 차이가 있는지 확인하기 위하여 추가적인 분석을 진행하였다. 먼저 모든 문서 종류(MIL-STD, MIL-HDBK, MIL-PRF, MIL-DTL)에 대해 텍스트 마이닝을 실시하여 Table 5에 정리하였다.

그 결과 Type이 가장 많이 사용되었으며, Document, Electrical, Connector가 그 뒤를 이었다. 이 단어들은 총 11,897개의 문서 제목 중에서 1,000회 이상 등장하였다. 조금 더 상세히 비교하기 위하여 각 문서 형태별로 빈도를 분석하고 어떤 단어가 많이 등장하는지 관찰하였으며, 이를 워드클라우드로 변환하여 한 눈에 알아보기 쉽게 표현하였다(Fig. 3).

Table 5. Military Standards Word Frequency Ranking over the Whole Period

Ranking	Word	Frequency
1	Type	1750
2	Document	1459
3	Electrical	1069
4	Connector	1065
5	Service	887
6	Unit	696
7	Load	693
8	Ribbon	612
9	Power	609
10	General	580
11	Style	566
12	Medal	563
13	Cable	509
14	Reliability	462
15	Air	453
16	Button	453
17	Frequency	446
18	Mod	435
19	Series	429
20	Lapel	417

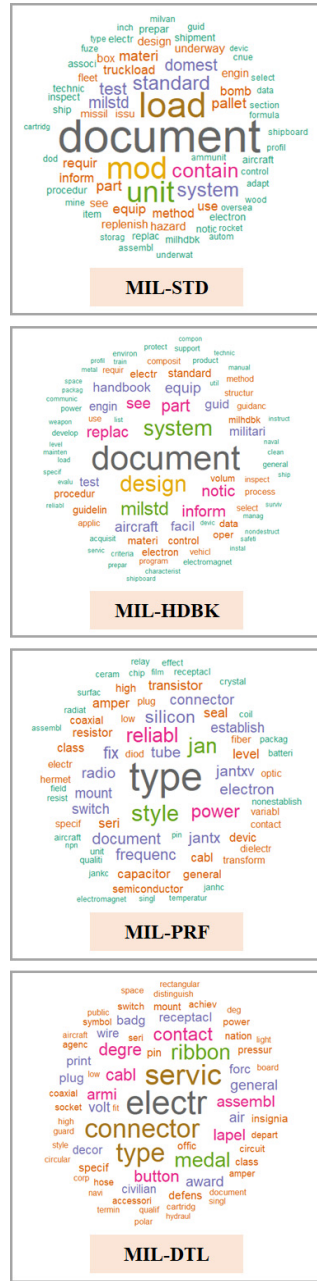


Fig. 3. Wordclouds by Military Standard Types

MIL-STD는 앞서 전체 단어뿐만 아니라 시대에 따른 분석을 진행한 바 있다. MIL-HDBK의 경우 Design과 System의 사용이 두드러진다. 절차나 장비의 정보 등을 알 수 있는 문서이기 때문에 그 설계나 체계를 뜻하는 단어가 자주 사용되는 것이다. 성능형 규격인 MIL-PRF에는 Type과 Style이 많이 등장하는데, 이는 MIL-STD에

서 사용되었던 Mod와 성격이 비슷하다. 특정 장비 타입을 명시하거나 스타일 번호를 지정하여 성능을 달성하는 형식 또는 방식을 규정하는 단어이다. Reliability도 자주 사용되는데, 이는 요구조건 성능 검증을 위해서 신뢰성을 중요하게 여긴다는 해석이 가능하다. 다른 문서들과 가장 눈에 띄게 차이가 나는 문서는 MIL-DTL이다. 상세형 규격인 만큼 다른 문서들에 비해서 장비나 체계를 특정하는 단어가 많이 등장하였다. 다른 문서에서는 잘 등장하지 않던 Electrical이나 Connector, Button, Ribbon 등 특정 장비를 규정하는 단어가 많다. 이를 통해 상세형 규격이 다른 문서들에 비해서 요구사항을 보다 상세하게 기술하는 것을 확인하였다. 이와 같이 제목에 등장하는 단어를 통해서 각 문서마다 그 특성에 부합하는 단어들끼리 쓰이는 것을 검증할 수 있었다.

앞서 살펴보았듯이 미국 국방 표준의 경우, 국방 표준의 유형을 다양하게 분류하고 효율적으로 관리하고 있다. 하지만 현재 한국 국방 표준은 상세형 규격 위주로 제정 및 관리되고 있다. 또한 국방규격과는 별개로 국방표준서가 존재하긴 하지만 이를 제정하고 관리하는 사업을 최근에 들어서 시작하였다. 2019년 10월에 최초로 국방표준서가 제정되었으며 2021년 1월 현재 145종이 등록되어 있다. 국방표준서의 성격에 맞는 국방규격을 우선 활용하여 편집 및 제정되고 있지만 아직 초기 단계이다. 또한 성능형 규격도 따로 관리하거나 분류하는 것도 아직 활성화되지 않고 있다. 따라서 우리나라의 국방 표준 문서체계를 보다 명확히 정립하고 관리할 필요가 있다. 또한 국방규격 위주, 그 중에서도 상세형 규격 위주의 표준 관리에서 벗어나야 한다. 전력에 영향이 없는 상세 요구조건을 과도하게 명시하여 업체의 자율성 및 경쟁력을 제한한다면 방위산업 전반의 질 저하가 발생할 수 있다. 따라서 국방표준서 및 성능형 규격을 활성화하고 관리하는 등 보다 효율적인 표준 관리에 투자할 필요가 있을 것이다.

3. 결론

국방 선진국인 미국의 국방 표준화 동향을 텍스트 마이닝을 통해 분석해보고자 하였다. 먼저 MIL-STD의 제목을 대상으로 1950년대부터 2010년대에 걸쳐 등장하는 단어 동향을 살펴보고 유형별로 분류해 보았다. 이를 통해 우리나라의 국방 표준도 탄약 및 폭발물 등 위험한 군수품을 안전하고 효율적으로 운반하는 표준에 대해 더

욱 고민하여야 할 것이다. 또한 인체나 환경에 유해한 재료를 관리하는 것에도 관심을 가져야 할 필요성을 느꼈다. 반면 미국 국방 문서에서도 아직 4차 산업혁명 시대에 기대되는 단어의 사용은 찾지 힘들었다. 이는 기술 발전 속도에 비해 표준의 제정이 뒤처지고 있다는 것을 의미한다. 혹은 군사력에 영향을 끼치는 최신 표준으로 인식되어 아직 ASSIST에 공개되지 않았을 수도 있다. 미국의 표준 제정의 지연 혹은 그 표준의 공개 여부와는 상관없이 우리나라 국방 분야에서 4차 산업혁명 등과 관련된 최신 기술을 적용하여 표준을 선점하고 규격을 마련하는 것이 중요할 것이다.

MIL-STD뿐만 아니라 다른 국방 표준 문서에 대해서도 텍스트 마이닝을 진행하였다. 이를 통해 미국은 표준화 추진을 위해 다양한 문서체계(MIL-STD, MIL-HDBK, MIL-PRF, MIL-DTL 등)를 갖추고 있으며 명확한 분리 및 관리가 되는 것을 확인하였다. 우리나라의 경우, 국방 표준서의 관리 및 운용이 아직 활성화되지 않고 있다. 따라서 우리나라도 국방표준 문서체계를 명확하게 재정립하고 효율적인 관리에 투자한다면, 국방 분야 표준의 질 향상을 기대할 수 있을 것이다.

다만 연구를 진행하면서 표준 문서의 본문에 대해서 진행을 하지 못한 것이 아쉬웠다. 제목을 대상으로 분석을 하다 보니 데이터의 절대적인 양이 부족하였으며, 파악할 수 있는 동향에 한계가 있었다. 또한 운반이나 화물 관련 표준이 많이 등장하는데, 이는 무기 체계에 비해 전력 노출이 희박하기 때문이라 생각된다. 비공개된 규격도 데이터를 확보할 수 있다면 보다 명확한 동향 분석이 가능할 것이다. 그리고 국내 규격이나 표준에 대해서도 분석을 진행한다면 보다 정확한 비교를 기대할 수 있다. 따라서 앞으로 표준서의 본문, 비공개된 표준, 국내 국방 표준에 대한 텍스트 마이닝 분석이 후속연구로 진행된다면, 더 많은 데이터를 바탕으로 보다 명확하고 유용한 결과를 낼 수 있을 것이다.

References

- [1] S. W. Lee, S-LIFE Vol.184, p.70, Korean Agency for Technology and Standards, 2020, pp.2-11
- [2] C. S. Kim, Y. M. Lee, S. H. Lee, D. S. Seo, "A Study on Activation Plan of Performance Military Specification", *Military Forum*, Vol.79, pp.164-184, 2014.
- [3] S. C. Choi, Y. H. Song, "A Study on the Development of Defense Performance Specification and Conversion

- into it”, *Journal of the Military Operations Research Society of Korea*, Vol.33, No.1, pp.135-155, 2007.
- [4] H. I. Shin, Y. J. Kim, “A Study on the Improvement of the policies and systems related to Performance Specification”, *Journal of the KIMST*, Vol.4, No.2, pp.97-104, 2001.
- [5] C. Yi, Defense Standardization Guidebook, p.94, Defense Agency for Technology and Quality, 2020.
- [6] S. Ree, “Analysis of Research Trends in Journal of Korean Society for Quality Management by Text Mining Processing”, *Journal of the Korean Society for Quality Management*, Vol.47, No.3, pp.597-613, 2019. DOI: <https://doi.org/10.7469/JKSQM.2019.47.3.597>
- [7] H. Kim, N. Jo, K. Shin, “Text Mining-Based Emerging Trend Analysis for the Aviation Industry”, *Journal of Intelligence and Information Systems*, Vol.21, No.1, pp.65-82, 2015. DOI: <https://doi.org/10.13088/IIIS.2015.21.1.65>
- [8] C. Lee, H. Moon, “Study on analysis of North Korea’s news trends associated with provocations using text mining”, *Journal of National Defense Studies*, Vol.59, No.4, pp.103-124, 2016.
- [9] S. Kim, Y. J. Lee, J. Shin, K. Y. Park, “Text Mining for Economic Analysis”, *SSRN Electronic Journal*, 2019. DOI: <http://dx.doi.org/10.2139/ssrn.3405781>
- [10] S. M. Kim, “Analysis of Press Articles in Korean Media on Online Education related to COVID-19”, *Journal of Digital Contents Society*, Vol.21, No.6, pp.1091-1100, 2020. DOI: <http://dx.doi.org/10.9728/DCS.2020.21.6.1091>
- [11] I. G. Go, “Big Data Analysis of Standards and Standards Education by Text Mining”, *Journal of Standards, Certification and Safety*, Vol.8, No.4, pp.19-32, 2018.
- [12] J. Yoon, W. Kim, H. Lee, “Analysis of Enactment and Utilization of Korean Industrial Standards(KS) by Time Series Data Mining”, *Journal of Technology Innovation*, Vol.23, No.3, pp.225-253, 2015. DOI: <http://dx.doi.org/10.14386/SIME.2015.23.3.225>
- [13] R. Feldman, I. Dagan, “Knowledge Discovery in Textual Databases (KDT)”, *KDD’95: Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, AAAI Press, pp.112-117, 1995.
- [14] Y. Zhang, R. Jin, Z. H. Zhou, “Understanding bag-of-words model: a statistical framework”, *International Journal of Machine Learning and Cybernetics*, Vol.1, No.1-4, pp.43-52, 2010. DOI: <http://dx.doi.org/10.1007/S13042-010-0001-0>
- [15] A. Aizawa, “An information-theoretic perspective of tf-idf measures”, *Information Processing & Management*, Vol.39, No.1, pp.45-65, 2003. DOI: [http://dx.doi.org/10.1016/S0306-4573\(02\)00021-3](http://dx.doi.org/10.1016/S0306-4573(02)00021-3)
- [16] G. Salton, C. Buckley, “Term-weighting approaches in automatic text retrieval”, *Information processing & management*, Vol.24, No.5, pp.513-523, 1988. DOI: [http://dx.doi.org/10.1016/0306-4573\(88\)90021-0](http://dx.doi.org/10.1016/0306-4573(88)90021-0)
- [17] M. A. Russell, Mining the social web, 2nd Edition, p.496, bpublic, 2015.
- [18] D. Helic, C. Trattner, M. Strohmaier, K. Andrews, “Are tag clouds useful for navigation? a network-theoretic analysis”, *International Journal of Social Computing and Cyber-Physical Systems*, Vol.1, No.1, pp.33-55, 2011. DOI: <http://dx.doi.org/10.1504/IJSCCPS.2011.043603>
- [19] DoD Manual 4120.24, Defense Standardization Program (DSP) Procedures, 2014, Available from: <https://www.dsp.dla.mil/Policy-Guidance/> (accessed Dec. 1, 2020)
- [20] K. K. Park, H. J. Oh, N. Y. Kim, Must Learning with R, WikiDocs, 2020, Available from: <https://wikidocs.net/book/4315> (accessed Dec. 1, 2020)
- [21] W. J. Yoo, Deep Learning for Natural Language Processing, WikiDocs, 2020, Available from: <https://wikidocs.net/book/2155> (accessed Dec. 1, 2020)

채 수 환(Soohwan Chae)

[정회원]



- 2020년 2월 : 한국과학기술원 신소재공학과 (공학석사)
- 2019년 12월 ~ 현재 : 국방기술품질원 표준인증연구부 연구원

<관심분야>

신소재공학, 표준

심 보 현(Bohyun Shim)

[정회원]



- 2013년 2월 : 광주과학기술원 광공학응용물리학과 (공학석사)
- 2013년 2월 ~ 현재 : 국방기술품질원 표준인증연구부 선임연구원

<관심분야>

형상관리, 표준화

염 슬 기(Seulki Yeom)

[정회원]



- 2014년 2월 : 부산대학교 기계공학부 (공학석사)
- 2017년 6월 ~ 현재 : 국방기술품질원 표준인증연구부 연구원

<관심분야>

품질경영, 표준화

홍 성 돈(Seongdon Hong)

[정회원]



- 2003년 8월 : 송실대학교 섬유공학과 (공학석사)
- 2013년 8월 : 송실대학교 유기신소재·파이버공학과 (공학박사)
- 2003년 7월 ~ 2010년 12월 : (주)이랜드 섬유연구소 과장
- 2011년 1월 ~ 현재 : 국방기술품질원 표준인증연구부 책임연구원

<관심분야>

섬유공학, 품질공학, 표준