

S&T Text Mining을 이용한 국방 유망기술 식별에 관한 연구

(A Study on the Identifying Emerging Defense Technology using S&T Text Mining)

이 태 봉(Tae-Bong Lee)*, † 이 춘 주(Choonjoo Lee)**

초 록

본 연구에서는 과학기술 텍스트 마이닝을 이용하여 국방 유망기술을 식별하는 방법론을 제안하고 있다. 그동안 국가차원에서 NTIS와 DTiMS를 포함한 과학기술 관련 정보체계를 구축하는데 많은 노력을 기울여왔는데 과학기술 정보체계는 연구자와 정책입안자, 또는 실무자들이 기술적 변화를 분석하고 효율적인 업무진행, 지식공유, 전략개발, 또는 조직의 경쟁력을 높이기 위한 정책개발에 활용성이 크다. 본 연구에서는 INSPEC 데이터베이스에 과학기술 텍스트마이닝 기법을 적용하여 미래 무인전투기술에 대한 지식네트워크 구조와 국방 유망기술을 식별하는 과정을 예시함으로써 구축된 과학기술 정보체계를 이용한 미래 유망기술의 식별 방법론을 제시하였다.

ABSTRACT

This paper tries to identify emerging defense technology using S&T Text Mining. As a national agenda, there has been much effort to build S&T information systems including NTIS and DTiMS that enable researchers, policy makers, or field users to analyze technological changes and promote the best policy practices for efficient workflow, knowledge sharing, strategy development, or institutional competitiveness. In this paper, the S&T Text Mining application to unmanned combat technology using INSPEC DB is empirically illustrated and shows that it is a feasible approach to identify emerging defense technology as well as the structure of knowledge network of the future technology candidates.

Keywords : 유망기술(Emerging Technology), 무인전투체계(Unmanned Combat System), 국방기술정보(Defense Technical Information), S&T Text Mining, Wikipedia

논문접수일 : 2009년 11월 27일 논문게재확정일 : 2010년 4월 13일

* 해군 진해 00전단, 소령(진)

** 국방대학교 국방과학학부 조교수

† 교신저자

1. 서론

2006년 이후 국가예산이 투입되는 연구개발의 성과는 국가과학기술 표준성과지표에 의해 관리가 되고 있다. 이와 관련하여 연구개발의 성과제고를 위한 정보체계 인프라가 구축되어 활용되고 있는데 국가과학기술지식정보서비스(NTIS: National Technical Information Service)와 국방기술정보통합관리체계(DTiMS: Defense Technology information Management System)가 대표적인 과학기술관련 정보체계 인프라라고 할 수 있다.

국방과학기술정보의 내용은 크게 기술정보, 기술관리정보, 무기체계 정보로 구분할 수 있다. 기술관리정보는 기술의 획득 및 관리에 필요한 현황 자료로 기술분류, 기술수준, 기술의 소재, 관련 부품 및 무기체계, 관련 민수 분야 기술, 중요도, 개발경위, 관련 프로젝트, 기술 획득방법 및 기술원천 등의 자료와 국내외 과학기술자의 전공, 학력, 연구경력, 소속 등 과학기술자 정보를 포함한다. 무기체계 정보는 무기체계 성능·제원·비용·보유·운용현황 및 거래현황 등에 관한 자료와 국내외 방산시장 및 업체현황 자료를 포함한다[1].

국방과학기술활동에 의한 산출물들은 DTiMS 상의 ‘국방기술기획정보DB’ 및 ‘국방연구개발정보DB’로 구축되어 각각 ‘국방기술기획정보서비스’ 및 ‘국방연구개발정보서비스’로 수요자에게 제공된다. 2009년 현재 국방인트라넷 망에서 제공이 되며 2010년경 인터넷으로 서비스를 제공할 계획에 있다.

국방기술기획정보DB는 국방기술기획문서정보, 기술수준조사분석정보, 국방기술정책정보, 국방기술동향정보로 구성된다. 기품원에서 관리하는 해외기술정보, 해군의 해군동향정보, 공군의 해외공군 무기체계 개발동향, 합참의 무기체계/군사기술 발전동향, 국과연의 해외과학기술, 국방과학기술플러스, 국방신기술동향분석, 기술보고서 등이 국방기술기획DB로 연계된다.

국방연구개발DB는 연구개발사업, 기술세부현

황, 기술자료, 성과물, 자료실로 구성된다. 연구개발 성과물인 논문, 연구보고서, S/W, 특허 등의 지식재산권 자료가 국방연구개발DB로 연계된다.

이처럼 DTiMS의 등장으로 국방과학기술정보를 체계적으로 관리·유통할 수 있는 인프라가 구축되었다. 하지만 수집된 정보의 가공 및 분석을 통한 새로운 지식창출체계로서 발전되어야 할 것이다.

현재 DTiMS에서 제공되고 있는 국방기술정보는 국방기술기획 관리자에게는 단편적이어서 기술기획업무에 즉각적인 이용이 불가하다. 즉, 효과적인 기술기획을 위해서는 분석정보를 이용하여 세계적인 연구개발 동향을 파악하고 전반적인 지적구조를 그려 부상하고 있는 기술을 분석함으로써 미래의 기술 발전추세를 예측할 수 있어야 한다.

따라서 국방과학기술정보가 점점 대용량으로 축적될수록 국방과학기술정보의 분석은 DB로부터 지식탐사 과정을 거쳐 의미있는 지식을 창출할 수 있어야 한다.

본 논문에서는 과학기술연구정보 분석방법론을 개괄하고, 국방유망기술 식별을 위한 프로세스를 4단계로 제안하며 지상무인전투체계기술에 대한 실증분석을 통해 국방과학기술 연구정보의 활용을 통한 과학기술혁신의 기회를 탐색하고 기술정책적 시사점을 제시하고자 한다.

본 논문의 순서는 먼저 2장에서 국방유망기술 식별을 위한 국방과학기술정보분석 프로세스를 제시하고, 3장에서 무인전투차량기술에 대한 실증분석을 통해 제시한 방법론의 실효성을 확인하고 4장 결론에서 정책적 시사점을 제시하는 것으로 구성하였다.

2. 국방과학기술정보분석을 통한 유망기술 식별 방법론

2.1 과학기술정보분석 방법론

과학기술정보분석을 위한 방법론으로서 biblio-

metrics(또는 scientometrics)의 인용분석과 최근 활발히 연구가 되는 co-word 분석이 있다. 기존에는 과학기술DB에서 제공하는 field 데이터를 이용하여 과학기술정보분석이 이루어졌지만 최근에는 텍스트마이닝 기술의 발달로 인하여 과학기술 문헌의 초록 및 문헌전체의 내용을 분석대상으로 삼기도 한다.

2.1.1 인용분석

인용분석(citation analysis)은 bibliometrics 영역의 하나로서 하나 이상의 주제분야에서 각각의 정보생산자가 인용하거나 참조한 각종 정보원들을 조사 분석하는 방법으로 정보생산자에 의하여 인용된 정보원들의 특성을 밝히고자 하는데 그 목적이 있다. 정보의 속보성이 다른 분야에 비하여 빠르고 정보수명이 상대적으로 짧은 과학기술분야에서 정보수집을 위한 정책과 방향을 수립하고, 특정분야의 이용형태를 파악하는데 인용분석이 높은 유용성을 제공하기도 한다.

과학기술계는 다양한 연구주체 사이의 정보교환에 의하여 움직인다. 이러한 정보교환의 매개체 중에서 중요한 역할을 하고 있는 것이 과학기술 문헌이며, 연구주체들은 다른 연구주체의 연구결과를 자신의 문헌에서 인용함으로써 그 가치를 부여하게 된다. 특정 문헌이 다른 문헌에서 많이 인용된다는 사실은 그 문헌이 과학기술적으로 우수하거나 영향력을 가진다는 것을 의미하므로, 과학기술정보분석에서 인용 건수를 분석하는 것은 충분한 의미를 가진다. 그러나 SCI를 제외한 COMPENDEX, INSPEC 등의 대부분의 학술데이터베이스 및 국내의 특허데이터베이스는 이러한 데이터를 정형화하여 제공하지 않으므로 인용분석시 어려움이 따른다.

한편 인용분석은 문헌인용건수, 특허인용건수 등의 통계적인 분석과 저자인용네트워크, 기관 및 국가 인용네트워크 등의 사회연결망 분석 등으로

구분할 수 있다.

2.1.2 co-word 분석

인용분석은 과학기술문헌에 내포되어 있는 인 용자료들에 대한 분석을 근간으로 이루어지며 인 용정보의 형태, 언어, 인용선호도, 정보이용현상을 파악하는데 유용하지만 논문을 이용한 연구주체와 직접적인 연관성을 찾는 데 활용하기에는 한계가 있다.

반면에 co-word 분석은 텍스트에서 제시된 특정 영역 안에서 아이디어간의 관계를 확인하기 위해서 텍스트의 언어 부분에서 아이템쌍이 동시에 출현하는 패턴을 사용하는 콘텐츠 분석기법으로서[2] 연관성 분석의 문제를 간결하게 처리할 수 있는 잠재력을 가지고 있다.

co-word 분석에서 핵심적인 가정은 단어들은 연구개발자가 연구를 수행하면서 사용한 개념을 나타내고, 연구개발자가 단어들을 한 문헌에서 동시에 연결하여 사용한다는 것은 곧 그러한 개념들이 연관되어 있음을 뜻한다는 것이다. co-word 분석은 일반적으로 데이터 선정, 유사도 행렬 작성, 클러스터링, 데이터 매핑 단계의 절차를 거친다.

2.1.3 Data Mining과 S&T Text Mining

과학기술정보분석은 정보학의 발전에 의해 새로운 전기를 맞고 있다. 데이터마이닝이라 불리는 새로운 기법에 의해 수많은 정보나 데이터베이스로부터 획득한 데이터로부터 유용한 가치를 가진 새로운 지식을 도출할 수 있게 되었다. 데이터마이닝은 1990년대 초부터 소개되기 시작한 대량의 데이터로부터 유용한 지식을 효과적으로 찾아내는 지식 탐사의 한 분야이다. 데이터마이닝은 “대용량의 데이터로부터 이들 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화함으로써 유용한 지식을 추출하는 일련의 과정들”

이라고 정의할 수 있으며, 이는 하나의 분석기법을 의미하는 것이 아니라 여러 기법과 방법들의 적절한 조합으로 이루어진 일련의 과정이라고 말할 수 있다.

과학기술정보분석에 있어서 데이터마이닝은 새로운 영역을 보여주고 있으나 수치데이터보다 텍스트데이터가 일반적인 기술정보분석에서는 데이터마이닝의 한 분야인 텍스트마이닝(Text Mining)이 더 활용되고 중요성이 부각된다. 텍스트마이닝은 전세계적인 인터넷의 보급에 따른 효율적인 검색엔진을 개발하는데 사용되어 왔다. 텍스트마이닝은 대량의 자유롭고 비구조화된 텍스트에 적용되는 절차로서 전문, 요약, 또는 색인된 용어와 같은 형태로 문서를 탐색하고 문서간의 복잡한 관련성을 탐색한다. 미국에서는 텍스트마이닝을 민간항공사의 사고리पोर्ट를 통해 결함패턴을 찾아내거나, 경찰의 범죄리पोर्ट를 통해 범죄유형 및 추세를 분석하는데 활용하였다.[4]

경찰은 지역 및 시간과 특정범죄유형이 상관관계에 있다는 것을 밝혀내어 순찰스케줄을 수립하는데 활용하였다. 미국의 911테러 이후 최근에는 e-mail의 키워드를 텍스트마이닝 기법으로 추출하여 분석함으로써 테러징후를 포착하는데 활용하고 있다.

최근에는 텍스트마이닝에 대한 학계적·실무적 관심이 급증하면서 전산기술이 발달함에 따라 텍스트마이닝을 데이터마이닝과 별도의 학문으로서 취급되고 있는 실정이다. 특별히 과학기술정보 분석을 위해서 사용하는 경우를 ‘과학기술 텍스트마이닝(S&T Text Mining)’이라고 표현한다.

2.2 국방 유망기술 식별과정

본 논문에서 제안하는 S&T 텍스트마이닝을 이용한 부상기술 식별을 위한 국방과학기술정보 분석의 프로세스는 자료수집, 자료정제, 기술정보분석, 부상기술 식별의 4단계로 구성된다.

1단계 자료수집에서 분석의 대상이 될 기술을 선택하고 자료수집 대상 DB를 결정한다. 자료수집 대상 DB는 해당기술의 특성에 따라 학술데이터베이스, 특허데이터베이스, 상용검색엔진, 국방과학기술 관련 전자발행물, DTiMS DB 등이 될 수 있다. 만약 해당기술에 대한 전문지식이 부족한 상태에서 분석을 한다면 자료수집단계와 자료정제단계를 여러 번 거쳐야 한다.

2단계 자료정제에서는 텍스트마이닝을 이용하여 수집자료에서 키워드를 추출하고 이중에서 제품, 성능, 기술적으로 의미를 갖는 키워드를 별도로 분류한다. 그리고 이후의 통계분석을 위하여 텍스트마이닝의 대상이 되는 종속변수를 제외하고 텍스트로 된 독립변수들을 숫자로 변환한다.

3단계 기술정보분석에서는 수집자료를 국가별, 기관별, 연도별로 분류하여 일차적인 통계분석을 실시함으로써 해당기술에 대한 개괄적인 연구개발동향을 분석하고, 필요시 인용분석을 통하여 국제협력관계, 저자네트워크, 국가별/기관별 영향력 등을 분석한다. 텍스트마이닝으로 추출한 키워드들의 증감양상을 분석하여 부상기술(Emerging Technology)을 탐색하고, 부상기술과 타기술간의 네트워크 분석을 통하여 부상기술의 지적구조를 분석한다.

그리고 마지막 유망기술 식별단계에서는 기술정보분석단계에서 도출한 부상기술 분석결과를 바탕으로 글로벌 온라인 백과사전인 위키피디아(Wikipedia)를 활용하여 유망기술을 식별한다.

3. S&T Text Mining을 이용한 국방 유망 기술 식별 실증연구

3.1 적용대상

제안하는 유망기술 식별 방법론은 적용 대상을 제한하지는 않으나 본 연구의 실증분석 대상기술은 미래전 양상에 대비하여 선진국을 중심으로 개발중

인 무인전투체계, 그중에서도 지상무인전투체계로 한정하였다. 구체적으로 2007년 국방과학기술조사서에서 분류한 지상무인전투체계의 핵심대상기술 [7] 중에서 선진국 대비 국내기술수준이 비교적 낮은 ‘지형인지장치기술’을 분석대상으로 정하였다.

3.2 자료수집 및 정제

3.2.1 자료수집 DB 선택

지형인지장치기술은 국방과학연구소에서 분류한 국방핵심 기반기술 11개 분야 중 ‘센서·신호처리 분야’에 해당하는 기술로서 민간분야 및 국방분야에 공통적으로 적용 가능한 민군겸용기술의 성격이 강하고, 특히 군보다는 민간의 산·학·연에서 선도하고 있는 첨단기술이다.

분석대상기술의 특성을 볼 때 학술데이터베이스 및 특허데이터베이스의 선택이 가능하나, 세계적으로 국방과학기술과 관련된 산출물은 특허로 등록하지 않는 경향을 고려하여 학술데이터베이스를 자료수집 DB로 선택하였다.

학술데이터베이스중에서 과학논문 선정에 있어 SCI 등재 학술지의 50% 정도가 의·약학과 생물·화학으로 편향되어 있고, 비교적 역사가 짧은 정보통신 분야는 등재 학술지가 적다. IT 기술에 관한 과학논문은 IEEE 계열의 학술지 중심으로 구성된 INSPEC이 SCI 보다 권위를 갖는다[5]. 이를 고려하여 본 연구에서는 INSPEC을 자료수집 DB로 선택하였다.

3.2.2 자료수집

(가) 1차 자료수집

지형인지장치기술에 대한 비전문가가 DB내에서 지형인지장치기술과 관련 있는 문헌정보를 검색하는 것은 쉽지 않다. 자율주행기술이 신기술로서 기술의 융합화·복합화가 빠르게 진전되고 있

기 때문에 지형인지장치기술과 관련된 문헌정보를 추출하기 위한 query keyword 식별을 위하여 먼저 자율주행기술과 관련된 문헌정보를 추출하였다.

INSPEC DB내에서 query keyword를 “autonomous navigation”으로 하여 검색결과에서 지상무인체계기술의 범주에 속하지 않는 AUV*, UUV*, underwater, sea, maritime, UAV*, aerial, flight*은 제외하였고, 검색 field는 제목(title), 주제어(keyword), 초록(abstract)으로 한정하였으며, 검색대상문헌은 저널논문으로 한정하였으며, 검색기간은 과거 10년의 자료인 1997년~2007년으로 설정하였다. 검색결과 총 922건의 저널논문을 수집하였다.

(나) 1차 텍스트 마이닝

수집한 저널논문 992개의 초록을 대상으로 텍스트마이닝 도구인 WordStat을 이용하여 초록내에 등장하는 단어들의 빈도수를 추출하였다. 이때 WordStat의 단어 추출 설정값(threshold)을 5개 이상의 서로 다른 문헌에서 7번 이상 출현하는 단어로 설정하였다. 이러한 방법으로 텍스트마이닝을 수행한 결과, 총 344개의 키워드를 수집하였다. 지형인지장치기술 관련 참고문헌 조사 및 해당기술 전공자의 자문을 거쳐 지형인지장치기술과 관련이 높은 20개의 키워드를 선별하였다.

(다) 2차 자료수집

1차 텍스트마이닝에서 선별한 20개의 키워드를 query keyword로 사용하여 1차 자료수집과 같은 방법으로 INSPEC DB내에서 검색 field를 제목(title), 주제어(keyword), 초록(abstract)으로 한정하였고, 검색대상문헌은 저널논문으로 한정하였으며, 검색기간은 과거 10년의 자료인 1997년~2007년으로 설정하였다.

총 2,698개의 수집한 저널논문은 EndNote의 ‘find duplicate’ 기능을 이용하여 저자, 발행연도,

제목이 일치하는 중복문헌 11개를 삭제하였다.

(라) 2차 텍스트 마이닝

2,687개의 저널논문의 서지정보를 WordStat으로 텍스트마이닝을 수행한 결과, 총 287개의 키워드를 수집하였다. 지형인지장치기술 연구개발을 담당하는 국방과학연구소 및 국방기술품질원 소속 연구원 3명에게 자문을 구한 결과, 지형인지장치기술과 직접적으로 관련이 높은 키워드 154개를 선별하였다.

3.3 부상기술 탐색

3.3.1 포트폴리오 맵 분석

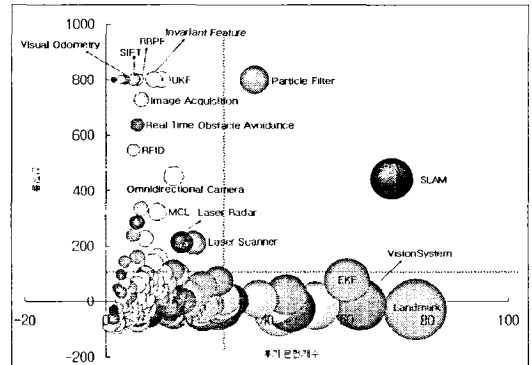
부상기술을 식별하기 위하여 지형인지장치기술 관련 154개의 키워드가 출현하는 문헌들의 수를 연도별로 집계하여 신장률을 다음과 같은 식을 통하여 계산하였다.

$$\text{신장율} = \frac{\text{후기문헌개수} - \text{전기문헌개수}}{\text{전기문헌개수}} \times 100$$

<그림 1>은 지형인지장치기술 키워드 포트폴리오 맵을 보여주고 있다. <그림 1>에서 키워드의 원형 크기는 키워드가 등장하는 전체문헌수(1997~2007년)의 빈도수를 나타내고, x축에 평행한 점선은 신장율의 평균인 82%와 교차하는 지점이며, y축에 평행한 점선은 후기 문헌개수의 평균인 28과 교차하는 지점이다. 점선에 따라 제 1, 2, 3, 4사분면으로 구분되고 있는데 이를 통하여 기술의 발전형태를 예측할 수 있다.

제 1사분면은 논문 발행건수가 많고 신장율이 높은 키워드로서 지형인지장치기술 중 발전기에 위치하는 기술이다. 제 2사분면은 논문 발행건수는 작으나 신장율이 높은 키워드로서 성장기에 위치하는 기술이다. 제 3사분면은 두 가지 측면에서 해석할 수 있다. 먼저 제 3사분면의 원편에 위치하면서 신장율이 높은 키워드는 성장기로 이동

할 가능성이 있는 탐색기의 기술이고, 오른쪽에 위치하면서 신장율이 음수의 값을 가지는 키워드는 성숙기를 지나서 쇠퇴하고 있는 기술이다. 제 4사분면은 논문의 발행개수는 많으나 신장율이 저하되는 키워드로서 성숙기에 위치하는 기술이다. 포트폴리오 맵을 통하여 기술의 수명주기의 양상은 제 3사분면의 원편 상단에 위치하는 키워드로부터 시작하여 제 2사분면, 제 1사분면, 제 4사분면을 지나 다시 제 3사분면의 마이너스 신장율의 위치로 이동할 것이라고 예상할 수 있다.



<그림 1> 지형인지장치기술 키워드 포트폴리오 맵

3.3.2 부상기술 키워드 및 후보 키워드 식별

<그림 1> 포트폴리오 맵에서 제 2사분면에 위치하는 키워드를 지형인지장치기술의 ‘부상기술(emerging technology) 키워드’로 식별하였다. 또한 제 1사분면에 위치하며 양의 신장율을 가지는 키워드를 성장 가능성이 있는 ‘부상기술 후보 키워드’로 식별하였다. 이렇게 식별한 키워드는 부상기술 28개와 부상기술후보 43개를 포함한 71개가 도출 되었다.

3.4 부상기술 식별

3.4.1 부상기술 후보군 검색

앞 절에서 도출한 부상기술 키워드 및 후보 키

워드는 기술용어를 표현하는 관련 키워드를 모두 포함한 것이다. 따라서 <표 1>의 부상기술 키워드 및 후보 키워드 리스트를 이용하여 Wikipedia

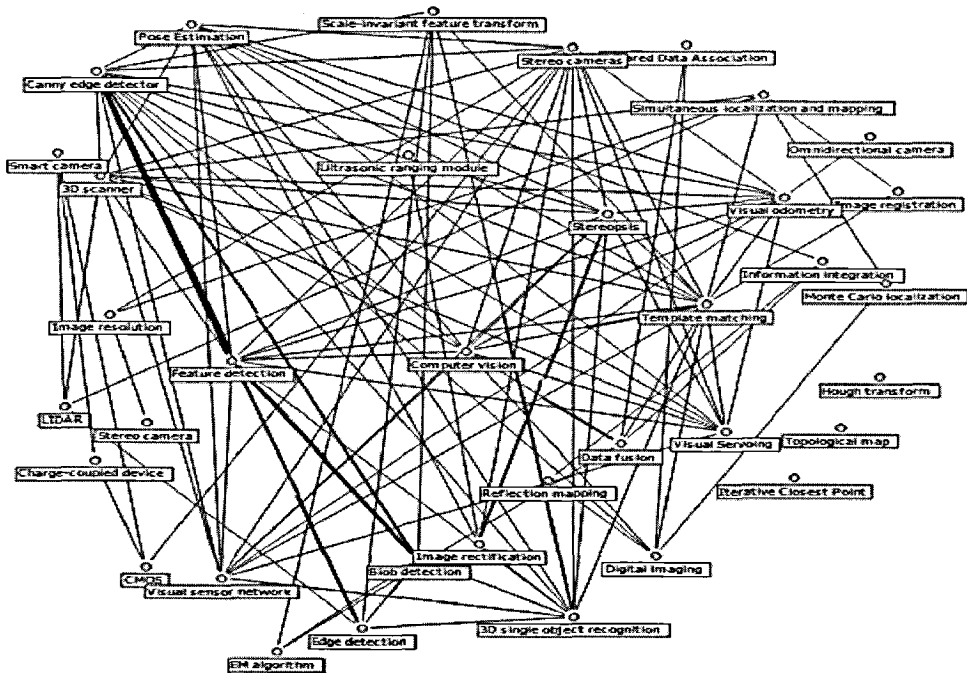
<표 1> 부상기술 후보군

3D scanner, reflection mapping, LIDAR, smart camera, visual sensor network, image resolution, visual odometry, image registration, monte carlo localization, SIFT, ICP, digital image acquisition(digital imaging), EM, 3D single object recognition, stereopsis(stereo vision), data fusion, control theory, visual servoing, topological map, computer vision, pose(pose estimation), infrared data association(IrDA), hough transform, digital image processing, charge-coupled device(CCD), stereo camera, stereo cameras, SLAM, digital signal processor, digital signal processing, omnidirectional camera, automatic data processing, CMOS, edge detection, edge detector, information integration, feature detection, blob detection, image rectification, ultrasonic ranging module, template matching

사전 검색을 통해 기술용어로 변환하는 과정을 거쳤다. 키워드가 Wikipedia 'article'에 등록되지 않는 경우에는 이 키워드를 포함하며 지형인지장치 기술과 가장 관련이 높은 'article'명을 부상기술 후보군으로 지정하였다. 이렇게 도출한 부상기술 후보군은 40개로 <표 1>과 같다.

3.4.2 자료수집

앞에서 도출한 부상기술 후보군 리스트를 이용하여 Wikipedia에서 순차적으로 검색을 실시하였다. 먼저 검색결과 창에서 'article' 페이지에 있는 'See Also' 탭의 연관키워드들을 목록화 하였다. 다음으로 'history' 페이지에서 제공하는 'Revision history statistics' 기능을 이용하여 총 편집회수, 연도별 편집회수, 최초편집일, 최종편집일, MTBE, 총 편집자수, 연평균 편집자수 등을 목록화 하였다.



<그림 2> 부상기술 후보군 네트워크 구조

3.4.3 연관키워드 네트워크 분석

Wikipedia를 통해 식별한 부상기술 후보군의 연관키워드는 총 111개이다. 부상기술 후보군간의 기술적 연관관계를 분석하기 위하여 먼저 부상기술 후보군과 연관키워드의 연관행렬을 작성하였다. 다음으로 부상기술 후보군끼리 같은 연관키워드를 공유하는 개수를 집계하여 <표 2>와 같은 연관행렬을 작성하였다. 이 연관행렬에서 행과 열이 교차하는 지점의 숫자는 해당 행과 열의 키워드가 같은 연관키워드를 가지는 개수를 의미한다.

이를 도식화하기 위하여 Cyram사의 'NetMiner 3'를 이용하여 네트워크 분석을 수행하였다. <그림 2>는 부상기술 후보군들의 상호 기술적 연관관계를 보여준다. <그림 2>에서 노드를 연결하는 링크의 굵기는 각 노드간의 기술적 연관성이 높다는 것을 의미한다. feature detection과 canny edge detector간의 연관성이 가장 높다. 반대로 hough transform, topological map, iterative closest point와 같이 고립된 노드들이 있는데 이들은 타 기술과의 연관성이 가장 낮은 기술들이다. 이

<표 2> 부상기술 후보군 연관행렬

구분	3D scanner	CMOS	Blob detection	...
3D scanner				
CMOS			1	
Blob detection		2		
Charge-coupled device				
3D single object recognition				1
Computer vision	1			
Digital imaging				
Edge detection		1	3	
Canny edge detector	1	1	3	
EM algorithm			1	
Feature detection	1	1	4	
...				

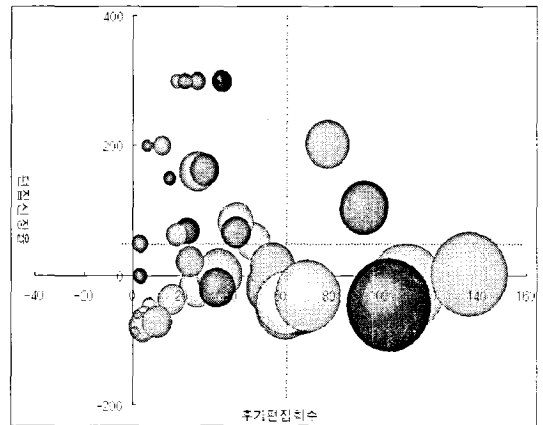
들은 지형인지장치와 직접적인 연관성은 부족하고 지형인지장치의 후행기술로 이루어지는 map building과 관련된 기술용어이다.

3.4.4 편집추세 분석

부상기술 후보군별로 Wikipedia의 'Revision history statistics'를 이용하여 'article' 페이지의 편집 추세를 분석하였다. 부상기술을 도출하기 위한 방법으로 포트폴리오 맵 분석을 시도하였는데, 편집 신장율은 아래와 같은 식을 이용하여 계산하였다. 전기편집회수는 2006년 11월 ~ 2007년 10월 기간 동안의 편집회수를, 후기편집회수는 2007년 11월 ~ 2008년 10월 기간 동안의 편집회수를 의미한다.

$$\text{편집신장율} = \frac{\text{후기편집회수} - \text{전기편집회수}}{\text{전기편집회수}} \times 100$$

기술용어가 Wikipedia에 처음 등록된 일자가 서로 상이하므로 편집이 발생한 기간도 달라지게 된다. 따라서 편집기간을 표준화하기 위하여 부상기술용어에 대한 편집은 최근에 들수록 점점 증가할 것이라는 가정으로 분석대상 기간을 최근 2년으로 설정한 것이다.



<그림 3> 부상기술 후보군 포트폴리오 맵

앞에서 <그림 1>을 작성한 방법과 마찬가지로 편집회수에 대한 포트폴리오 맵을 <그림 3>과 같이 도식화하였다. Fig. 3. 에서 점선 축을 중심으로 제 1, 2, 3, 4분면으로 구분하여 <표 3>과 같이 정리하였다.

<표 3>에서 제 1사분면에 있는 후보군들은 발전기에 있는 기술용어로, 제 4사분면에 있는 후보군들은 성숙기에 있는 기술용어로 간주한다. 따라서 제 2, 3사분면에 있는 부상기술 후보군 중에서 부상기술을 도출한다.

<표 3> 부상기술 후보군 분류

사분면	부상기술 후보군
1	Edge Detection, SIFT
2	3D single object recognition, Active pixel sensor, blob detection, feature detection, image registration, Iterative Closest Point, Monte Carlo localization, omnidirectional camera, Pose Estimation, stereo camera, template matching, topological map, visual odometry, visual sensor network, Visual Servoing
3	Articulated body pose estimation, Automatic data processing, canny edge detector, digital imaging, EM algorithm, hough transform, image rectification, information integration, Infrared Data Association, Reflection mapping, Sensor fusion, SLAM, Smart camera, stereo cameras, stereo vision, ultrasonic ranging module
4	3D scanner, CCD, computer vision, digital image processing, digital signal processing, DSP, image resolution, LIDAR

제 2, 3사분면에 있는 부상기술 후보군들의 Wikipedia 최초등록일을 조사하여 2005년 1월 1일 이후에 등록된 것을 부상기술로 식별하였다. 이렇게 최종적으로 식별된 부상기술은 <표 4>와 같다.

<표 4>에서 MTBE는 부상기술 후보군에 대해

여 얼마나 자주 편집이 이루어 졌는지를 의미한다. 즉, MTBE가 짧을수록 해당용어에 대한 편집이 자주 이루어졌고 전문가의 관심이 높음을 의미

<표 4> 부상기술 식별

부상기술	편집회수	최초등록일	MTBE	편집자수	편집율
omnidirectional camera	18	2008-07-11	5.6	6	3.0
visual odometry	21	2008-07-06	5	5	4.2
Visual Servoing	36	2008-04-14	4.2	5	7.2
3D single object recognition	26	2007-12-12	11.1	5	5.2
Articulated body pose estimation	20	2007-09-15	16.4	9	2.2
ultrasonic ranging module	9	2007-03-06	36.2	6	1.5
image rectification	20	2007-03-05	28.3	13	1.5
visual sensor network	18	2006-12-15	32.3	9	2.0
stereo cameras	7	2006-12-01	96	4	1.8
blob detection	116	2006-09-04	6	22	5.3
Active pixel sensor	102	2006-05-08	8.8	49	2.1
Monte Carlo localization	12	2006-02-16	72.8	8	1.5
Pose Estimation	42	2005-11-12	25.4	17	2.5
information integration	29	2005-09-20	36.2	23	1.3
topological map	25	2005-05-09	43	13	1.9
stereopsis	120	2005-05-04	10.6	65	1.8
ICP	28	2005-05-02	45.1	22	1.3
Reflection mapping	69	2005-04-17	17.9	38	1.8
stereo camera	81	2005-02-15	15.5	49	1.7
Sensor fusion	70	2005-02-03	17.9	37	1.9
Smart camera	79	2005-01-20	17.4	45	1.8

한다. 편집율은 총 편집회수를 총 편집자수로 나눈 값이며 해당용어의 편집활동자들이 어떻게 분포되어 있는지 보여준다. 편집율이 높을수록 특정 소수 편집활동자들의 분포가 높다고 할 수 있다.

4. 결론

과학기술정보는 연구개발 및 기술혁신의 과정에서 필수적인 인프라이다. 국방 분야에서도 이러한 과학기술정보를 관리하기 위한 제도적 방안과 정보관리시스템을 효과적으로 구축하고 있는 중이다.

하지만 축적되어 가는 국방과학기술정보를 계량적으로 분석하여 정보체계 구축의 성과제고를 위한 방법론에 관한 연구노력의 활성화가 요구된다.

본 연구에서는 대용량의 국방과학기술정보를 효과적·효율적으로 분석하기 위한 대안으로 텍스트마이닝을 이용한 국방과학기술정보분석 프로세스를 제안하고 무인전투체계의 자율주행을 위한 핵심기술인 ‘지형인지장치기술’을 대상으로 실증분석 하였다.

1997-2007년 기간에 대하여 INSPEC DB에 수록된 지형인지장치기술 관련 과학기술 저널논문을 수집·정제하여 부상기술의 단서가 될 키워드를 도출한 후 Wikipedia를 활용하여 최종적으로 지형인지장치기술 중 부상기술을 식별하는 프로세스를 제안하였다.

제안하는 방법론은 해당기술의 비전문가라 할 지라도 전문가의 참여를 최소화하면서 국방과학기술정보분석을 수행하여 국방핵심 기반기술 기획시 응용할 수 있는 방법론으로서 의미가 있다. 또한 민간분야의 데이터베이스 뿐만 아니라 국방과학기술활동을 통하여 발생하는 산출물들에 대하여 텍스트마이닝을 이용하여 분석할 수 있는 방법을 개괄적으로 제시함으로써 국방 핵심기술 기

획시 활용할 수 있도록 하였다. 그리고 인터넷 사용자라면 누구나 쉽게 접근할 수 있는 사용자 기반의 온라인 백과사전인 Wikipedia를 활용한 부상기술을 도출함으로써, 향후 이러한 방법에 대한 논의의 가능성을 창출하였다.

향후 이러한 국방과학기술연구정보의 분석시에 사용DB에 따라서 국방부문의 기술지식 맵을 분석하는데도 확장 적용할 수 있을 것이다.

참고문헌

- [1] 방위사업청 법률 제10218호, 방위사업법, 2010. 3.31.
- [2] 이우형·김윤명·박각로·이명호, “Co-word Analysis를 통한 신기술 분야 도식화 방법에 관한 연구”, 경영과학 23권, 2006.
- [3] 윤문섭·이우형·김윤명·오해영·손성혁, 신기술 연구기획 사전 타당성분석을 위한 지식맵 작성 방법론 개발 및 활용방안, 과학기술정책연구원, 2003.
- [4] Olson, D., and Shi, Y., Introduction to Business Data Mining, New York: McGraw-Hill, 2007.
- [5] 임치환, Knowledge Map의 활용에 관한 연구, 한국외국어대학교 석사학위 논문, 2006, p.55.
- [6] 대통령령 제 20793호, 국가연구개발사업의 관리 등에 관한 규정, 2008.5.27.
- [7] 국방기술품질원, 2007 국방과학기술조사서 제 4권(일반본), 국방기술품질원, 2008.
- [8] Yoon, B., and Park, Y., “A systematic approach for identifying technology opportunities: Keyword-based morphology analysis”, Technological Forecasting & Social Change, 2004.
- [9] Shmueli, G., and Patel, N. R., and Bruce, P. C., Data Mining for Business Intelligence, Wiley Interscience, 2007.

【 저자소개 】

이 태 봉 (E-mail: Itb6405@naver.com)

- 2000 해군사관학교 군사전략학과 졸업(학사)
- 2009 국방대학교 국방과학학부 무기체계전공 졸업(석사)
- 2009 방위사업청 사업관리본부
- 2010 현재 해군 진해 00전단 근무
- 관심분야 기술정책, 데이터마이닝, OR

이 춘 주 (E-mail: sarang90@kndu.ac.kr)

- 1991 육군사관학교 물리학과 졸업(학사)
- 1997 미국 UC Berkeley 핵공학과 졸업(석사)
- 2005 KDI국제정책대학원 국제정치경제학과 졸업(정책학 석사)
- 2006 서울대학교 기술정책대학원 졸업(기술정책 박사)
- 현재 국방대학교 국방과학학부 조교수
- 관심분야 무인무기체계, 기술혁신, 연구개발 및 기술관리, DEA-STATA