

민·군규격용어 표준화를 위한 로드맵 구축 연구

Research to establish a road map for the standardization in military and commercial terminology

박정호, 최영호*, 임익순**, 장효준***
국방기술품질원, 송실대학교*,
충남대학교**, (주)다이퀘스트***

Park jeong-ho, Choi young-ho*, Im ik-soon**,
Jang hyo-jun***
Defense Agency for Technology and Quality,
Soongsil University*, Chungnam University**,
Diquest Inc.***

요약

본 연구는 국방규격서의 전문어, 오용어, 어문규정 및 순화어 미(未)준수 어휘를 추출, 정의 또는 순화어로 정제하는 맵핑구조를 제시, 민·군규격용어 표준화를 위한 정보업무 로드맵을 구축하여 민간용어와의 호환성 및 일관성을 유지할 수 있는 지원체계를 연구하였다. 대상 규격용어는 KS용어표준 원칙을 기본으로 한 신뢰도 평가와 텍스트 마이닝(text mining)빈도분석을 이용하여 선정하였으며, 시소러스(thesaurus) 체계를 삽입, 개념기반 서비스의 확장성을 제시하였다. 이를 기반으로 산출된 규격용어 DB는 민간 및 국방 관련분야의 용어표준관리 정보체계에 검색 및 용어설명에 활용될 수 있다.

I. 서론

1. 민·군규격용어 표준화의 배경과 목적

민·군규격용어 표준화는 군용어 순화추진정책 부응, 민·군규격용어와의 호환성을 제고하며 행정 및 대민지원업무의 효율성을 위하여 오용되고 어문규정을 준수하지 못한 용어들을 현 시제에 맞추어 순화시키는 표준 활동이다. 민·군규격용어 표준화는 비주기적인 용어의 추출, 선정, 편찬, 심의과정을 거쳐 제정 및 개정이 되고 있지만, 국방규격이 운용된 70년대 이후 현재까지도 비순화어 및 민간규격용어와의 상이성이 산재해 있다. 이에 대한 교정의 즉시성을 확보, 보급·활용의 확산을 위한 표준화 절차 로드맵 구축이 본 연구의 목적이다.

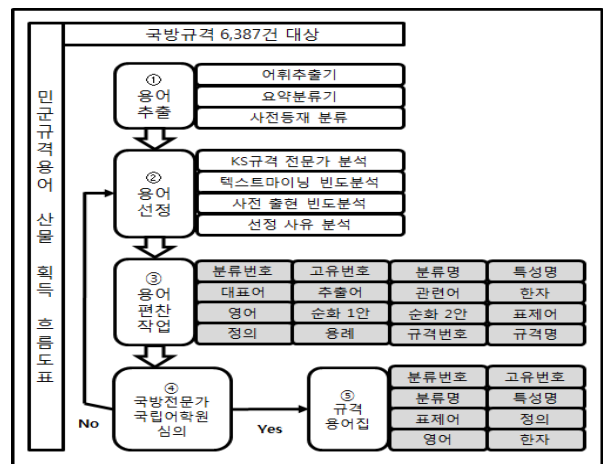
II. 본론

1. 민·군규격용어 표준화 대상 범위

국방규격은 약 3.3만여 건으로 구성되며, 이 중 한글 문서로 전산화된 6,387건을 대상으로 약 100만여 개의 용어를 한글형태로 자동 추출하였고, 이를 다시 지상무기, 탄약, 항공·함정, 기동, 통신·전자, 시설, 전력지원으로 분과, 총 50명의 국방, 국어, IT 전문가들이 참여, 약 10만 개의 대상 용어를 추출, 수집하였다¹⁾.

2. 민·군규격용어 표준화 관리 절차

민·군규격용어 중 표준화 대상 용어 발제 시 이해관계자, 분과위원회, 심의위원회에서 적합성 평가와 함께 [그림-1]의 절차에 따라 편찬 작업이 수행된다.



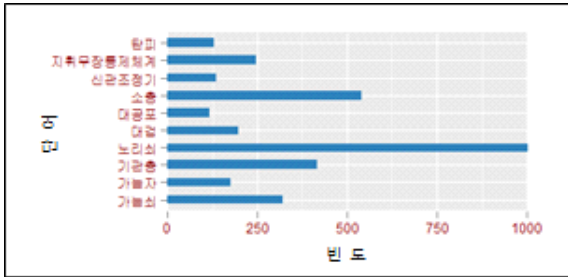
▶▶ 그림 1. 민군규격용어 표준화 절차 로드맵

① 용어 추출: 국방규격 문서는 어휘추출기를 통해 추출 어휘, 대분류, 중분류, 용례가 필터링(filtering) 되며, 요약분류기를 통해 중복용어 제거 및 전문용어사전에 등록여부를 확인, 추출용어가 선정된다.

② 용어 선정: 추출용어의 신뢰도평가는 정성적, 정량적 두 가지 기법을 혼용, 편찬 대상 용어를 추출한다. 첫째, KS용어 표준원칙을 기반으로 하는 어휘평가기법을 이용, 명료성, 일관성, 적합성, 파생성, 모국어성, 언어학적 경제성, 언어학적 정확성 요인을 3점 만점으로 전문가 그룹이 쌍대비교 형식으로 책정, 총합 12점 이상을 획득한 어휘를 중심으로 선정한다. 둘째, 텍스트 마이닝 빈도분석을 실시한다. 아래는 통계분석 프로그램 'R-project'의 ggplot-2 package를 이용한 빈도 분석 프로그래밍 사례이다[2].

표 1. "R" ggplot-2 package 빈도 분석 코딩

```
ggplot(dta, aes(word, freqs)) + geom_bar(stat="identity", fill="darkblue")+
coord_flip(ylim=c(0,1000)) + labs(x="빈도", y="단어", face="bold") +
theme(axis.text.x = element_text(colour="darkred", size=rel(0.9)),
axis.text.y = element_text(colour="darkred", size=rel(0.9)))
```



▶▶ 그림 2. "R" ggplot-2 package 빈도 분석 결과물

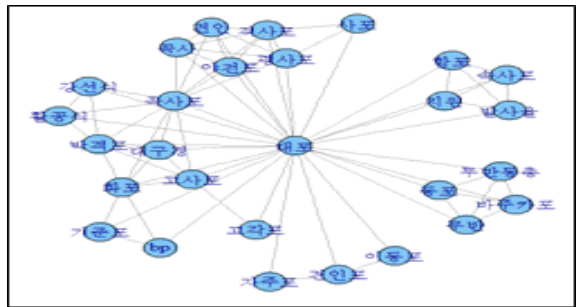
셋째, 국방과 민간사전의 출현빈도를 기준으로 중요도를 선정한다. 넷째, 선정사유는 (a)국방규격 전문용어, (b)어려운 한자어, (c)일본식 한자어, (d)외래어 표기 미 준수, (e)동의어로서 하나의 개념을 다르게 표기한 어휘, (f)변경어로서 제도, 규정, 기술, 시대에 따라 변경이 필요한 어휘, (g)신조어로 새로이 생성된 어휘, (h)띄어쓰기 어문규정 미 준수, (i) 기타 어문규정 미 준수, (j)유의어 관계에 있는 어휘, (k) 어휘추출 오류 및 원문참조가 필요한 어휘[3]로 분류하며, 선정사유의 중요도가 낮은 용어는 정비대상용어¹⁾로 이전한다.

③ 용어편찬 작업: 선정된 용어는 국방분야 및 국립국어원용어사전 등을 참조하여 표제어(순화어), 용어의 정의, 시소러스 등의 작업을 수행한다. 시소러스 작업은 온톨로지(ontology) 개념기반 서비스에 확대 적용, 유의어 및 파생어의 공통영역에 대한 이해를 돕게 한다. 아래는 통계분석 프로그램 'R-project'의 tm, KoNLP, igraph packages 를 이용한 시소러스 프로그래밍 사례이다.

1) 정비대상용어: 전문어휘는 아니나 교정이 필요한 용어로, 오타, 띄어쓰기 미 준수, 외래어 순화 표기 오류 등이 있는 용어를 의미한다.

표 2. "R" tm, KoNLP, igraph packages 시소러스 코딩

```
names(docs) <- paste("doc", 1:length(docs), sep="")
docs.corp <- Corpus(VectorSource(docs))
konlp_tokenizer <- function(doc) { extractNoun(paste(doc, collapse = " ")) }
tdmat <- TermDocumentMatrix(docs.corp, control=list(tokenize=konlp_tokenizer,
wordLengths=c(2,Inf)))
tdmat1 <- as.matrix(tdmat)
tdmat1[tdmat1>=1] <- 1
tdMatrix <- tdmat1 %*% t(tdmat1)
#Result1
tdG <- graph.adjacency(tdMatrix, weight=T, mode="undirected")
tdG <- simplify(tdG)
td_layout <- layout.fruchterman.reingold(tdG)
plot(tdG,layout=td_layout)
```



▶▶ 그림 3. "R" tm, KoNLP, igraph packages 시소러스 결과물

- ④ 국방전문가/국립어학원 심의: 표준화작업을 마친 용어는 국방전문가 및 어학원 심의를 거친 후 규격용어집에 실리게 되며, 필요시 삭제 또는 수정작업을 수행한다.
- ⑤ 규격 용어집: 최종 결과물을 의미한다.

Ⅲ. 결론

본 연구에서 제시한 로드맵은 민·군규격용어 편찬업무의 즉시성, 상시성을 위하여 단순화 하였고, 최적화 작업의 구조를 위한 기본적인 신뢰성을 제시하였다. 또한, 편찬된 규격용어 DB는 보급·확산을 위해 국방기술품질원 국방기술정보서비스(Defence Technology Information Service, DTiMS)에 탑재, 온라인 검증이 가능하며, 상향식 민간용어표준 환경에 기여할 것으로 사료된다.

■ 참고 문헌 ■

- [1] 최중환 외 "국방분야 전문용어 사전 및 워크벤치 개발", 한국콘텐츠학회 논문지, 2012
- [2] 오장욱 외, "A comparative study of positive and negative keywords based on text mining", ICIC Express letters, Vol. 6, No. 3, pp. 665~671, 2014
- [3] 이길섭 외 "민군규격용어 표준화 방안 및 협업지원체계 구축", 국방기술품질원, pp. 61~64, 2013