

국방과학기술 전문용어 사전 구축을 위한 프로세스 및 워크벤치 개발

Science and Technology Terminology Dictionary Building Process and Workbench Development in Defense Area

최중환*, 박정호*, 김경선**, 김 평***
국방기술품질원*, ㈜다이퀘스트**, 전주교육대학교***

Jungwhoan Choi(cjhjkw@gmail.com)*, Jeongho Park(smilepjh1@naver.com)*,
Kyungsun Kim(kksun@diquest.com)**, Pyung Kim(pyung@jnue.kr)***

요약

용어 의미 표준화를 통한 업무의 효율성을 높이기 위해서 다양한 분야에서 전문용어가 구축·활용되고 있으며, 국방분야에서도 국방용어 사전 발간을 통해 각 군의 효율적인 업무 연계 및 표준화된 국방 용어 유통 체계를 정립하는데 노력하고 있다. 국방기술품질원에서도 3년을 주기로 국방과학기술분야의 전문용어 사전을 발간함으로써 전문용어 구축 표준화 연구는 물론, 전문용어의 활용성 제고에 힘쓰고 있다. 본 연구에서는 국방기술품질원에서 전문용어 사전을 실제로 구축하는데 필요한 구축 프로세스와 조직, 이를 지원하기 위한 워크벤치에 대해서 기술하고 있다. 즉, 국방과학기술 전문용어 사전을 구축하기 위한 전문용어 사전 구축 프로세스 및 조직의 구성, 표제어 정의, 전문용어 후보 추출을 위한 대상 문서의 선정, 전문용어 후보 추출, 전문용어 후보군 생성, 워크벤치 등록, 사전 구축, 사전 검증 과정에 대해서 설명하고 있다. 또한 전문용어 사전의 활용성 제고를 위해서 시소러스를 구축·활용하는 내용은 물론, 전문용어 사전 구축을 위해서 개발된 워크벤치의 기능을 중심으로 효율적인 전문용어 사전 구축 지원 방법에 대해서도 기술한다.

■ 중심어 : | 전문용어 사전 구축 | 워크벤치 | 시소러스 |

Abstract

To improve the efficiency of business, it is important to standardize the meaning of terminology. And then, terminology dictionaries have been actively being built and used in various fields. In defense area, the publication of defense terminology dictionary is useful for information exchange of each army and distribution of standardized terminology. Defense agency for technology and quality(DTaQ) publishes terminology dictionary of defense science and technology on a three-year cycle. DTaQ tries to standardize the construction process of terminology dictionary and improve service efficiency by using terminology dictionary in defense area. This proposed method is based on the result of previous study about standardization of terminology dictionary. We suggest the practical steps including terminology dictionary constructing process, composition and role of organization, definition of headword, selection of target documents to be extracted terminology candidate, terminology extraction, generation of terminology candidate group, workbench registration, construction and validation of terminology dictionary. Thesaurus and workbench are developed to use and support terminology dictionary effectively.

■ keyword : | Terminology Dictionary Construction | Workbench | Thesaurus |

I. 서론

전문용어란 어떤 분야에서 특화된 어휘로서, 전문용어의 표준화는 업무 의사소통의 효율화를 위해 핵심적으로 수행되어야 하는 기능이다[1]. 과학기술의 발전 속도가 점점 빨라짐에 따라 업무 의사소통의 효율성 향상을 위해서 다양한 분야에서 전문용어 사전이 구축되어 활발하게 활용되고 있다. 국방분야에서도 국방부, 합동참모본부, 국방과학연구소를 중심으로 다양한 전문용어 사전들이 구축되었다, 국방기술품질원에서는 국방과학기술 전문용어 사전 구축 과정을 표준화하기 위한 연구[2]를 수행하였고, 국방과학기술 전문용어 사전을 3년을 주기로 발간하고 있으며, 2011년에는 약 1.5만개의 국방과학기술 용어를 대상으로 전문용어 사전을 발간하였다[3].

특정 분야의 전문용어 사전을 구축하는 작업은 해당 분야의 전문가는 물론, 국어학 전문가, IT 전문가들의 협업이 필요하며, 전문용어 사전 구축 과정의 효율화를 위해서는 전문용어 후보군 선정부터 정제, 사전 발간, 활용을 포함하는 일련의 프로세스를 지원하기 위한 워크벤치의 개발 및 활용이 요구된다. 또한 구축된 전문용어 사전의 활용을 극대화하기 위해서는 온라인서비스에서 연계·활용이 가능해야 한다. 국방기술품질원에서는 전문용어 사전의 활용성 제고를 위해서 전문용어 사이의 관계정보(USE/UF, BT/NT, RT 등)를 포함한 시소러스를 추가 구축하고, 국방기술정보통합서비스인 DTiMS 시스템[4]과 시소러스를 연계하고, 사용자들이 DTiMS 서비스 사용시 전문용어와 시소러스를 활용함으로써 정보의 접근성을 개선하였다.

본 연구에서는 국방기술품질원에서 전문용어 사전 구축 표준화 프로세스[2]를 준용해서 국방과학기술 전문용어 사전을 구축하는 과정에서 구축 단계별 상세 프로세스는 물론, 각 프로세스에서 나타난 이슈들과 해결 방안을 제시함으로써, 실제 연구자들이 사전 구축 작업을 수행하는데 도움이 되고자 한다. 즉, 전문용어 사전을 구축하는데 투입된 인력과 역할, 전문용어 후보 추출을 위한 대상 문서를 선정하는 방법, 전문용어 후보를 선정하는 방법, 표제어 기술 방법, 시소러스의 활용

방법과 워크벤치 기능을 중심으로 전문용어 사전 구축에 필요한 실제 프로세스들을 구체적으로 기술하고자 한다. 2장에서는 전문용어 선정을 위한 연구와 표준화 연구, 워크벤치에 관련된 기존 연구들을 살펴보고, 3장에서는 전문용어 사전 구축을 위한 조직과 프로세스, 4장에서는 전문용어 사전 구축을 위한 워크벤치의 기능, 5장에서 결론과 향후 연구내용을 중심으로 기술한다.

II. 관련 연구

전문용어는 어떤 분야의 특화된 어휘로서 자연어 처리 기술들을 이용해서 전문용어를 자동인식하기 위한 연구들은 지속적으로 연구되어 왔다[5][6]. 생의학 분야에서 전문용어를 자동인식하기 위한 연구[5]는 다양한 전문용어의 특징을 고려하고 이들 중 전문용어 선택에 효과적인 특징을 도출한 후, 이들을 기계학습방법으로 통합하고 순위화하여 전문용어 여부를 결정하는 연구이다. 분야간 유사도와 통계기법을 이용한 전문용어 인식 연구[6]는 기계가독형 전문분야 사전들을 이용하여 사전 간의 계층관계를 구축하고 이를 이용하여 전문용어를 추출하는 연구로서, 전문용어 사전에서 나타나지 않는 전문용어를 추출하기 위하여 용어의 빈도수, 외래어 및 외국어, 내포관계 등을 포함한 통계기법을 사용하였다.

전문용어와 시소러스 구축 과정 및 이를 표준화하기 위한 연구는 학술 전문용어 정비 및 표준화 연구[7], 전문용어 표준화의 원칙과 평가기준 연구[8], 전문용어 정비 지침 연구[9], 시소러스 구축 및 활용방안 연구[10] 등이 있으며, 국내는 물론 해외에서도 활발하게 연구되고 있다. 특히 국방기술품질원의 국방과학기술분야의 전문용어 사전 구축 표준화 연구[2]는 국방과학기술분야 전문용어 후보 선정 및 표제어 기술 방법, 표준화를 위한 표준화 조직 및 구성원의 역할을 망라하고, 과학기술용어 사전에 수록된 단어를 중심으로 전문용어 사전을 시범 및 구축된 사전의 활용 방안에 대해서 제시하고 있다. 국방분야에서도 국방대학교 주관으로 국방군수 용어 사전이 2008년 만들어졌으며 약 9천개의 군

수 용어들을 포함하고 있다. 2006년에 발간된 합동참모 본부의 군사용어 사전은 군사용어와 영문약어를 포함해서 약 1만개의 단어를 포함하고 있다. 그러나 국방분야에서 국방용어에 대해서 시소러스를 구축하고 활용한 사례는 찾아보기 어렵다.

워크벤치는 사전 구축, 품사 태깅, 시물레이션 작업 등을 효과적으로 지원하기 위해서 개발된 시스템으로 다양한 분야에서 활용되고 있다. 한국어 개념망 구축을 위한 지능형 워크벤치 연구[11]는 어휘들의 뜻풀이와 상위어간의 관계성, 복합명사와 상위어간의 관계성을 통계적으로 분석한 후, 분석된 결과를 기반으로 확률모델을 이용하여 상위어 추천 기능을 구현한 워크벤치를 통해서, 다양한 도메인에서 요구되는 개념망 구축 시간을 절감하였다. 합성명사 의미해석용 사전 구축을 위한 워크벤치 연구[12]는 합성명사를 구성하고 있는 구성명사간 의미관계의 결합 종류를 규명하기 위해서, 워크벤치를 통해서 의미관계 패턴을 정의하고, 정의된 의미관계 패턴을 이용해서 합성명사를 자동으로 추출하고, 추출된 합성명사 사전을 이용해서 각 명사의 상위 개념에 대해서도 합성명사의 의미관계를 해석할 수 있는 기능을 제공하고 있다. 국립국어원에서는 다양한 학술분야의 전문용어를 심의하고 관리하기 위한 도구로서 워크벤치를 활용하고 있으며, 가톨릭 의대에서도 의학분야의 전문용어를 정비하고 관리하기 위해서 워크벤치를 개발해서 활용하고 있다.

III. 전문용어 사전 구축

국방과학기술 전문용어 사전 구축을 위한 조직, 역할, 구축 프로세스는 전문용어 사전 구축 표준화 연구[2]에서 제시한 기준을 준용하지만, 사전 구축 프로세스의 효율화를 위해서 워크벤치를 개발·활용함에 따라 일부 프로세스를 수정하여 적용하였다. 2011년 수행된 국방과학기술분야 전문용어 사전은 총 1.5만개의 전문용어를 포함하고 있으며, 국방분야 전문가, 국어학 분야 전문가, 전산 전문가를 포함한 약 50명의 인력이 9개월(2011.4~2011.12) 동안 사업을 수행하였다.

1. 전문용어 사전 구축 프로세스 및 조직의 구성

전문용어 사전은 다양한 분야에 속한 전문가들의 협업을 통해 구축될 수 있다. 사전 구축 프로세스는 전문용어의 기준 선정, 전문용어 수집 및 사전 데이터 구축, 검증 등의 복잡한 과정들을 포함하고 있다. 전문용어 사전 구축을 위한 조직들은 연구 총괄팀, 용어 수집팀, 용어 선정팀, 사전 구축팀, 사전 검증팀과 시소러스 구축팀으로 구성되며, 각 팀은 중복 인적 구성을 허용하고 있다. 작업 기간의 단축을 위해서 전문용어의 선정, 전문용어 데이터 입력, 사전 구축, 검증, 시소러스 구축 작업들은 병렬적으로 진행된다.

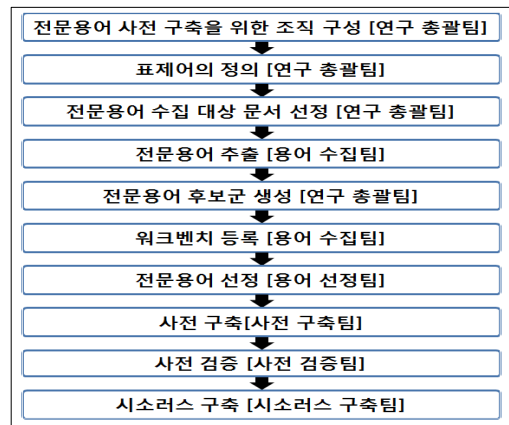


그림 1. 전문용어 사전 구축 프로세스

사전 구축 프로세스와 프로세스별 담당 팀은 [그림 1]과 같고, 각 팀의 구성 및 역할은 다음과 같다.

- 연구 총괄팀: 국방분야 전문가 2명, 전산 전문가 2명, 국어학 전문가 1명으로 구성되어, 사업의 전체 프로세스 관리, 조직의 구성, 용어의 수집·선정 기준 마련 등의 작업을 담당한다.
- 용어 수집팀: 전산 전문가 3인으로 구성되어, 국방분야 서비스, 저널, 각종 용어 사전을 대상으로 전문용어 후보를 수집하고, 용어의 선정 기준에 적합한 용어들을 워크벤치에 등록하는 작업을 담당한다.
- 용어 선정팀: 국방과학연구원, 국방기술 품질원, 합동참모본부, 육군본부 등의 국방전문가 32명으

로 구성되어, 용어의 활용도와 중요성을 고려해서 사전으로 구축될 용어를 선정한다.

- 사전 구축팀: 용어 사전 구축 분야 전문가 5인으로 구성되어, 용어의 정의와 사전 발간에 필요한 정보를 입력한다.
- 사전 검증팀: 국방분야 전문가 2인과 국어학 전문가 2인으로 구성되어, 사전 구축팀의 결과를 교차 검증한다.
- 시소러스 구축팀: 시소러스 구축 분야 전문가 5인으로 구성되어, 전문용어 시소러스를 구축한다.

2. 표제어의 정의

표제어는 사전에서 각각의 분리되는 개념을 대표하는 말로서, 전문용어 사전의 표제어 원칙이 필요하다. 일반 어휘가 특정 분야의 특정 개념을 표기하는데 사용되는 경우에도 표제어로 사용될 수 있으며, 연구 총괄팀에서는 다음과 같은 5개의 표제어 표기 원칙과 4개의 국방과학기술분야 표제어 선정 기준을 정하였다.

- 표제어 표기 원칙 1: 표제어는 하나의 핵심 개념을 대표하는 것을 전제한다.
- 표제어 표기 원칙 2: 표제어는 하나 이상의 한글, 영어, 숫자만으로 구성하며, 다음의 문자는 표제어에 포함될 수 있다. "-", "/"
- 표제어 표기 원칙 3: 표제어는 1개 이상의 어휘로 구성되며, 어휘란 띄어쓰기 혹은 언어간에 구분될 수 있는 낱글자의 합을 뜻하며, 개별 형태소를 뜻하지 않는다.
- 표제어 표기 원칙 4: 표제어 원칙1에 의거하여 표제어를 선정하였으나 형태가 동일한 경우(동형의 의미)에는 뜻 번호를 부여하고, 뜻 번호를 부여하는 경우에는 많이 사용하는 의미에 1번을 부여한다.
- 표제어 표기 원칙 5: 영어 관용구를 이용하여 한글화 하거나 줄여쓰는 경우에는 줄여쓰기 이전의 풀어 쓴 영어 관용구를 표제어로 한다.
- 표제어 선정 기준 1: 국방과학기술분야의 용어로 분야별 필수 개념 용어와 신규 용어를 선정한다.
- 표제어 선정 기준 2: 사전 구축에 필요한 정보가

충분한 용어를 선정하되, 너무 일반적인 용어, 단순한 형식·모델명, 조직·직급·역할 등을 의미하는 용어 등은 배제한다.

- 표제어 선정 기준 3: 표제어는 한국어 용어를 기준으로 하되, 영어로 기재된 용어는 가급적 한글화된 영어를 기재한다. 한글과 영어가 동시에 추출된 경우 한글로 발음된 용어를 선정하며, 용어의 한글화는 가장 많이 사용되는 개념을 선택한다.
- 표제어 선정 기준 4: 특정 모델의 애칭은 배제되며, 민군겸용용어와 현재 군에서 통용되는 용어를 우선하여 선정한다.

3. 전문용어 후보 수집 대상 문서

전문용어 사전 구축에 필요한 후보 용어군을 만들기 위해서 대상 문서를 선정하고 전문용어 후보를 수집하는 작업은 매우 중요하기 때문에 사전, 저널, 용어집 등을 포함한 다양한 출처들을 대상으로 전문용어 후보를 수집해야 한다. [표 1]은 국방과학기술 전문용어 후보를 수집하기 위해서 사용된 국방 관련 사전, 국방 저널, 국방 관련 용어집들을 보여준다.

표 1. 국방과학기술 전문용어 후보 수집 대상

구분	대상
국방 관련 사전	국방과학기술용어 사전
	군사약어 사전
	Jane's 약어 사전
	나토용어 사전
	항공용어 사전
	해군 C4I 사전
국방 저널	연합합동용어집 등
	Jane's 저널 국방과학기술 저널
국방 관련 용어집	국방기술품질원 정보기술팀 검색용어
	DTIMS 사용자 용어
	국방과학기술연구소 시소러스
	국방기술품질원 전자도서관 논문 키워드
	KISTI IT 용어

4. 전문용어 후보 추출

기존 사전에 등재된 용어와 용어집으로 관리되는 전문용어들은 그대로 전문용어 후보군에 포함된다. 저널

이나 웹 문서와 같은 일반 문서를 대상으로 전문용어 후보를 추출하는 경우에는 다양한 문서 포맷(MS Office, 한글, XML, HTML 등)에서 텍스트를 추출하는 문서 필터와 추출된 텍스트의 파싱을 위해서 (쥘)다이퀘스트[13]에서 제공하는 형태소 분석기, 개체명 분석기 등의 언어 분석기를 사용한다.

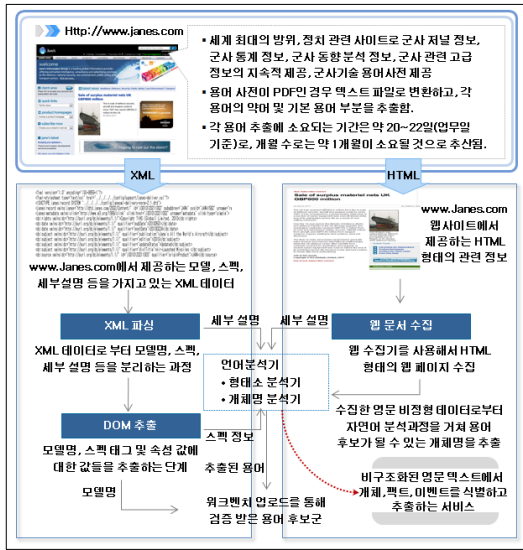


그림 2. 웹 문서에서 전문용어 후보 추출

[그림 2]는 세계 최대의 군사 정보 사이트인 Jane's 홈페이지[14]에서 전문용어 후보를 추출하는 방법으로, XML 문서의 경우는 XML 파싱을 통해 모델명, 스펙, 속성 등의 정형 데이터에서 전문용어 후보를 추출하고, HTML 문서의 경우는 웹 수집기를 통해 웹 문서를 수집한 후 언어 분석을 통해 개체를 구분한 후 전문용어 후보를 추출한다.

5. 전문용어 후보군 생성

전문용어 후보군 생성 과정에서 중요한 것은 다양한 출처에서 매우 많은 용어들이 수집되기 때문에 최종 전문용어 후보군으로 선정될 최종 전문용어 후보들을 선정하기 위한 기준이 필요하다. 용어 사전에 포함될 수 있는 용어의 수가 제한적이기 때문에 사전의 발간 시기를 전후로 활발하게 사용되는 중요한 용어들을 전문용

어 사전 구축을 위한 최종 후보 용어군으로 포함시켜야 한다. 전문용어의 출처, 발생빈도, 전문용어 데이터, 발생 문장 등 다양한 정보들은 전문용어 추출시 같이 수집해서 워크벤치에 등록될 전문용어 후보군을 선정하는데 활용한다.

[표 2]는 출처별로 추출된 용어의 개수를 보여주며, 용어들의 중복을 고려하더라도 약 10만개 이상의 용어들이 추출되는 것을 알 수 있다. 약 10만개 이상의 용어들을 대상으로 1.5만개의 용어를 선정하는 작업은 매우 어렵기 때문에, 용어 선정법에서는 다음과 같은 선정 규칙을 적용해서 약 3만개의 용어 후보군을 선정된 후, 선정된 약 3만개의 용어와 용어 데이터를 워크벤치에 등록한다.

- 선정 규칙 1: 국방과학저널, Jane's 저널, DTIMS Query Log 등 사용자의 사용성을 반영하는 데이터 출처와 국방과학기술용어사전, 기술정보탐 용어집, 나토사전 등 용어의 중요도를 반영하는 출처가 2곳 이상인 용어를 선정 (1.6만개)
- 선정 규칙 2: 국방과학기술용어사전, 기술정보탐 용어집, 나토 사전에 포함된 용어는 이미 검증된 중요한 용어로 표제어 후보군으로 선정 (1.5만개)
- 선정 규칙 3: DTIMS 검색어 등에서 추출된 용어 중 사용자가 10번 이상 사용한 용어를 선정 (1.2천개)

표 2. 출처별 전문용어 후보 용어

출처	용어	출처	용어
국방과학기술용어 사전	7,521	나토용어 사전	6,210
군사약어 사전	16,451	항공용어 사전	93
Jane's 약어 사전	23,963	해군 C4I 사전	367
연합함동용어집 등	5,959	Jane's 저널	51,899
국방과학기술 저널	21,950	DTIMS 사용자 용어	1,841
국방기술품질원 정보기술 팀 검색용어	419	국방기술품질원 전자도서관 논문 키워드	467
KISTI IT 용어	30,000		

6. 워크벤치 등록

전문용어 선정에 사용되는 전문용어 메타데이터와 사전 구축에 참여하는 모든 구성원들, 구성원들의 작업 권한 부여 등, 워크벤치를 활용하는데 필요한 모든 정

보를 데이터베이스화하여 워크벤치에 등록한다. 워크벤치의 세부 기능에 대해서는 다음 장에서 설명한다.

7. 전문용어 선정

전문용어 선정 작업은 32명의 국방 관련 전문가로 구성된 용어 선정팀에서 수행한다. 용어 선정팀은 워크벤치에 등록된 약 3만개의 전문용어 후보군을 대상으로 사전에 등재될 전문용어를 선정하게 되며, 기타 무기체계 분야를 포함한 11개의 국방 무기 체계 분야별 15인의 전문가들로 구성된다. 국방 무기 체계 분야별 복수의 전문가들이 워크벤치를 통해서 전문용어의 표제어 표기, 내용과 분야를 확인한 후, 사전에 구축될 최종 전문용어를 선정하였다. 이 과정에서 약 1천 이상의 전문용어를 추가적으로 선정함으로써, 사전 구축 과정에서 전문용어 데이터의 부족으로 일부 전문용어가 누락되어도 전체 개수에는 차질이 없도록 하였다.

8. 사전 구축

전문용어의 사전 표기는 [그림 3]과 같이 “한글(한자) [분야] 영어 한글 정의문”의 형식으로 구성된다.

가계약	
<p>가간섭성 干渉性 [지휘통제/통신] <i>Coherence; Coherency</i></p> <p>(1) 공간상 또는 시간상으로 분리되어 있는 지점에서 전자파의 사이의 상관관계. (2) 전자파가 파장이 같은 가지에서 진행하여 집합된 상태. ≡ 코히런스(Coherence)</p> <p>가간섭성 광 干渉性空間 [지휘통제/통신] <i>Coherence Area</i></p> <p>전파 방향에 수직인 평면상에서 빛이 충분히 가간섭성이라고 할 수 있는 영역. 일반적으로 가간섭성으로 이루어진 공간을 말한다. 이 영역으로부터의 빛은(나) 영역에 놓인 2개의 바로 구멍을 통과하면) 간섭에 의한 명암(明暗)의 패턴(무늬 모양)을 만든다.</p> <p>가간섭성 레이저 干渉性 레이저 [지휘통제/통신] <i>Coherence Ray</i></p> <p>복사 광파(光波)의 위상과 주파수가 같아서 공간적으로나 시간적으로 일치할 때의 광파. 통신에 이용되는 광은 가간섭성 광이어야 하나 일반 광은 그렇지 않다. 가간섭성의 발견기를 사용하여 특별한 가간섭성 광을 생성시켜 이용하는 것이 레이저나 레이저광(holography)이다.</p> <p>가간섭성 광파 전송 干渉性光波傳送 [지휘통제/통신] <i>Coherent Light Wave Transmission</i></p> <p>전파와 마찬가지로 광파의 가간섭성을 이용하여 빛이 주파수와 위상에 신호를 실어서 전송하는 방식. 이 경우에 가간섭성인 주파수가 일정하여 위상이 시간적, 공간적으로 일정한 빛의 파, 즉 가간섭성 광파를 정보를 실는 파로 이용한다는 의미이다. 빛의 강도에 따른 송수의 코딩, 변조, 광필수 전송 방식보다 고속으로 대용량 전송이 가능하다. 장래의 광대역 종합 정보 통신망(B-ISDN)이나 가입자의 가정까지 광선망을 포설하는 광가입제 시스템인 광이더 두 더 호프(FTTH)의 중요한 기술의 하나이다.</p> <p>가격 고사 (價格高售) [공통] <i>Quotation</i></p> <p>증권, 외환 시장 및 부에 등에서 광범위하게 사용되는 용어로서 시세 기준 가격, 가격 산정 등 의 의미로 통용된다. 증권 시장에서는 거래되는</p>	<p>유가 증권의 매매 계약이 체결될 때의 가격을 가리키며, 외환 시장에서는 외국환 은행이 각종 외환에 대해 제시하는 외환 시세를 말한다. 한편 부에에 있어서는 부에 재산의 기초가 되는 제시 가격의 산정 또는 그 결과로서의 산정 가격을 의미한다.</p> <p>가격 및 가용성 자료 (價格-可用性資料) [공통] <i>Price And Availability Data</i></p> <p>특히 군사 판매를 통해 물자 및 용역을 구매하기 위해서 가격 및 가능한 납기를 판단하여 계획과 예산을 검토하고, 오퍼 요청서를 작성하는 데 필요한 기초 자료를 작성하기 위해 요구하는 자료로서 절차 및 수확시(COA)와 동일한 수준의 신의상과 상세한 가격 정보를 가지며, 이는 미국 방무 해당 기관에서 요청 받은 날로부터 45일 이내에 구매자에 제공됨. ≡ PSA(Price And Availability Data)</p> <p>가격 정보 (價格情報) [공통]</p> <p>외국 거래관리의 주요 구성요소로서 외화 산인 과정에서 가격정보는 계약실적, 일체권(계약), 마크로 리와 더불어 계약단계 결정에 중요한 자료로서 일일금액 이상 품목에 대해서는 반드시 가격정보를 요청, 획득하도록 하고 있음.</p> <p>가격 조사 (價格調査) [공통] <i>Price And Cost Investigation</i></p> <p>적정한 거래가 형성된 경우에 거래 실제 가격을 (명목)의 규정에 의하여 가격이 결정된 경우에는 결정 가격의 범위 안에서 거래 실제 가격을), 거래 실제 가격 조사 또는 원가 계산 방법에 의한 수 없는 경우에는 경쟁 가격 및 유사한 거래 실제 가격 또는 전의 가격을 조사하여 계약 가격을 결정하기 위한 기초 자료를 작성하는 것을 말한다.</p> <p>가계약 (假契約) [공통] <i>Draft Contract</i></p> <p>계약의 효력이 아직 가지지 않거나 특정 조건 하의 증서시 발표기로 약정하고 체결한 계약을 말</p>

그림 3. 전문용어 사전 예제

사전 구축에 필요한 전문용어 데이터를 입력하는 작업은 사전 구축팀에 의해서 수행되며, 사전 구축팀은 기존 전문용어의 정의문과, 문헌에서 추출된 내용을 활용해서 사전 구축에 필요한 데이터를 워크벤치를 통해 입력하며, 전문용어 데이터에 대한 작업 이력은 워크벤치를 통해 관리된다.

국방분야 전문가들이 사전 구축팀에 참여해서 전문용어의 표제어인 한글과 한자, 영어, 분야, 정의문을 작성하는 것이 가장 효과적이겠지만, 실제로는 국방분야 전문가들이 많은 시간이 소요되는 사전 구축 작업에 참여하기 어렵기 때문에, 사전 구축팀은 사전 구축 전문가인 국어학분야의 전문가들로 구성된다. 사전 구축팀은 전문용어의 정의문을 생성하기 위해서 기존 용어의 정의문과, 전문용어 선정시 전문가들이 등록한 전문용어 데이터, 웹 자원 검색, 전문용어의 출처 정보 등을 활용한다.

전문용어 사전 구축 프로세스를 진행하면서 수집된 모든 전문용어 데이터는 데이터베이스를 통해 저장·관리되며, 사전 구축팀은 워크벤치를 통해 저장된 전문용어 데이터를 확인하고, 사전에 출판되는 내용에 동영상, 이미지를 포함한 저작권에 문제가 있는 데이터를 제외하는 것은 물론, 추가적으로 필요한 전문용어 데이터를 수집해서 전문용어 데이터를 보강하는 작업을 수행한다.

9. 사전 검증

사전 구축팀에 의해서 구축된 전문용어 데이터에 대한 감수는 1차와 2차로 구분되어 수행된다. 1차 감수는 국방분야 전문가들로 이루어진 용어 선정팀에 의해서 수행된다. 즉, 국방분야 전문가들이 직접 사전 구축팀으로 참여하지 못했지만, 사전 구축 결과에 대한 1차 감수를 통해서 사전 구축 결과를 확인하고 보정하는 작업에 참여하게 된다. 2차 감수는 별도의 국방분야 전문가 그룹에 의한 내용의 감수, 국어학 전문가 그룹에 의한 국문학적 감수를 통해 진행된다.

10. 시소러스 구축

전문용어 사전의 활용성 제고를 위해서 1.5만개의 전문용어를 대상으로 시소러스를 구축하였으며, 시소러

스에 포함된 용어는 국방과학기술용어사전에 등재되는 전문용어 약 1.5만개와 시소러스 구축과정에서 추가된 확장용어 및 동의어를 포함하여 약 7.5만개의 용어로 구성되었다(통계숫자에서 외국어는 제외).

표 3. 국방과학기술 전문용어 시소러스 항목

속성	속성명	반복 여부	필수/재량	설명/표기	스킴
용어	용어		필수	한자 혹은 원어를 '[]' 로 묶어 부기	
정의	정의	0	해당시 필수		text
대응 외국어	언어 코드	0	해당시 필수	1) 상이한 언어 코드의 반복 2) 같은 언어 코드의 반복	ISO 639-3: 2007
분류	분류	0	필수	1) 상이한 분류체계명의 반복 2) 같은 분류체계명 내에서의 반복	분류 체계명
관계	동의어	0	재량	UF	
	관련개념어	0	재량	RT	
	상위개념어 (속관계)	0	재량	BT	
	상위개념어 (부분관계)	0	재량	BTP	
	하위개념어 (속관계)	0	재량	NT	
	하위개념어 (부분관계)	0	재량	NTp	
	이전/이후용어	0	해당시 필수	PT/LT	

용어(우선어) →	무인기(無人機)
용어의 해설 →	정의 1) 『항공』 레이더, 발파비전 따위의 전자 공격적 장치에 의하여 자동 조종이 되거나 원격 조종이 되는 항공기. 2) <국방> 무인 비행체. 일반적으로 사전에 경계권 경로를 따라 비행하고 1회 사용하면 회수하지 못하는 무인기(drone)와, 외부 조종에 의해 상황에 따라 임의의 비행경로를 비행한 후 회수하여 재사용할 수 있는 원격 조종 무인기(RPV: Remotely Piloted Vehicle)로 구분한다. <국방> UAV(무인기)(unmanned aerial vehicle) 무인 비행기 [無人飛行機] 무인 비행체 [無人飛行體] 무인 항공기 [無人航空機] 무인 장비 [無人裝備] 항공기 [航空機] 무인 항공 체계 [無人航空體系] 공격용 무인기 [攻擊用無人機] 기만용 무인기 [欺騙用無人機] 대공 제압 무인기 [對空制壓無人機] 드론(비행체)(drone) 무선 조종 항공기 [無線操縱航空機] 무인 감시기 [無人監視機] 무인 공격기 [無人攻擊機] 무인 전투기 [無人戰鬥機] 무인 경찰기 [無人偵察機] 무인 표적기 [無人標的機] 소형 무인기 [小型無人機] 원격 조종기 [遠隔操縱機] 전기 구동 무인기 [電氣驅動無人機] 전술 무인 항공기 [戰術無人航空機] 제트 추진식 무인기 [jet推進式無人機] 중계 UAV [中繼UAV] 중고도 무인 항공기 [中高度無人航空機] 재투입 무인기 [滯留無人機] 함정 탑재 무인기 [艦艇搭載無人機]
동의어 →	UF
상위개념어 →	BT
하위개념어 →	NT

그림 4. 시소러스 구축 예제

시소러스 구축 항목은 전문용어 구축 표준화 연구[2]에서 제시된 전문용어 사전의 기술 구조 항목중에서 관리 항목을 제외하고 선정하였다. [표 3]은 본 사업에서 구축된 시소러스 항목과 항목에 따라 반복 여부, 필수/재량 여부, 설명과 표기, 표기 스킴을 보여준다. 시소러스 정보는 용어, 용어에 대한 정의, 대응 외국어, 분류, 용어간 관계 정보를 포함하고 있으며, 시소러스의 효율적인 관리와 전자적 활용성 제고를 위해서 XML을 이용해서 저장·관리된다. [그림 4]는 전문용어 '무인기'를 중심으로 시소러스 구성 항목에 대한 예제를 보여주고 있다.

IV. 전문용어 사전 구축 워크벤치

전문용어 워크벤치는 용어 관리 및 관리 절차를 효율적으로 지원하기 위해서 많이 개발 및 활용되어 왔다. 본 연구에서 개발된 전문용어 사전 구축을 위한 워크벤치는 선정된 용어 대상 후보군을 등록하여 국방 무기체계 전문가들이 워크벤치를 활용하여 사전에 등재될 용어를 선정하는데 이용된다.

그림 5. 전문용어 워크벤치

워크벤치는 용어에 관한 정의, 분류 등의 데이터 등을 입력하는 편집도구로 활용될 뿐만 아니라 입력된 데이터를 검수하는 작업에도 활용되며, 용어의 정보 변경에 따른 이력 관리 기능도 제공한다. 워크벤치는 그림 5와 같이 구성되어 있으며, 작업자들은 로그인을 통해 작업 권한을 부여받고, 작업자의 자격 권한에 따라서 할당된 기능을 수행할 수 있다. 워크벤치에서 지원하는 기능들은 다음과 같다.

- 전문용어 선정: 다양한 정렬 및 필터를 제공함으로써, 용어 선정팀 전문용어를 선정하는 과정에서 전문용어 데이터를 확인하고 의견을 제시하는 기능
- 전문용어 사전 구축을 위한 데이터 입력: 용어 구축팀에 속한 복수의 구축자들이 전문용어의 데이터 입력하고, 데이터 수정 이력을 관리하는 기능
- 전문용어 관리를 위한 통계 데이터 제공: 전문용어 구축 현황을 파악할 수 있도록 분야별, 작업자별, 전문용어별 작업 통계 정보 데이터를 제공하는 기능
- 전문용어 검수: 국방 무기체계 전문가와 국어 전문가에 의한 전문용어 검수 과정 및 전문용어 데이터 수정 작업을 지원하는 기능

V. 결론

전문용어는 다양한 분야에서 용어 의미 표준화를 통한 업무의 효율성을 높이기 위해서 구축·활용되고 있으며, 전문용어에 관한 다양한 연구들이 수행되어 왔다. 그러나 기존의 연구에서는 전문용어를 추출하는 방법이나 전문용어의 전문도 측정, 또는 표준화에 대한 연구들로서, 실제로 특정 분야의 전문용어 사전을 구축하는 과정에서 필요한 구체적인 내용들을 기술하고 있지 않다. 본 연구에서는 국방기술품질원에서 3년을 주기로 발간하고 있는 국방과학기술 용어 사전의 실제 구축 사례를 중심으로 전문용어 사전을 구축하는데 필요한 구축 프로세스와 조직, 이를 지원하기 위해 개발된 워크벤치에 대해서 상세하게 기술하고 있다. 즉, 국방과학기

술 전문용어 사전을 구축하기 위한 전문용어 사전 구축 프로세스 및 조직의 구성, 표제어 정의, 전문용어 추출을 위한 대상 문서의 선정, 전문용어 추출, 전문용어 후보군 생성, 워크벤치 등록, 사전 구축, 사전 검증 과정에 대해서 설명하고 있다. 또한 전문용어 사전의 활용성 제고를 위해서 시소러스를 구축·활용하는 내용은 물론, 전문용어 사전 구축을 위해서 개발된 워크벤치의 기능을 중심으로 효율적인 전문용어 사전 구축 지원 방법에 대해서도 기술하였다.

본 연구의 결과물로 구축된 시소러스는 DTIMS 서비스와 연계·활용되어, 시소러스에 구축된 용어가 사용자의 검색 질의로 사용된 경우 용어에 대한 분야와 정의, 용어간 관계 정보를 제공함으로써, 사용자의 정보 접근성 향상에 기여하고 있다. 또한 신규 전문용어 후보들을 워크벤치에 지속적으로 등록하고 관리함으로써, 3년 주기로 발간되는 전문용어 사전 구축 작업을 보다 쉽고, 빠르게 구축할 수 있는 인프라가 형성되었다.

전문용어는 시대성을 반영하기 때문에 전문용어 사전의 효율성 증대를 위해서는 특정 시기에 적합한 전문용어들로 구성된 전문용어 사전을 구축하는 것이 필요하다. 따라서 전문용어 사전이 구축·활용되는 기간을 단축하는 것은 매우 중요하며, 향후 연구에서는 국방분야 전문용어의 자동 인식 및 정의문 추천, 대역어 추천 등을 포함한 전문용어 사전 구축 자동화 방안을 보장하는 것은 물론, 이런 기능들이 추가된 워크벤치를 개발하고 다른 분야에 적용할 수 있도록 범용화하는 것을 목표로 하고 있다.

참고 문헌

- [1] <http://ko.wikipedia.org/wiki/전문용어>
- [2] 최중환, 최석두, 김이겸, 박영욱, 정종희, 안희정, 정한민, 김평, “국방과학기술 전문용어 사전의 구축 프로세스 표준화 및 활용 방안”, 한국콘텐츠학회논문지, 제11권, 제8호, pp.247-259, 2011.
- [3] 국방기술품질원, “국방과학기술용어 사전”, 2011.
- [4] <https://dtims.dtaq.re.kr>

- [5] 오종훈, 최기선, “기계학습에 기반한 생의학분야 전문용어의 자동 인식”, 정보과학회논문지, 제33권, 제8호, pp.718-729, 2006.
- [6] 오종훈, 이경순, 최기선, “분야간 유사도와 통계기법을 이용한 전문용어의 자동 추출”, 정보과학회논문지, 제29권, 제3호, pp.258-269, 2002.
- [7] 이현주, 조동성, “학술 전문용어 정비 및 표준화의 특징 및 과제”, 한국어의미학회지, 제35집, pp.245-283, 2011.
- [8] 리수락, “전문용어 표준화의 원칙과 평가기준”, 정보처리학회지, 제13권, 제5호, pp.54-66, 2006.
- [9] 박창원, “전문용어 정리지집”, 국립국어원.
- [10] 황순희, 정한민, 성원경, “페이스(Facet)을 이용한 과학기술분야 시소러스 구축과 활용방안”, 정보관리연구, 제37권, 제3호, pp.61-84, 2006.
- [11] 허정, 최미란, 장명길, “한국어 개념망 구축을 위한 지능형 워크벤치”, 2005년 한국컴퓨터종합학술대회, pp.472-474, 2005.
- [12] 이경순, 김도완, 최기선, “합성명사 의미해석용 사전 구축을 위한 워크벤치”, 2000년 한글 및 한국어 정보처리 학술대회, pp.145-155, 2000.
- [13] <http://www.diquest.com>
- [14] <http://www.janes.com>

저 자 소 개

최 중 환(Jungwhoan Choi)

정회원



- 1987년 : 부산대학교 금속공학과 학사
- 1992년 : 부산대학교 금속공학과 석사
- 2000년 : 부산대학교 금속공학과 박사

▪ 1988년 ~ 현재 : 국방기술품질원

<관심분야> : 금속 열처리, 표면처리, 정보기술, 정보 분석

박 정 호(Jeongho Park)

정회원



- 1987년 : 공군사관학교 산업공학과 학사
- 1997년 : 고려대학교 산업공학과 석사
- 2011년 ~ 현재 : 국방기술품질원

<관심분야> : 시맨틱 웹, 정보검색, 정보분석

김 경 선(Kyungsun Kim)

정회원



- 1997년 : 서강대학교 전자계산학과 학사
- 1999년 : 서강대학교 컴퓨터공학과 석사
- 2006년 : 서강대학교 컴퓨터공학과 박사

▪ 2006년 ~ 2008년 : 삼성전자 통신연구소

▪ 2008년 ~ 현재 : 다이렉스트 기술연구소

<관심분야> : 자연어처리, 정보검색, 시맨틱 웹

김 평(Pyung Kim)

정회원



- 1997년 : 충남대학교 전산학과 학사
- 1999년 : 충남대학교 컴퓨터학과 석사
- 2004년 : 충남대학교 컴퓨터학과 박사

▪ 2000년 ~ 2003년 : (주)엔퀘스트테크놀로지 기술이사

▪ 2004년 ~ 2012년 : 한국과학기술정보연구원 선임연구원

▪ 2012년 ~ 현재 : 전주교육대학교 조교수

<관심분야> : 시맨틱 웹, 정보검색, 정보교육