

언어자원을 활용한 스마트 국방정보서비스

이 길 섭*

◆ 목 차 ◆

- | | |
|---------------|----------------|
| 1. 서 론 | 3. 스마트 국방정보서비스 |
| 2. 언어자원의 기술동향 | 4. 맺음말 |

1. 서 론

과학기술의 급속한 발전으로 새로운 전문용어가 생성되고 있으며, 이들 전문용어에 대하여 국제, 국가, 민·군간 의사소통을 지원하기 위한 표준화가 이루어지고 있다. 그 결과물은 용어사전, 백과사전, 시소러스, 개체명, 어휘의미망 등 언어관련 자원으로 발전이 되어 정보서비스에 활용이 되고 있다.

이들 전문용어를 효율적으로 관리하기 위하여 정부에서는 국어기본법[1]을 제정하여 전문용어의 표준화를 시행하고, 국가정보화 기본법[2]으로 지식정보자원의 효율적 관리, 기본 방향, 활용, 표준화, 공동이용, 유통체계 구축 등의 기준을 제시하고 있다.

국방 분야의 방위사업법[3]에서는 연구개발 정보, 방산물자관련 국외 도입 기술정보, 절충교역에 의해 이전받은 기술정보, 산업교육기관, 연구 기관, 산업체 협력에 의한 기술정보, 그밖에 국내외 수집된 자료 등 국방과학기술정보의 관리를 명시하고 있다. 또한, 방위사업관리규정[4]을 통해 국방과학기술 용어의 표준화를 규정하고 있다.

그렇지만 대부분 전문용어는 용어사전으로 발간되어 관리되고 있다. 현재 정부, 공공, 국방, 민간 분야의 가용한 전문용어사전 30종을 분석한 결과 책자 또는 파일 형태로 제공되고 있으며, 온라인 용어사전과 모바일 용어사전은 각각 10종 및 1종에 불과하다.

또한, 전문용어는 단순히 용어사전 수준으로만 관리되고 있으며, 정보서비스에 적용을 위한 언어자원으로 활용은 백과사전 2종, 시소러스 사전 2종에 그치고 있다. 대부분의 국방정보서비스는 데이터베이스에 저장된 정보 또는 그 정보의 조합으로 얻어진 정보를 제공하고 있으며, 언어자원을 활용한 예를 찾아보기 어려운 상황이다.

반면에 학계에서는 어휘의미망[5]을 구축하여 의미처리 기반의 대규모 지식베이스 구축, 정보 검색, 기계 번역, 시맨틱 웹 등의 정보서비스에 응용을 연구하고 있다. 그리고 포털 및 언론사 등에서는 인물, 조직, 사건과 관련한 개체명 사전을 구축하여 웹서비스를 통하여 사용자에게 유익한 정보를 제공하고 있다.

따라서 본 논문에서는 전문용어에서 출발하는 언어자원을 활용하여 지식정보 검색, 추출, 분석, 번역이 가능한 스마트 국방정보서비스에 대하여 고찰하고자 한다. 이를 위하여 언어자원의 기술동향과 활용 사례를 살펴보고, 언어자원을 활용하는 스마트 국방정보서비스를 제안한다.

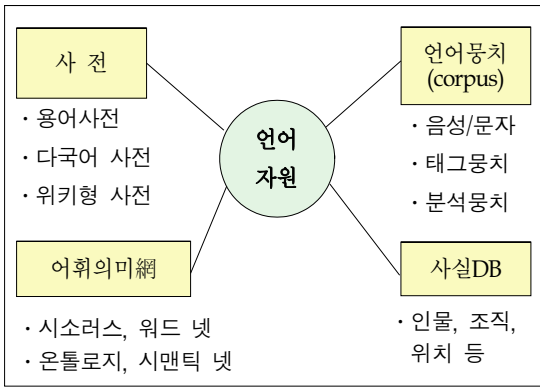
이를 위하여 이전에 사용하던 용어사전을 시소러스, 온톨로지 등 언어자원으로 개념을 확장하였다. 또한, 언어자원을 표준화하여 국방정보서비스 및 일반 정보서비스에 적용하는 방안을 제시하였다. 그 결과 사용자는 보다 편리하게 고부가 정보를 획득하게 되어 업무의 생산성이 향상될 것으로 기대된다.

* 국방기술품질원

2. 언어자원의 기술동향

2.1 언어자원의 개념 및 활용

언어자원은 인간의 모든 언어 행위의 결과물을 다양한 형태로 가공 저장하여 개발한 음성언어, 문자언어, 원시·분석 말뭉치, 전자사전, 어휘망, 온톨로지 등의 언어자원[6]으로 정의하고 있다. 언어자원은 인간이 이해하는 자연언어를 기계가 이해할 수 있도록 연구하기 위하여 필요하다. 이러한 언어자원의 개념은 (그림 1)에 제시되고 있다.



(그림 1) 언어자원의 개념

일반인이 이해하기 쉬운 대표적 언어자원은 사전이다. 사전은 표제어에 대한 뜻을 제공해주는 것이 일반적이며, 다국어 표현을 지원하거나 백과사전에서 발전되어 사용자가 지식을 정의하는 위키피디아[7]처럼 개방형 지식백과사전이 존재한다.

어휘의미망(lexical network)은 어휘간의 관계를 설정하는 망이다. 시소러스는(thesaurus) 전통적인 어휘간의 관계를 나타내는 유의어 모음이다. 예를 들어 ‘유도탄’이라는 용어에 대하여 ‘미사일(동의어), ‘장거리 유도탄(하위어), ‘유도무기체계(상위어), ‘유도탄 발사대(관련어), ‘추진체(부분어)’ 등이 각각의 관계를 가지고 있는 유의어들이다.

워드 넷[8]은 명사, 동사, 형용사, 부사 등의 어휘가 인식적 동의어인 신셋(synsets) 형태로 그룹화 된 대규모의 영어 어휘 데이터베이스이다. 신셋은 개념적 의

미와 어휘적 관계로 상호 연결되어 있다. 그 결과 브라우저를 통해서 관련된 용어와 개념을 검색할 수 있다. 그리하여 워드 넷은 컴퓨터에 의한 언어학과 자연어 처리에 활용되고 있다.

온톨로지(ontology)는 공유되는 개념화 대상을 일정한 형식을 가지고 명시적으로 정의하는 것이며, 인공지능, 시맨틱 웹, 시스템 공학, 정보공학 등에서 사용되는 정보를 구조화 하는 골격이다[9]. 예를 들어 ‘중-속-과-목-강-문-계’로 분류되는 종의 관계, 용어와 용어 사이의 관계를 정형화된 기법으로 기술한 것이다. 시맨틱 넷은 어휘와 어휘의 의미적 관계를 의미한다. 예를 들어 회사와 직원은 고용관계로 연결된다.

기타 언어자원으로 언어뭉치(corpus)와 사실(fact) DB가 있다. 언어뭉치는 음성 언어와 문자 언어로 구성되며, 어휘관계, 의미관계, 분석관계 등으로 구분된다. 사실 DB는 포털 또는 언론사에서 기사에 나타난 인물, 조직, 위치 등에 대하여 상세 정보를 제공하는 데 이용된다.

이러한 언어자원은 자연어 처리를 기반으로 하는 정보 검색, 기계 번역, 언어 교육, 언어 이해 등 기술에서 응용된다. 언어자원은 더 나아가 전자도서관, 전자상거래, 전자 뉴스, 시맨틱 웹, 정보 분석, 인공지능, 지식관리, 업무시스템 등의 분야에서 활용이 된다.

2.2 언어자원의 국방분야 활용

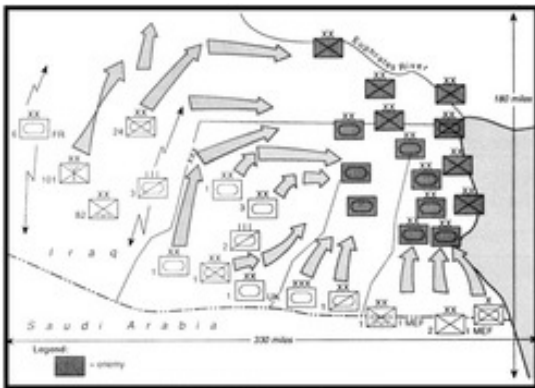
국의 국방 분야에서 언어자원의 활용은 대부분 용어사전, 정보검색 등 정보서비스 전반에 적용되고 있다. DoD Techipedia[9]는 위키피디아 형식으로 사용자가 참여하여 국방과학기술정보를 구축한 개방형 지식백과사전이다. 이를 통하여 연구자, 기술자, 과학자, 사업관리자, 군인, 정책분석가, 과학자문가, 계약자가 상호 협업이 가능하도록 지원한다.

또한, 국방 온톨로지는 국방 분야의 실체에 대한 속성, 사건, 프로세스, 관계들을 분류하여 공통적이고 공유가능한 정보로 구축하여 정보의 통합과 새로운 지식의 발견을 촉진하는 도구로 활용하고 있다. 대표적인 사례로 정보 분석(military intelligence)은 외국, 적대국, 잠재적 적대국 또는 작전지역에 대한 가용한 정보를 수집, 처리, 통합, 평가, 분석, 해석 등을 통해 작

전에 필요한 정보를 산출하는 활동이다.

여기서 온톨로지는 정보 분석 영역에서 구성 요소를 지원, 표현, 정의, 연계시키는 자원으로 활용된다. 그 결과 정보 분석기는 전장에서 획득된 이동물체의 부분적인 첩보를 통해서 무기체계의 명칭, 재원, 특징, 소속부대, 전투 배치 등 관련 상황정보를 신속하게 파악할 수 있게 된다.

정보 분석을 통해서 작전을 계획하는 장교는 (그림 2)와 같은 전장 상황도[10]를 통해서 지휘관의 작전의 도, 단위제대의 기동, 작전의 목적, 최종 상태, 기동로 등을 그래픽으로 표현할 수 있다. 여기에 나타난 그래픽 개체마다 속성정보가 연결되어 있으며 이들 간의 관계는 온톨로지에 의해 연계되어 신속한 작전계획을 수립하고 공유할 수 있게 된다.



(그림 2) 전장 상황도의 예

한편, 한국군은 분야별 전문용어사전을 통하여 국방용어에 대한 뜻을풀이를 제공하고 있다. 국방과학기술 정보서비스(DTiMS)[11]에서는 1.5만개의 전문용어의 뜻풀이와 6만여개의 시소러스를 정의하여 정보 검색에 활용하고 있으며, 메타데이터 저장소(MDR)[12]는 국방정보서비스의 상호운용성을 향상하기 위하여 전군의 데이터베이스에 사용되는 용어를 서비스 하고 있다.

또한, 빅 데이터 관점으로 방위사업 DB[13]를 구축하기 위해서는 언어자원의 필요성이 제기되고 있으며, 국방부는 ‘군 용어 순화 추진단’[14]을 구성하여 일반 및 전문용어의 순화 및 정비를 추진하고 있다. 이를

위하여 행정, 인사, 동원, 시설, 군수, 과학기술·전력, 작전 등의 분과를 구성하고 구체적인 추진방안을 마련하고 있는 상황이다.

3. 스마트 국방정보서비스

3.1 구축 개요

본 논문에서는 국방 전문용어의 순화 및 정비 차원에 머물러 있는 수준을 보급 및 활용 차원까지 확대한다. 특히, 언어자원을 활용한 스마트 국방정보서비스는 국방 전문용어를 언어자원으로 구축한 후에 응용을 위한 정보서비스이다.

먼저 국방 용어의 순화는 첫째, 일제 강점기 잔재어, 외래어, 군내 인격모독적인 비속어, 은어 등을 국어 어문규정에 맞게 정비하는 것이다. 둘째, 3군, 민·군, 연합군, 남북한, 외국어 등간의 유통성을 강화하는 것이다. 셋째, 국방 외교, 군사, 정보 분석, 작전, 군수, 방위사업, 방산수출, 교육훈련 등의 업무 분야를 원활하게 지원하는 것이다. 넷째로 장애인, 다문화 국내인 또는 외국인 불편을 해소하는 것이다.

다음으로 국방 전문용어는 국어기본법[1]에 따라 공문서, 간행물, 홍보물, 교육자료, 공식 시험자료 등에 우선적으로 활용이 되어야 하며, 신문, 방송, 잡지, 인터넷, 또는 전광판 등을 활용한 홍보 및 교육이 이루어져야 한다.

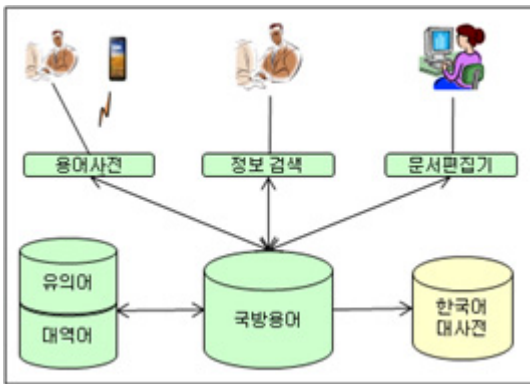


(그림 3) 스마트 국방정보서비스 개념도

국방 언어자원은 2장에서 논의 되었던 시소러스, 온톨로지 등 언어자원을 구축하는 것이다. 스마트 정보서비스는 온라인·모바일 서비스, 정보 검색·분석 서비스, 다국어 번역 서비스, SNS 분석 서비스, 사무 자동화 서비스, 국방업무지원 서비스로 구분하여 제시한다. (그림 3)은 스마트 국방정보서비스 구축 개념도를 보여주고 있다.

3.2 스마트 국방정보서비스

본 논문에서는 6개의 스마트 국방정보서비스를 제시한다. 이 중에서 온라인·모바일 용어사전, 정보 검색, 사무자동화 등 공통 스마트 국방정보서비스의 운용개념은 (그림 4)와 같다.



(그림 4) 공통 스마트 국방정보서비스 운용개념

국방용어 통합 DB는 한국어대사전과 행정, 인사, 동원, 시설, 군수·전력, 과학기술, 작전 등 국방용어사전, 국내외 전문용어사전 간에 호환성과 일관성을 유지한다. 이를 기반으로 업무적, 기술적, 용례적 관계를 고려하여 시소러스를 구축한다. 또한 다국어 번역을 지원하기 위하여 대역어 사전을 구축한다.

이러한 토대 위에서 온라인 용어사전은 하나의 용어에 대하여 다국어, 분야별 뜻을 조회할 수 있다. 또한 업무적, 기술적, 용례적 유의어를 조회할 수 있다. 이러한 용어의 조회를 통해서 얻어진 결과에서 선택되어진 용어를 이용하여 통합 정보검색을 수행할 수 있다. 더 나아가 위키피디아 형태로 개방형 지식 정보서비스를 통하여 국방 분야 이해관계자들 간의

협업을 지원할 수 있다. 이러한 서비스는 모바일 환경에서도 동일하게 제공이 가능하다.

둘째, 정보 검색 서비스는 검색 창에 주어진 키워드에 대하여 유의어와 대역어를 고려한 검색 결과를 제시한다. 예를 들어 ‘유도탄’이라는 키워드로 검색을 하는 경우에 ‘미사일,’ ‘유도탄,’ ‘○○미사일,’ ‘유도탄 발사대,’ ‘유도무기’ 등 관련되는 유의어별로 검색결과를 제시할 수 있다. 이러한 기능은 인간의 기억력의 한계를 극복하게 해준다. 사용자가 관련되는 부분 정보를 입력하면, 서비스가 관련되는 일체의 지식 정보를 제공하여 사용자는 기억하고 있지 못하던 정보에 대한 검색이 가능하게 된다.

셋째, 언어자원을 활용하는 대부분의 환경은 문서를 편집하는 상황이다. 기존의 문서편집기에서는 내장된 용어사전을 통해서 국어사전 정보를 제공한다. 이를 통하여 맞춤법 검사 등을 지원한다. 그렇지만, 국방 전문용어 등은 문서편집기 내부 사전에 등록되지 않아서 대부분 오타자로 처리되어 왔다.

본 논문에서는 문서편집기 등 사무자동화 도구에 국방용어사전을 연계하여 용어 조회, 맞춤법 검사 외에도 선택된 용어에 대하여 오용어에 대한 순화어, 다국 대역어 지원, 약어에 대한 원어, 유의어 등의 새로운 서비스가 제공 가능하다. 이를 통하여 기존의 책자형 또는 온라인 용어사전을 조회하는 번거로움을 제거할 수 있게 된다.

넷째, SNS(Social Network Service) 분석 서비스는 소셜미디어 이슈 탐지 및 모니터링 기술[15, 16]을 기반으로 주요 온라인 뉴스, 블로그, 트위터, 전문 홈페이지 등에서 국방관련 이슈를 탐지하고 추적하는 서비스이다. 이를 통하여 국내의 방산 거래, 가격, 예산, 신기술 개발, 무기체계 개발, 전력화 정보 등의 관심 정보를 추출하거나 관련 출처별 연관성과 신뢰도에 대한 분석이 가능하다.

다섯째, 다국어 번역 서비스이다. 최근 국방과학기술의 발전으로 다양하고 방대한 기술 자료들이 제공되고 있다. 특정 전문분야에 익숙한 독자에게는 원문이 편리할 수 있겠으나 대부분의 독자들은 한국어로 번역을 통해서 더 많은 정보를 정확하게 습득하게 된다. 대부분의 상용 번역 서비스는 한·영, 영·한, 일·한, 한·일 등 단방향 일대일 대응 방식으로 서비스가 제

공된다. 장기적으로 다국어에 대한 대역어를 워드 넷 [8] 개념으로 구축하여 양방향 n : m 방식의 번역 서비스의 개발이 요구된다.

여섯째, 국방업무지원 서비스는 국방 전 분야에 해당이 된다. 예를 들어 국방기술기획 분야에서는 인터넷상에서 획득된 기술 정보를 통하여 세계 각국의 기술 수준, 개발 동향, 제품 제원 및 거래 정보 등을 분석할 수 있다. 또한, 앞의 2장에서 소개된 정보 분석과 작전계획 수립 등의 서비스를 제공할 수 있다. 이러한 서비스들에 대해서는 시범사업의 추진 또는 각 분야의 서비스 공모를 통하여 보다 구체적이고 실현 가능한 서비스를 구축할 수 있을 것이다.

3.3. 기대효과

본 논문에서는 6개의 스마트 국방정보서비스 영역을 제시하였다. 이들 서비스 중에서 앞의 세 가지는 2012년에 구축을 추진하여 2013년에 서비스를 제공할 예정이다. 나머지 세 가지는 장기적인 목표를 가지고 추진이 요구된다. 이들 각각의 서비스에 대한 기대효과를 정량적으로 제시하는 것은 별도의 연구가 필요한 사항이다.

한편, 앞서 진행한 전문용어에 관한 정책연구[17]에 의하면 국민 1인당 연 평균 5.93회로 용어사전을 찾아보고 있으며, 용어의 의미를 해결하는 데 소요되는 시간은 52~660초로 제시하고 있다. 이러한 기초자료를 근간으로 국방 분야는 연 21.8억원, 방산, 조달, 교육관련 민간 분야는 57억원이 절감되는 것으로 산정된다.

다음으로 국방용어 통합 DB 구축을 통하여 약 120만 어휘에 해당하는 국방용어를 국어 어문규범과 한 국어대사전과 호환성·일관성을 유지하게 되어 국민의 올바른 국어사용에 기여하게 된다. 또한, 정보 검색, 사무자동화 도구에 적용 등을 통하여 언어자원의 실생활화에 접목되어 국방용어의 보급 및 활용이 실제적으로 가능하게 된다.

SNS 분석 서비스를 통해서도 온라인상에 가용한 정보를 통해서 관심 정보의 획득이 가능하다. 또한 발표된 국방 정책에 대하여 온라인상에서 긍정, 중립, 부정적인 반응의 분포와 구체적인 그 사유를 분석하여 정책에 피드백 할 수도 있다. 다국어 번역 서비스

는 국방 외교, 방산수출, 선진 기술 및 교육 등에 필요한 정보를 온라인상에서 수집이 가능하게 한다. 이를 위하여 다국어 번역 서비스에 대한 지속적인 투자가 요구된다.

국방업무지원 서비스는 궁극적으로 언어자원이 활용되어야 하는 분야이다. 국방 행정, 인사, 동원, 군수, 전력, 과학기술, 군사 분야에 대하여 언어자원을 근간으로 하는 지식정보의 구축을 통하여 각 분야의 업무 생산성과 효율성의 향상이 기대된다.

4. 맺음말

과학기술의 급속한 발전으로 국방 분야의 다양한 전문용어가 생성되고 있다. 이들을 정보서비스에 적용하기 위하여 시소러스 또는 온톨로지 등의 언어자원의 개념을 소개하였다. 현재 한국군의 국방정보서비스는 대부분 데이터베이스를 활용하여 구축 정보 또는 그 저장된 정보를 이용한 연산 결과 정보를 제공하고 있으며, 언어자원을 효과적으로 활용하고 있지 못하는 실정이다.

따라서 본 논문에서는 국방 언어자원을 표준화 하고 통합 DB를 구축한 이후에 확보 가능한 온라인·모바일 용어사전, 정보 검색, 사무자동화 도구, SNS 분석, 다국어 번역, 국방업무지원 등 6개의 스마트 국방정보서비스 영역을 제시하였다. 이 중에서 앞의 3가지는 구축을 통하여 서비스를 준비 중에 있으며, 나머지 3가지는 중·장기적 목표를 가지고 추진 중이다.

본 논문의 제안을 통하여 국방 전문용어의 보급 및 활용체계를 정보서비스를 통하여 실제 업무에서 활용이 가능하도록 획기적인 개선이 이루어지게 되었으며, 연 78억원 규모의 비용 절감효과도 발생하는 것으로 추정되고 있다. 또한 국방용어의 표준화 및 활용으로 올바른 국어 생활에 대한 지원과 국방 업무지원을 통하여 생산성 향상도 기대되고 있다.

참고 문헌

- [1] 법제처, 국어기본법 제17조 (전문용어 표준화), 법률 제11424호, 2012. 5월.

- [2] 법제처, 국가정보화 기본법 제25조 (지식정보자원의 관리), 법률 제10629호, 2011년 5월.
- [3] 법제처, 방위사업법 제31조 (국방과학기술의 관리), 법률 제10907호, 2012년 1월.
- [4] 방위사업청, 방위사업관리규정 제650조 (국방과학기술정보의 관리), 방위사업청 훈령 제188호, 2012년 8월.
- [5] 옥철영, 한국어 어휘망 국방 분야 적용 방안, 국방기술품질원 민군규격용어 표준화 세미나, 2012. 11월.
- [6] 옥철영, 한국어 워드넷 개발과 상위 온톨로지 - 사용자취지능망(U-WIN)을 중심으로, 제1회 언어 중립적 온톨로지 워크숍, 2012년 6월.
- [7] 위키피디아, <http://www.wikipedia.org/>, 2012년.
- [8] Princeton University, WordNet - A lexical database for English, <http://wordnet.princeton.edu>, 2010.
- [9] Defense Technology Information Center, DoD Techipedia, <https://www.dodtechipedia.mil/>, 2012.
- [10] Williams S. Mandrick, Military Ontology, <http://militaryontology.com/>, 2012.
- [11] 국방기술품질원, 국방과학기술정보서비스 (DTIMS: Defense science & Technology Information Management Service), 2012년.
- [12] 국군통신사령부, 메타데이터 저장소 (MDR: Meta Data Registry), 2012년.
- [13] 이길섭, “빅데이터 관점의 방위사업 DB 구축 및 활용,” 제8회 국방기술학술대회, 서울시립대학교, 2012년 7월.
- [14] 국방부, 군 용어 순화 종합대책, 2012년 12월.
- [15] 류법모, 김현진, 김현기, 박상규, “심층 언어분석 기반 소셜미디어 이슈탐지 및 모니터링,” 정보과학회지, 제30권, 제6호, 통권 제277호, 2012년 6월, pp. 47-58.
- [16] 국방기술품질원, 실시간 국방 이슈 및 동향 정보 분석, 연구 보고서, 2012년.
- [17] 문화체육관광부, 전문용어 표준화 제도 정책연구보고서, 2011년.

● 저 자 소 개 ●



이길섭

1985년 금오공과대학 전자공학과(공학사)
 1987년 한국과학기술원 전산학과(공학석사)
 1992년 한국과학기술원 전산학과(공학박사)
 1990년~2003년 국방부/육군 정보화부서 근무
 2003년~2006년 국방대학교 국방과학학부 교수
 2006년~2011년 방위사업청 전산정보관리소장
 2011년~현재 국방기술품질원 기술정보센터 근무
 관심분야 : 정형기법, 온톨로지, 시멘틱 웹, 정보 분석