

# 빅데이터 정보분석을 통한 연구개발 이슈 및 부상 연구분야 도출

김소영\*·이재민\*\*·여운동\*\*\*·고병열\*\*\*\*

## I. 서론 (1p)

기업이나 공공기관의 연구개발 기획 시 가장 먼저 시작하는 것이 연구개발의 영역을 선정하고 연구개발 목표를 설정하는 일이다. 조직마다 주 활동 분야가 어느 정도 정해져 있으면, 우선 해당 기술과 주변 기술의 최근 동향을 파악하고, 시장 환경에 대한 정보를 수집하고 분석하는 작업에 착수하게 된다. 사회적인 측면, 기술적인 측면, 경제적, 정치적 측면 등 여러 측면을 고려하여 기술 및 시장 환경을 조사하고 최근 이슈를 분석하는 작업은 많은 정보를 수집하여 종합적으로 분석하는 통찰력이 필요한 작업이다. 연구개발 분야별로 아이디어의 탐색과 연구개발 주제의 선정 및 목표 설정 등은 이런 통찰력을 보유하고 있다고 평가되는 해당 분야의 전문가들의 손을 거쳐 이루어지는 것이 일반적이다.

분야의 전문가들은 해당 분야에 대해 오랜 기간 동안 축적된 전문성과 통찰력이 있지만, 한편으로는 저마다 다른 기준이 있어 탐색과 선정에 주관이 개입할 수 있다는 단점이 있다. 또한 경험에 의한 판단을 하게 되기 때문에 연구개발 분야에 대한 선입관에서 자유롭기 어렵기 때문에 새로운 방향성에 대해 보수적일 수 있으며, 전문가 자신의 속한 분야와 관련이 적은 분야와의 협력이나 융합에 대해 충분한 고려가 어려울 수 있다.

구조화된 지식의 결과물을 정량적으로 분석하는 방법은 전문가의 통찰과 주관을 뒷받침해줄 수 있는 근거가 되거나 놓칠 수 있는 부분을 보완할 수 있는 수단이 될 수 있다. 이런 흐름은 ‘과학계량학’, ‘정보계량학’ 등의 분야 내에서 지속적으로 발전해오면서 그 방법론이 점차 견고해져 왔다. 여기서 구조화된 지식의 결과물이란 연구개발 분야에서 발생하는 지식과 노하우를 명문화된 형태로 출판하는 과정을 거치는 논문, 특허 등을 의미한다. 논문과 특허는 축적된 양이 방대하고 해마다 더욱 많은 양의 문서가 생산되고 있으며, 전문가의 평가를 거쳐 출판되므로 비교적 양질의 정보를 담고 있다고 할 수 있다. 또한 제목, 초록, 키워드 등 약속된 형태로 생산되며, 인용의 형태로 지식의 흐름에 대한 정보도 담고 있어 이를 분석하는 것은 지식, 특히 기술의 진화과정(이재민 외, 2011)에 대한 정보를 제공하는 좋은 원천이 된다.

최근에는 구조화된 논문, 특허뿐만 아니라 구조화되지 않은 대규모의 정보를 활용하여 사회 환경을 분석하는 데 활용하는 빅데이터 분석도 활발해지고 있다. 빅데이터는 크기(Volume), 속도(Velocity), 다양성(Variety)의 특징을 갖는다고 정의되며, 인터넷 상에서 생산되는 대량의 문서들 뿐만 아니라, 지리, 교통, 기상 등 여러 종류의 과학기술 데이터 및 각종 기기와 센서들이 생성하는 수많은 신호 정보 등 구조화되지 않았지만 대량으로 생산되는 데이터들을 모두 포함하는 포괄적인 개념이다. (Russom, 2011) 특히, 인터넷에서 생산되는 문서들을 활용하는 빅데이터 분석은 사회적 이슈를 감지하고, 변화의 방향성과 잠재적 위험 등을 분석하여 미래 예측을 위한 근거자료를 제시하는 방향으로 진화하고 있다. (Chen et al, 2012)

\* 김소영, 한국과학기술정보연구원 연구원, 02-3299-6128, sykim8171@kisti.re.kr

\*\* 이재민, 한국과학기술정보연구원 선임연구원, 02-3299-6292, jmlee@kisti.re.kr

\*\*\* 여운동, 한국과학기술정보연구원 선임연구원, 02-3299-6017, wdyeo@kisti.re.kr

\*\*\*\* 고병열, 한국과학기술정보연구원 책임연구원, 02-3299-6039, cohby@kisti.re.kr

인터넷에서 실시간으로 생산되는 빅데이터, 예를 들어 인터넷 블로그나 신문 기사, 소셜 네트워크 상의 기록들은 현재 사회에서 활발히 논의되고 있는 사회적 이슈를 감지하여 분석하는 데 좋은 재료가 된다. 기업들은 제품의 개발 방향이나 마케팅 전략을 짤 때 이들 자료를 수집, 분석하여 소비자의 수요를 정확히 파악하는데 활용하고 있다(윤미영, 2013). 과학기술과 연구개발 분야에도 인터넷 빅데이터를 활용할 수 있는데, 특정 기술이나 기술이 체화된 제품에 대한 인터넷 상의 논의는 사용자의 수요를 의미할 수도 있고, 혁신의 필요성이나 잠재적 위험성 등을 의미할 수 있어 이를 분석하는 것은 미래 연구개발의 방향성이나 전략을 수립하는 데 참고 자료가 될 수 있다.

구조화된 자료로서 특허나 논문을 분석하는 작업도 어떤 측면에서는 빅데이터 분석의 방법론이 적용될 수 있다. 특허나 논문 자체는 구조화된 형태로 구성되어 있으나 분석할 데이터의 크기가 점차 커지고, 텍스트를 기반으로 과학기술의 지식맵을 구성하는 작업은 방대한 데이터의 추출과 정제가 필요하다. 또한 텍스트마이닝을 통해 추출된 복잡한 형태의 데이터 간 연관 관계를 재구조화하는 작업은 기존의 통계적 방법론을 통한 분석 결과 이상의 새로운 정보를 제공한다는 측면에서 빅데이터 분석으로 분류될 수 있다.

본 연구는 논문, 특허, 웹 등 서로 다른 출처의 정보를 분석하여 기술의 성숙 단계마다 부상하고 있는 이슈를 찾는 방법을 제안하고 적용하였다. 이를 위해 문서 내 기술키워드의 동시발생 패턴과 논문 인용패턴의 유사성을 이용하여 기술 군집을 찾고, 기술군집의 특성을 분석하였다. 군집의 특성 분석에는 소셜네트워크 분석 지표와 부상성 지표를 사용하였다. 각 분석을 위해서는 국내외 관련 상용 소프트웨어 및 자체 개발한 소프트웨어를 사용하였다. 본 연구가 제안하는 방법론을 통해 특정 기술 분야를 연구개발 단계, 실용화 단계로 구분하여 각 단계별로 부상하고 있는 연구 영역과 관련 이슈를 도출할 수 있으며, 웹 정보 분석을 통해서도 사회적으로 주목받고 있는 관련 이슈와 기술 영역을 도출할 수 있다. 개발된 방법론을 항공전자 시스템 설계 및 시스템 소프트웨어 기술 분야에 적용하여 부상하고 있는 연구개발 이슈를 선정하였다.

## II. 연구방법론

### 1. 기존 연구

논문과 특허의 구조적 정보, 즉 서지정보를 활용하여 다양한 계량적 지표를 통해 과학기술 연구개발의 현상을 분석하고 평가하는 학문이 과학계량학(Scientometrics) 분야이다. 과학계량학 분야는 20년대 중반 이후 지난 수십 년 간 꾸준히 연구되어왔으며, 최근에는 방대한 과학기술관련 문서들의 생성과 컴퓨팅 능력의 향상, 분석 기법의 고도화 등과 함께 더욱 비약적으로 성장하고 있다. 특히, 텍스트 마이닝이나 소셜 네트워크 분석 방법론 등 빅데이터 분석 방법론의 발전과 함께 다양한 시도가 이루어지고 있다.

과학계량학은 전통적으로 과학기술 연구개발활동을 분석하고 평가하는 데 활용되어 왔는데, 논문과 특허의 서지정보의 통계적 처리를 통해 산출하는 지표들, SCI(Science Citation Index), IF(Impact Factor) 등이 대표적인 사례이다(Garfield and Merton, 1979). 그 외에도 서지정보로부터 나타나는 정보들 간의 관계를 통해 산출하는 지표들을 활용하기도 하는데, 공동 저자 수(Co-authorship), 공동 인용 문헌(Co-citation) 분석 등이 그 사례이다.

서지정보의 처리를 넘어서서 문서 자체에 산재한 정보를 재구조화하여 정보 간의 관계를 통해 새로운 통찰력을 얻고자 하는 시도도 많이 이루어지고 있다. 한 문서 상에 출현하는 단어들 간의 동시 출현을 기반으로 과학기술 개념, 활동, 영역 간의 관계를 분석하는 동시단어(Co-word) 혹은 동시 출현(Co-occurrence) 분석이 그것으로, 과학기술 지식의 지도(map)를 구축하고자 하는 연구에 적용되고 있다(Lee and Jeong, 2008).

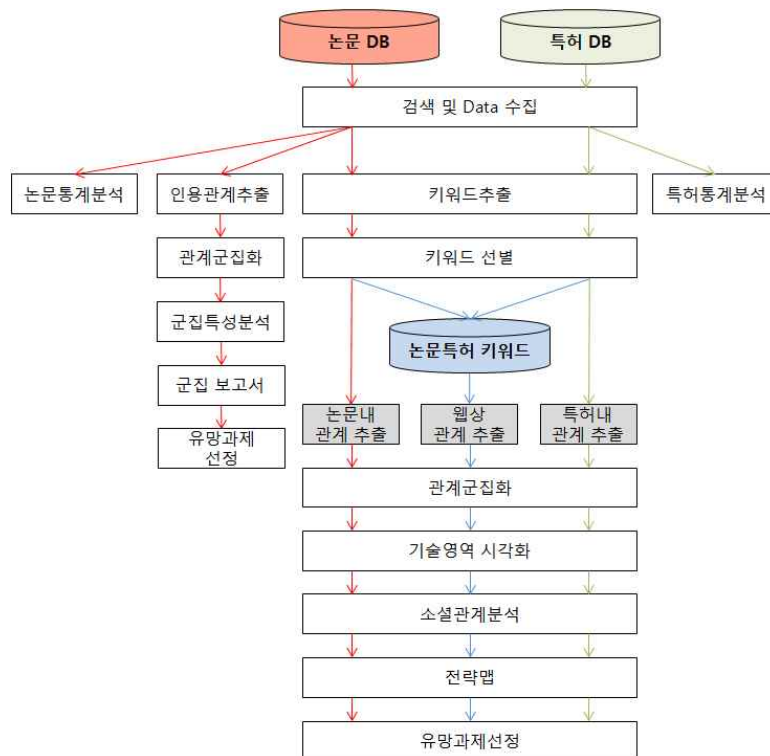
빅데이터 기반의 과학계량학을 활용하여 미래기술예측이나 연구개발 기획에 활용하는 시도도 세계 각국에서 이루어지고 있다. 한국과학기술정보연구원은 2005년부터 특허 및 논문 데이터베이스 분석을 통해 매년 국가차원의 미래유망기술을 선정하여 발표하고 있다(한국과학기술정보연구원, 2012). 주로 사용되는 데이터기반의 분석기법은 인용 분석과 동시단어분석이며, 데이터 기반으로 선정된 유망기술 후보군은 전문가 검증을 거쳐서 최종 선정한다. 일본의 미래기술 예측연구에서는 델파이 조사 방법과 함께 동시인용분석을 급부상 연구 영역을 도출하기 위한 방법론으로 활용하였다(NISTEP, 2005)

Shen et al.(2010)은 유망기술을 선정하는 과정에 종합적인 접근이 필요하며, 퍼지 델파이 (fuzzy Delphi) 방법론, AHP(analytic hierarchy process) 등의 전문가 통찰 기반 분석과 함께 동시 인용 분석을 활용하는 것이 효과적임을 보였다.

미국 조지아텍의 유망기술발굴 방법론 연구(Cozzens et al. 2010)는 정량적 방법론과 정성적 방법론을 동시에 활용하였다. 정성적 방법론으로 델파이, 정량적 방법론으로는 동시저자, 동시단어, 동시인용 분석을 함께 활용하여 군집 분석 및 요인 분석을 수행하여 기술의 유망성을 판단하고자 하였다.

## 2. 연구 프로세스

본 연구는 논문, 특허, 웹 등 서로 다른 출처의 정보를 분석하여 기술의 성숙 단계마다 부상하고 있는 이슈를 찾는 방법을 제안하고자 하였다. 이를 위해 문서 내 기술키워드의 동시발생 패턴과 논문 인용패턴의 유사성을 이용하여 기술 군집을 찾고, 기술군집의 특성을 분석하였다. 군집의 특성 분석에는 소셜네트워크 분석 지표를 사용하였다. 전체 연구 프로세스를 도식화하면 그림 1과 같다.



(그림 1) 전체 연구 프로세스

## 1) 논문 인용패턴 분석을 통한 부상연구영역 도출

학술 논문은 인용정보를 밝힘으로써 지식의 흐름을 알게 해준다. 이로 인해 나타나는 논문의 인용문헌 일치하는 두 개의 다른 논문(지식)이 같은 지식의 원천으로부터 진화한 것임을 말해준다. (Wouters, 1999) 따라서 비슷한 인용의 패턴을 가지고 있는 논문들을 군집화하여 관심 기술의 하위 분류, 즉 연구 영역을 분류해낼 수 있으며, 이를 시각적으로 표현할 수 있다. 최근에는 일본 과학기술정책연구소(NISTEP)와 KISTI가 이 방법론을 활용하여 부상 연구영역을 발굴하고 있다.

논문의 인용패턴을 분석하는 전통적 방법인 ‘서지적 결합분석(Bibliographic coupling)’은 두 개의 논문이 서로 공통된 논문들을 인용하는 경향이 크다면 해당 논문이 유사한 연구 내용을 다루고 있을 것이라고 가정한다. 따라서 이 방법을 활용하면 각 논문들이 인용한 논문 목록의 유사성을 비교하여 논문들을 군집화할 수 있다. (Garfield, 1964)

논문 인용패턴 분석에서 유사성은 코사인 계수 측정을 통해 도출하였다. 코사인 유사성은 다음 식과 같이 계산된다.

$$Sim(d_i, d_j) = \frac{\sum_{k=1}^n r_{ik} \cdot r_{jk}}{\sqrt{\sum_{k=1}^n r_{ik}^2 \sum_{k=1}^n r_{jk}^2}}$$

여기서  $d_i$ 와  $d_j$ 는 서로 다른 문서를 나타내며,  $r_i$ ,  $r_j$ 는 각 논문이 인용한 문서를 나타내는 n차원 벡터로, 분석대상 논문들이 인용한 논문 집단 전체에 대해 인용 여부를 0 또는 1로 표시한다.

인용패턴 유사성 도출의 결과를 활용하여 각 논문을 군집화한 후, 특정 군집 내 핵심 논문과 주요 키워드를 통해 군집의 기술적 특성을 지정한다. 군집화는 Single-linkage 방식을 활용하였는데, 이 방법은 비유사성(거리)의 행렬을 활용하여 두 항목 간 최단거리를 기준으로 군집을 구성하는 방법이다. 군집의 최대 크기는 논문 50건으로 제한하였다.

## 2) 키워드 네트워크 분석을 통한 부상이슈 도출

인용패턴분석으로 논문의 출판을 통한 지식의 흐름과 진화 과정을 알 수 있다면, 주제어(키워드)의 동시 출현 관계를 통해서도 주어진 기간 동안 누적된 지식의 구조, 즉 세부 지식 간의 관계를 알아볼 수 있다. 이 분석은 두 개 이상의 키워드가 하나의 문서(논문, 특허를 비롯한 완결된 문서)에서 동시에 활용되고 있다면, 이 키워드가 나타내는 개념이 서로 연관되어 있다는 가정을 바탕으로 하고 있다. (Noyons, 1999) 또한, 동일한 키워드의 군집들을 함께 활용하고 있는 두 개의 논문은 결과적으로 동일한 개념들을 함께 사용하는 유사한 연구 개발을 수행하고 있다는 가정 하에 분석하는 것이다.

분석 대상 문서에 나타나는 키워드의 동시 출현 관계를 분석하기 위해서는 우선 문서의 텍스트로부터 키워드 (명사 혹은 명사구)를 추출하고, 이들 키워드가 각 문서에 공통적으로 출현한 빈도를 계산한다. 각 단어 간의 공통 출현 빈도를 적절한 유사도 지수 계산 방법에 적용하면 각 키워드 간의 유사도를 측정할 수 있으며, 이를 통해 각 키워드 간의 관계를 네트워크 형태로 시각화할 수 있게 된다. 결과적으로 키워드 간의 시각화된 네트워크 형태와 키워드의 군집 및 군집 간 연결 패턴을 분석하여 연구 영역 간의 관계를 분석할 수 있다. 키워드 네트워크의 시각화는 VOSviewer(Nees and Waltman, 2009)를 활용하였다.

키워드 네트워크 분석은 직접적인 연구와 기술개발에 대한 내용을 담고 있어 연구개발 단계나

기술 실용화 단계 분석에 적합한 논문, 특히뿐만 아니라 거시적인 사회 트렌드나 사용자의 관심사와 수요 등을 관찰하기에 반드시 적합한 웹문서 분석을 포함한다. 웹문서 데이터는 신문기사, 개인의 블로그나 커뮤니티의 사적인 기록, 산업과 시장분석 자료 등 다양한 사회적 시점을 통합적으로 반영한다.

### 3) 지표 분석과 부상이슈, 연구 영역 도출

논문, 특허 및 웹으로부터 키워드 네트워크가 제작되고, 키워드 간 유사성 분석을 통해 키워드 군집이 도출되고 나면, 각 군집별로 네트워크 속성값을 계산한다. 특히 사회관계네트워크의 분석 지표인 밀도와 중심성을 계산하는데, 밀도는 해당 연구영역 내 논문들의 연결강도를 의미하여, 연구의 발전정도를 나타내며 중심성은 해당연구영역이 전체연구영역에서 중심적인 위치를 차지하는 정도를 나타낸다. 추가적으로 해당연구영역의 최신성을 나타내는 평균 출판연도를 고려할 수 있다.

밀도와 중심성의 크기에 따라 각 군집은 4분면 상에 위치하게 되는데 (An et al., 2011), 높은 중심성과 낮은 밀도를 가진 제4사분면에 위치한 기술 군집은 관련 연구 주제 가운데 중심적인 위치를 차지하지만 아직 많은 연구가 되지 않은 분야로 향후 부상이 기대되는 연구 영역으로 해석할 수 있다.

## III. 사례 분석

### 1. 분석 대상 설정

본 연구에서는 항공전자분야의 핵심 세부 기술군에 대해 제시한 연구방법론을 적용하여 연구개발 이슈와 부상하고 있는 연구 영역을 분석하여 제시하고자 한다.

#### 1) 분석 대상 기술의 특징

항공전자(avionics)란 용어는 항공(aviation)과 전자(electronics)의 합성어로 1930년대부터 사용되기 시작하였으며, 항공기에 장착되는 구성품이나 세부계통 중에 전자기술에 의해 동작하는 것들을 통칭한다. (송찬호, 2009) 항공기의 첨단화와 함께 장비들의 정밀성과 신뢰성 등이 요구되면서 점차 IT의 비중이 높아지고 있으며, 최근에는 항공기 생산 비용의 30%에서 70%까지가 항공전자 장비에 투입되고 있다.

국내 항공산업은 1980년대부터 글로벌 항공기 제작사(Boeing, Airbus 등)의 기체 면허 생산을 시작하여 현재에 이르고 있으나, 최근 기체제작 분야는 인건비가 낮은 중국이나 동남아 등으로 이전됨에 따라 국내 가격경쟁력이 약화되고 있다. 따라서 후발주자들의 진입이 어려워지면서 한국이 기술경쟁력을 가지고 있는 IT분야가 핵심인 항공전자 분야가 국내 항공산업의 미래 핵심 연구개발 분야가 될 수 있다.

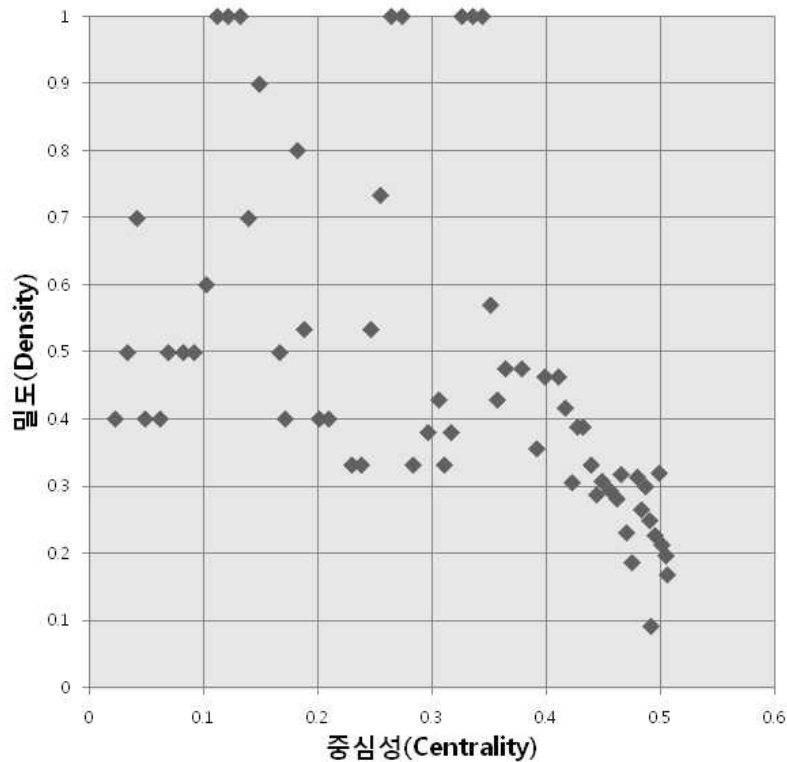
항공전자분야는 매우 광범위한 영역을 아우르는 기술로, 매우 많은 세부 기술과 부품, 장비들로 이루어진다. 또한 다른 항공관련 기술과 마찬가지로 항공전자 관련 기술은 첨단, 핵심기술로 분류하여 선진국이 기술이전을 기피하는 분야로, 동 분야 진입을 위해서는 자체적인 연구개발이 반드시 필요한 분야이다. 따라서 국내기업에서 이 분야에 진출하고자 할 때, 정성적인 판단과 조사만으로는 연구개발목표나 전략을 수립하기 어려우며, 논문, 특허 등을 통해 최근 부상하는 연구 영

역과 기술 개발의 이슈를 정량적으로 수집, 분석하는 것이 필수적인 분야이다. 또한 향후 기술의 사업화 등을 고려할 때 사회적 트렌드와 산업 및 시장의 이슈를 살펴볼 필요도 있다.

## 2. 분석 결과

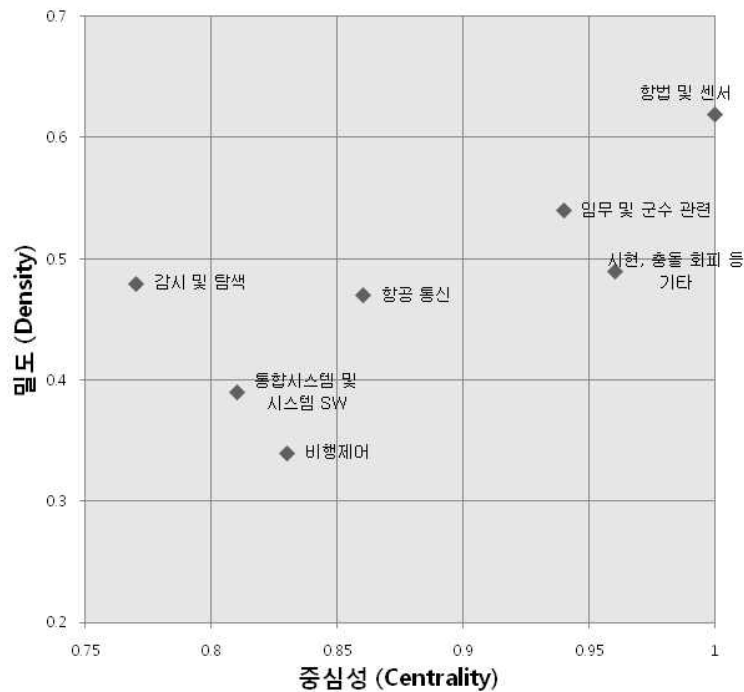
### 1) 논문 인용패턴 유사성 분석

항공전자분야는 크게 분류할 때 항공전자 시스템 설계 및 시스템소프트웨어, 항법 및 비행제어, 항공통신과 데이터처리, 임무 및 무장체계, 생존과 감시/탐색/안진 등으로 분류할 수 있다. 각 부분을 종합하여 SCOPUS 데이터베이스로부터 2003년부터 2012년까지 총 3,672건의 논문이 검색되었으며, 인용패턴분석을 통해 총 63개의 연구영역 군집으로 구분하였다. 각 연구영역은 5개에서 50개의 논문으로 구성되어 있다. 군집별 지표 분석 결과를 보면, 최근 여러 시도가 이루어지고 있는 항공전자분야의 특성이 반영되어 모든 군집이 밀도와 중심성이 그리 높은 편은 아니다. 그러나 그림 2에서 보는 바와 같이, 일부 군집이 아직 밀도는 높지 않으나 중심성은 비교적 높은 편으로, 앞으로 활발한 연구 활동이 이루어질 것으로 예상된다. 표 1은 부상연구영역 8개를 제시하고 있는데, 이들은 높은 중심성과 낮은 밀도를 보여 향후 부상이 기대되고, 핵심 논문 수의 규모가 일정 수준이상인 동시에, 논문 출판 평균연도가 2007년 이후 연구영역들이다.



(그림 2) 인용패턴분석을 통한 연구영역 전략맵





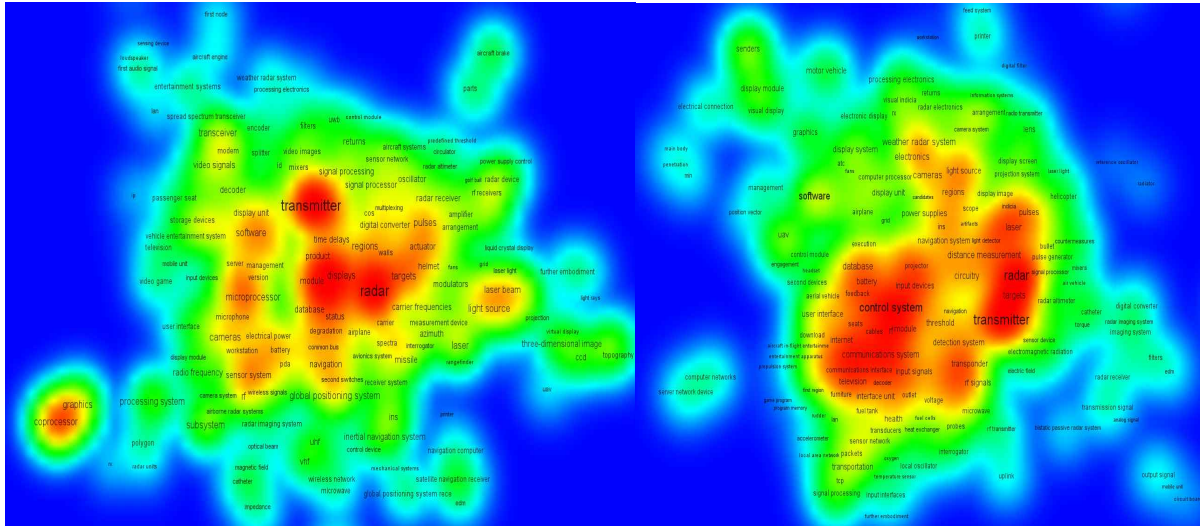
(그림 4) 논문 키워드 군집의 전략맵

그러나 밀도 측면에서는 약간의 차이가 있었는데, 그림 4의 전략맵에서 보는 바와 같이, Integrated modular avionics(IMA)를 포함한 통합시스템과 시스템 SW 관련 클러스터, 무인항공기(UAV), flight control 등을 포함하는 비행제어, communication systems를 포함하는 항공 통신분야가 단일 세부분야에 대한 군집이면서 비교적 높은 중심성과 낮은 밀도를 보이는 기술 분야로 나타났다. 반면 GPS/INS를 포함하는 항법 및 센서 분야, military operations를 포함하는 임무(군수) 관련 분야 및 RADAR를 포함하는 감시 및 탐색 분야는 높은 중심성과 높은 밀도를 보여 이미 잘 구축되어 있는 연구 영역으로 나타났다. 그 외에 시현, 충돌 회피 등을 포함하는 여러분야가 혼재되어 있는 키워드 군집이 중심성은 높으나 낮은 밀도를 보이는 분야로 나타났다.

한편, 항공전자분야의 각 세부기술분야를 종합하여 미국특허청(USPTO) 데이터베이스로부터 특허를 검색한 결과, 2003년부터 2012년까지 총 2,259건의 특허가 수집되었다. 이들 전체 특허문서의 제목, 초록, 청구항으로부터 자연어 처리를 통해 총 8,195개의 명사와 명사구를 추출하였다. 이중 동의어 처리작업과 불용어 및 빈도가 낮은 키워드를 제거하는 과정을 통해 총 541개의 분석대상 핵심 키워드를 선별하였다. 또한, 앞서 밝힌 바와 같이 동시 출현 빈도 및 유사도 지수 계산을 통해 유사 핵심키워드별로 군집화하는 과정을 거쳤다. 특히, 특허는 기술의 실용화 단계에서 주요 세부 기술개발 영역과 동 영역에 대한 부상성을 살펴볼 수 있기 때문에 2003년부터 2007년까지의 전반기와 2008년부터 2012년까지의 후반기로 나누어 각 기술군의 동향 변화를 살펴보았다.

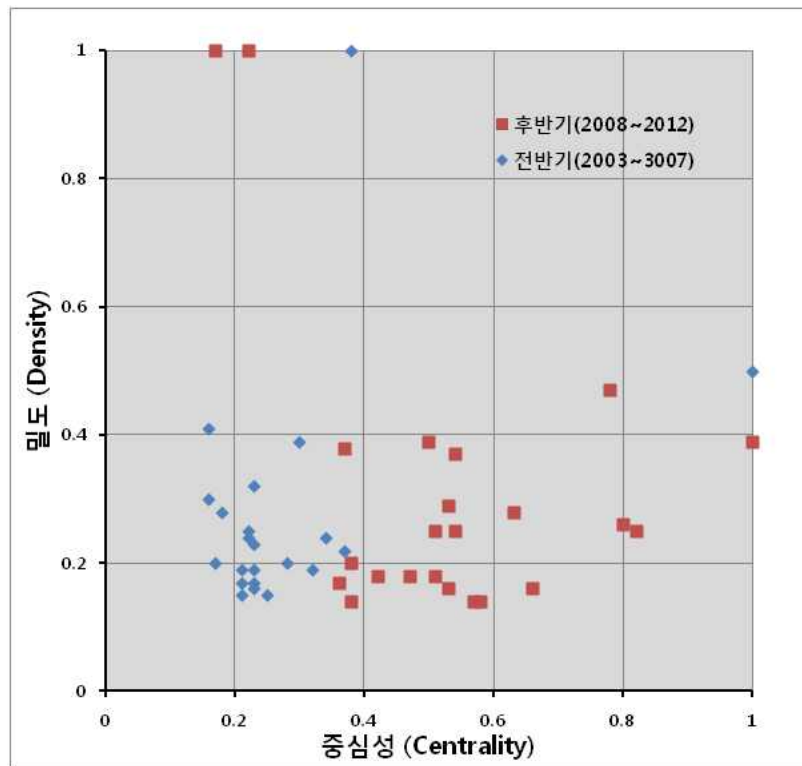
그림 5는 전반기와 후반기에 항공전자분야 특허의 핵심 키워드의 관계 네트워크 형태가 변하는 모습을 보여준다. 후반기가 되면서 두드러지게 나타난 부분은 'control system'을 중심으로 하는 비행제어 분야의 부상과 항공통신분야의 영역 중심으로의 이동이다. 또한 항공전자 기술 영역 전반에 걸쳐 세부 영역별로 고른 발전을 보이면서 각 세부 기술영역 간의 연결 정도가 더욱 밀접해지는 경향을 알 수 있다.





(그림 5) 특허 키워드 네트워크 형태의 변화 (전반기 vs. 후반기)

이런 경향은 그림6의 전략맵을 통해 더욱 확실하게 알 수 있다. 2007년까지 항공분야 특허를 통해 본 세부 기술 영역은 중심성과 밀도가 모두 낮아 각 세부 분야별로 산재된 개발이 이루어지면서도 분야 내의 긴밀한 연결과 집중적인 기술 개발이 이루어지지 않는 모습을 보인다. 그러나 2008년 이후 세부 분야 내의 긴밀한 연결성은 높아지지 않더라도 서로 간의 융합적 성향이 커지는 것을 알 수 있다. 이런 부분은 항공전자 기술 전반이 통합 모듈러 시스템으로 진화해 나가는 경향과 일치한다고 볼 수 있다.

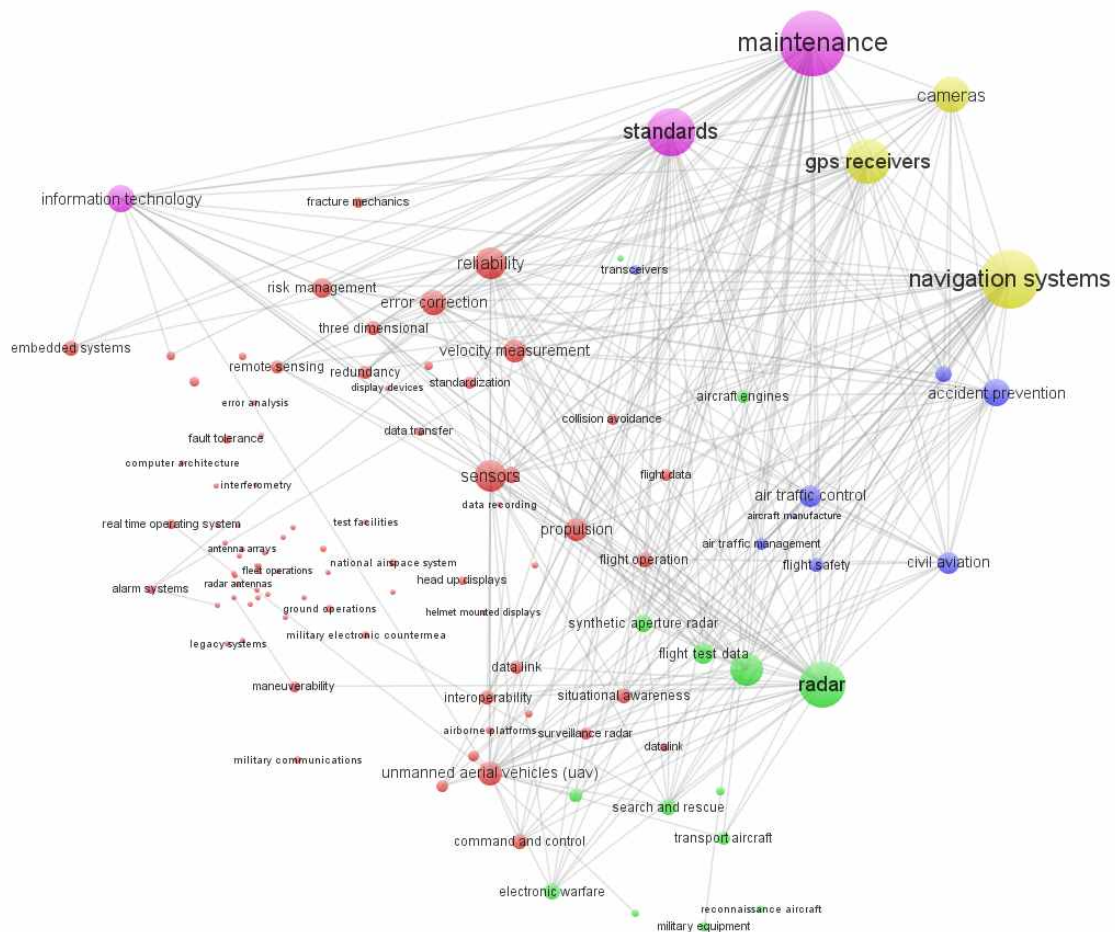


(그림 6) 특허 키워드 전략맵의 변화

마지막으로 논문과 특허에서 수집된 키워드를 다시 한 번 정제하여 인터넷을 통해 연구 영역의 군집 특성과 부상 정도를 다시 한 번 살펴보았다. 항공전자분야 논문과 특허에서 선별된 키워드 중요도에 따라 107개의 키워드를 재선정하여 2009년부터 2013년까지의 모든 웹문서를 대상으로 키워드의 동시 출현 분석을 실시하였다.

항공전자분야의 핵심 키워드에 대한 사회적 관심 경향은 총 5개의 군집으로 분류되었다. 그림 x는 키워드 간의 동시 출현을 기준으로 연관성이 높은 키워드들을 2차원 평면에 분포시킨 후 군집화한 결과이다. 노드의 색으로 각 군집을 구별할 수 있다.

인터넷을 통해 항공전자분야의 핵심 키워드 군집 경향을 살펴본 결과, 기술적인 용어 보다는 일반적인 용어들의 출현이 두드러지는 것을 확인할 수 있다. 특히 그림 7 우상단의 maintenance, navigation systems를 비롯한 소수의 일반적인 용어들이 매우 높은 출현 수를 보이고 있다. 반면, 좌하단 부분에는 출현 수가 많지 않은 기술적 용어들이 모여 있으며, 이들은 출현 수는 많지 않으나 다른 핵심 키워드 간을 연결하는 역할을 하고 있는 것을 알 수 있다. 사회적으로는 세부 기술 자체 보다는 상위 개념의 키워드에 대해 많은 이슈가 제기되고 있는 것을 보여주는 사례라고 할 수 있다.



(그림 7) 웹 키워드 분포

### 3) 부상이슈 및 유망연구영역 도출

제시된 방법론을 항공전자 핵심기술분야에 적용한 결과, 데이터 원천의 특성에 따라 빅데이터 분

석의 결과는 다른 결과를 제시해주는 것으로 나타났다. 논문데이터를 활용한 분석은 연구개발 단계에서는 항공전자분야가 세부기술분야 전반에 걸쳐 고루 발전하고 있다는 결과를 보였다. 그러나 좀더 세부적으로 보면, 차세대 IMA나 소형SAR과 자율비행 등 아직 실용화 단계에 이르지 않은 최신 기술 분야가 향후 부상할 가능성이 있는 연구 영역으로 나타났다. 특히 분석 결과에서는 항공통신분야와 비행제어 분야가 기술 실용화 단계에서 부상 가능성이 있는 연구 영역임이 나타났다, 또한 분석기간 후반기에 세부기술 간의 연결이 긴밀해진 것은 항공전자분야에서 통합시스템(IMA)가 점차 실용화 단계에 이르고 있음을 말해준다. 반면, 웹 분석의 결과는 구체적인 기술 발전보다는 최근 산업적 이슈인 항공기 정비/수리/점검(MRO; maintenance, repair, and overhaul)이나 민간 항공기의 트래픽 제어 등이 좀더 관심을 받고 있음을 보여준다.

#### IV. 결론

본 연구에서는 논문, 특허 및 웹문서에 대한 통합적 접근과 분석을 통해 기술개발의 각 단계별로 연구 영역의 특성과 부상성을 분석할 수 있는 방법론을 제시하였다. 제시한 방법론은 항공전자분야의 세부 핵심기술군에 적용하여 사례 분석을 수행하였다. 논문 분석을 통해서 연구개발 단계에서 부상이슈와 유망 영역을 도출할 수 있었으며, 특허 분석을 통해서 기술 실용화 단계에서 부상이슈를 전, 후반기로 분리하여 특성과 변화를 분석할 수 있었다. 마지막으로 웹 분석을 통해서 해당 분야에서 사회적으로 주목받는 이슈를 도출하였다.

과학계량학에 기반을 둔 논문, 특허, 웹 빅데이터 분석은 기업이나 연구조직에서 목표로 하는 기술의 단계에 따라 부상하는 이슈를 관찰하고 적합한 연구 영역을 찾는 데 도움이 될 수 있다. 연구개발 단계에서는 논문데이터를 원천으로 분석하여 부상 연구 영역을 관찰하고, 기술 실용화 단계를 목표로 한다면 특허데이터를 기반으로 분석하는 것이 적합할 것이다. 만약 시장 진출과 산업의 부상 이슈를 고려하고자 한다면 웹데이터를 함께 고려하는 것이 더욱 적절할 수 있을 것이다.

#### [참고문헌]

- 송찬호, (2009), “항공전자 기술 발전 동향”, *IT SoC Magazine*, 2009년 9월호
- 양혜영, (2012), “빅데이터를 활용한 기술기획 방법론”, Issue Paper 2012-14, 한국과학기술기획평가원.
- 윤미영, (2013) “더 나은 미래를 위한 데이터 분석 - Big Data 글로벌 선진사례 II”, 한국정보화진흥원
- 한국과학기술정보연구원, (2013), *KISTI 미래기술백서*.
- An, X.Y. and Wu, Q.Q., Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics*, 2011. 88(1): p. 133-144.
- Bani-Ahmad, S., A. Cakmak, G. Ozsoyoglu, (2005) “Evaluating Publication Similarity Measures,” *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 28(4): 21-28.

- Chen, H., R. H. L. Chiang, and V. C. Storey (2012), "Business Intelligence and Analytics: From Big Data to Big Impact", *MIS Quarterly*, 36(4): 1165-1188.
- Cozzensa, S., Gatchaira, S., Kang, J., Kim, K., Lee, H., Ordóñez, G., Porter, A., (2010), "Emerging technologies: quantitative identification and measurement", *Technology Analysis & Strategic Management*, 22(3): 361-376.
- Garfield, E. (1964), "'science citation index' – A New Dimension in Indexing", *Science*, 144: 649 - 654.
- Garfield, E. and R. K. Merton (1979). *Citation indexing: Its theory and application in science, technology, and humanities*, Wiley New York.
- Wouters, P.F. (1999), *The Citation Culture*, PhD thesis, University of Amsterdam.
- Lee, B. and Y.-I. Jeong (2008). "Mapping Korea's national R&D domain of robot technology by using the co-word analysis", *Scientometrics* 77(1): 3-19.
- Nees, J., Waltman, L. (2009) "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics* 5: 14 - 17.
- NISTEP, (2005), "Study on Rapidly-Developing Research Area", *NISTEP REPORT* No.95
- Russom, P. (2011), "Big Data Analytics", *TDWI Best Practices Report*, Fourth Quarter.
- Shen, Y., Chang, S., Lin, G., Yu, H., (2010), "A hybrid selection model for emerging technology", *Technological Forecasting & Social Change*, 77(1): 151-166.