

4차 산업혁명 관련 기술의 정부R&D과제 현황과 관심도 분석 -그래핀, 빅데이터, 바이오마커를 중심으로

정재웅*, 박현우**, 성태응***

논문 요약

바야흐로 4차 산업혁명의 시대가 도래하면서, 4차 산업혁명과 관련된 기술에 대한 관심 또한 날이 높아져 가고 있다. 이런 시대적 상황에 맞춰, 본 연구에서는 웹 검색트래픽을 활용하여 4차 산업혁명 관련 기술의 국내·외 트렌드를 비교하고 관련된 기술에 대한 국내 정부 R&D 과제 현황에 대해 분석해 보고자 한다.

본 분석은 먼저 관련 기술에 대한 정부 R&D 현황 과제 현황을 네트워크 분석하고 분석된 기술을 4차 산업혁명의 3가지 카테고리로 구분한다. 이후 카테고리별로 선별된 각각의 기술에 대해 구글의 웹 검색트래픽을 활용하여 관련된 기술의 국내·외 트렌드를 비교·분석한다.

본 연구를 통해 4차 산업혁명 관련 기술에 대한 과거와 현재의 관심 동향을 확인할 수 있을 것으로 예상된다. 또한, 해당 기술과 관련된 정부 R&D 과제 현황을 네트워크 분석하고 시각화하여 4차 산업혁명 관련 R&D 현황에 대한 직관적인 정보를 제공하고자 한다.

본 연구는 인류 최대의 혁명으로 정의되는 4차 산업혁명 기술에 대한 국내·외 트렌드 정보와 함께 현재까지 진행되어온 정부의 R&D 현황을 직관적으로 볼 수 있게 제공함으로써, 4차 산업혁명과 함께 다가오는 미래에 대한 올바른 정책 방향 설정에 기여할 것으로 기대된다.

Keyword : 4차 산업혁명, 웹 검색트래픽, 네트워크 분석, 정부 R&D 과제 현황

* 과학기술연합대학원대학교 과학기술경영정책 박사과정, jj@kisti.re.kr, 02-3299-6269

** 한국과학기술정보연구원 책임연구원, haprk@kisti.re.kr, 02-3200-6051

*** 한국과학기술정보연구원 책임연구원, ts322@kisti.re.kr, 02-3299-6172

I. 서론

작년 이세돌과 알파고(AD)의 바둑대결은 사람들에게 큰 충격을 주었고 인공지능(AD)은 더욱 세상에 두각 되었다. 이를 계기로 인공지능을 포함한 4차 산업혁명에 대한 관심도 급증하게 되었다. 4차 산업혁명(The 4th industrial revolution)은 2016년 1월 20일 스위스 다보스세계경제포럼에서 클라우스 슈밥(Klaus Schwab)회장이 처음으로 언급한 개념으로 융합을 핵심으로 하여 물리적 시스템, 디지털 시스템, 바이오 시스템이 융합한 인류 최대의 혁명으로 정의 되었다. 먼저 4차 산업 혁명의 시대에 메가트렌드(Megatrend) 기술로 물리학, 디지털, 생물학으로 나누어 보자면 첫 번째 물리학(Physical) 기술로는 무인운송수단, 3D프린팅, 로봇공학, 신소재의 네 가지 영역에서 발전을 기대하고 있다. 첫 번째 기술은 두 번째로 디지털(Digital)기술과 함께 발전하게 되는데 사물인터넷과 같이 실물과 가상네트워크를 연계하는 것으로 새로운 관리 방식의 영향을 얻게 된다. 예를 들어, 우버는 소유하고 있는 자동차가 없고, 페이스북은 콘텐츠를 생산하지 않는다, 또한 알리바바는 물품을 만들지 않고 에어비앤비는 소유한 부동산이 없다. 세 번째 생물학(Biological) 기술은 유전학의 혁신으로 게놈프로젝트를 통해 시작된 것이 합성생물학의 영역으로 나아가 유전자 편집까지 나아가고 있다. 이러한 4차 산업혁명이 현재 이루어지고 있는 가운데, 4차 산업혁명과 관련된 정보를 수집하고 이해하여 다가오는 미래에 보다 적극적으로 대처할 수 있도록 도움을 주고자 한다. 이를 위해, 국내 정부 R&D현황을 네트워크 분석하여 시각적으로 보여준다. 이렇게 도출된 기술을 다시 4차 산업혁명 카테고리에 맞게 선별 후, 웹 검색트래픽을 활용하여 선별된 기술의 국내·외 관심 트렌드를 비교한다.

II. 이론적 배경 및 선행연구

본 연구에 대한 이론적 배경 및 선행연구는 <표 1>과 같이 정리된다. 먼저 최창우(2004)는 텍스트 마이닝과 사례기반 추론 방법론을 활용하여 연구제안서 스크리닝 시스템 알고리즘을 제안하였다. 하지만, 데이터마이닝 기법 제안서 및 연구개발 관련 자료에서 추출된 키워드 정보가 실질적인 내용에 대해 얼마나 잘 설명할 수 있는지에 대한 문제와 유사성이 존재할 경우, 두 문서 간의 관계를 정확히 파악하기가 어렵다는 한계점이 있다.

박동진 외(2009)는 국가과학기술지식정보서비스(NTIS)의 하위시스템 중 하나인 유사과제 검색시스템에 필요한 알고리즘을 제안하였으며, 이는 단순 키워드매칭 기반검색을 벗어나 연구과제 제안서를 기반으로 수행중/완료된 과제의 문서 및 연구 성과물과 비교하여 유사한 건들을 찾아내는 시스템이다. 이 알고리즘은 기존 키워드 매칭 방법에 비해 정확도가 향상되었으며, 우선순위별로 배열이 가능하다는 장점 및 검색시간은 줄이고 전문가

의 판단을 객관적이고 일관성 있게 판단할 수 있다는 장점이 있지만, 불용어 처리 방법 등에 대한 개선의 필요성, 실제 제안서와는 다른 연구보고서를 사용하여 실험한 부분들이 한계점으로 지적되었다.

정옥남 외(2011)는 NTIS에서 사용되는 유사과제 검색 시스템에 대한 추출방법 및 분석방법을 개선한 복합키워드 활용 모형을 제안하였다. 개선된 모형은 유사도의 정확도가 개선된 것을 입증하였으며, 과제 산출물인 연구보고서와 연구주제망이 유사도의 정확도에 유의미한 영향을 미친다는 사실 또한 확인하였다. 하지만, 해당 모형은 항목별 가중치 적용이 유사도에 많은 영향을 미치는 점을 고려했을 때, 체계적인 가중치 부여 방법이 부재하다는 한계점이 있다.

김주호(2012)는 국가R&D사업에 대한 유사검증방법을 위한 부처별 예산 요구서 항목에서 주요키워드를 추출 후 유사도를 측정하는 방식에 대해 사업별 예산 요구서 항목의 상이함과 기술수준 편차가 심해 유사성이 매우 낮게 측정되는 한계점을 지적하고 이를 극복하기 위해 과학기술표준분류를 이용한 모형을 생산한 후 이를 사업간 유사도 측정에 활용하는 방법을 제안하였다. 새로운 방식을 통한 실험 결과, 기존의 텍스트 기반 키워드 추출 방식보다 유사도가 높게 측정되어, 기존방식의 한계점을 극복하기 위한 대안이 될 수 있음을 확인하였다.

박근철 외(2011)는 기존 검색어 기반의 단순 검색이 가지고 있는 한계를 극복하기 위해 사업, 과제정보를 활용하여 유사 성과검색 모델을 제시하였다. 이 모델에서는 성과 및 과제, 사업정보를 기본항목과 특성화항목으로 나누었으며, 특성화항목의 각 정보에 가중치를 적용하였다.

김종배(2014)는 R&D 관련 특허를 조사, 수집하는 정부 R&D 특허기술 동향조사사업의 특허분석 DB를 활용하여 집합이론 및 확률이론을 기반으로 한 R&D 과제간의 유사도 분석 방법을 제시하였다. 이 모델에서 제안된 척도들을 통해 유사한 과제를 우선 순위화할 수 있지만 척도의 해석으로 우선순위가 높거나 낮음을 표현하지 못하는 한계점이 존재한다.

〈표 1〉 선행연구 및 연구내용

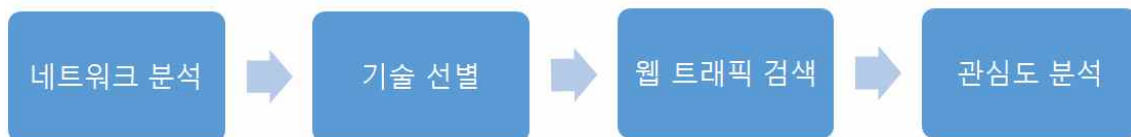
선행연구	연구내용
최창우(2004)	텍스트 마이닝과 사례기반 추론 방법론을 활용한 연구제안서 스크리닝 시스템 알고리즘 제안(Structural CBR 방법 사용)
박동진 외(2009)	NTIS의 유사과제 검색시스템에 필요한 알고리즘 제안
정옥남 외(2011)	NTIS의 유사과제 검색 시스템의 추출방법 및 분석방법을 개선한 모형 제안(복합키워드 사용)
김주호(2012)	국가 R&D 사업의 유사검증방법 개선 방법 제안(국가과학기술표준분류를 이용하여 각 사업의 고유한 벡터모델 생성 후 이를 사업간 유사도 측정에 사용)
박근철 외(2011)	유사 성과검색 모델 제안(사업정보를 기본항목/특성화 항목으로 구분 후 별도의 가중치 부여)
김종배(2014)	R&D 과제간의 유사도 분석 방법 제시(집합이론 및 확률이론 기반)

구글의 웹 검색 트래픽을 활용하기 위해 Google Trends(<https://trends.google.com>)를 사용하였다. Google Trends는 해당 주제에 대한 대중적 관심도를 보여주는 것으로 대상 기간 중 검색횟수가 가장 많았던 때를 100으로 정하고 시기별로 상대적 수치를 환산해 나타낸다.

III. 연구 사례 및 방법

본 연구에서는 네트워크 분석과 웹 트래픽 검색의 2가지 분석 방법이 활용되며, (그림 1)과 같이 연구는 4개의 단계로 진행된다. 먼저 NTIS에서 추출한 NST 산하 25개의 출연연이 2015년에 수행한 2,954개 과제에 대해 한글 키워드를 대상으로 동시단어출현 분석을 수행한다. 분석을 위해 2,954개의 과제를 정제한 결과, 한글 키워드 3,223개 중 동시출현횟수가 3회 이상인 키워드는 182개이며 이 키워드들이 동시 출현하는 과제는 총 761개로 나타났으며, 이를 기반으로 매트릭스를 생성하여 VOSviewer로 네트워크 분석을 실시하였다. 총 182개 키워드 중 연결강도가 낮은 데이터를 제외한 167개 단어를 기반으로 클러스터링 및 매핑 옵션은 전체과제 대상으로 분석한 것과 동일하게 진행하여 총 9개의 클러스터가 도출되었다.

여기서 도출된 군집결과를 토대로 클라우스 슈밥(Klaus Schwab)회장이 제시한 4차 산업혁명의 물리적, 디지털, 바이오 시스템에 해당되는 기술 3개를 선별하였다. 첫 번째로 물리적 시스템은 그래핀, 두 번째는 디지털 시스템은 빅데이터 그리고 마지막으로 바이오 시스템의 바이오마커를 선별하였다. 이렇게 카테고리별로 선별된 3개의 기술들에 대한 국내·외 관심도의 상이점을 확인하기 위해 구글의 웹 검색트래픽을 활용하였다. 이를 위해, Google Trends(<https://trends.google.com/trends>)에 해당 키워드를 각각 입력 후 최근 5년간의 트렌드를 국내(South Korea)와 국외(Worldwide)로 구분하여 수집하였다. 최종적으로 이렇게 수집된 정보를 비교하여 선별된 기술의 국내·외 관심도를 비교하였다.

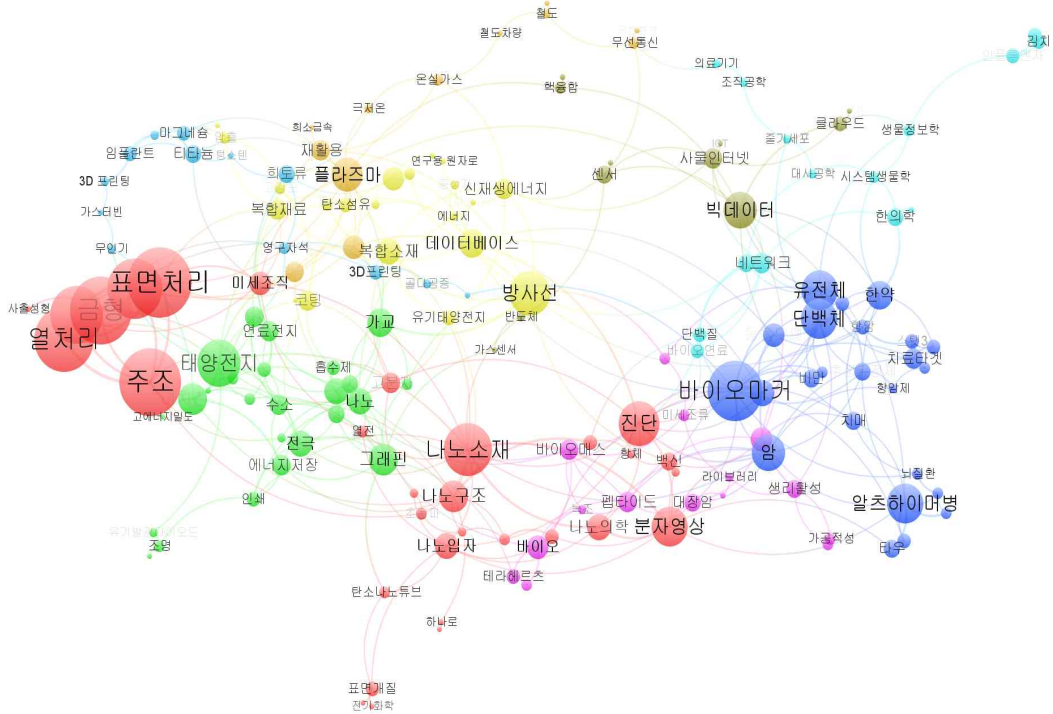


(그림 2) 연구방법의 도식화

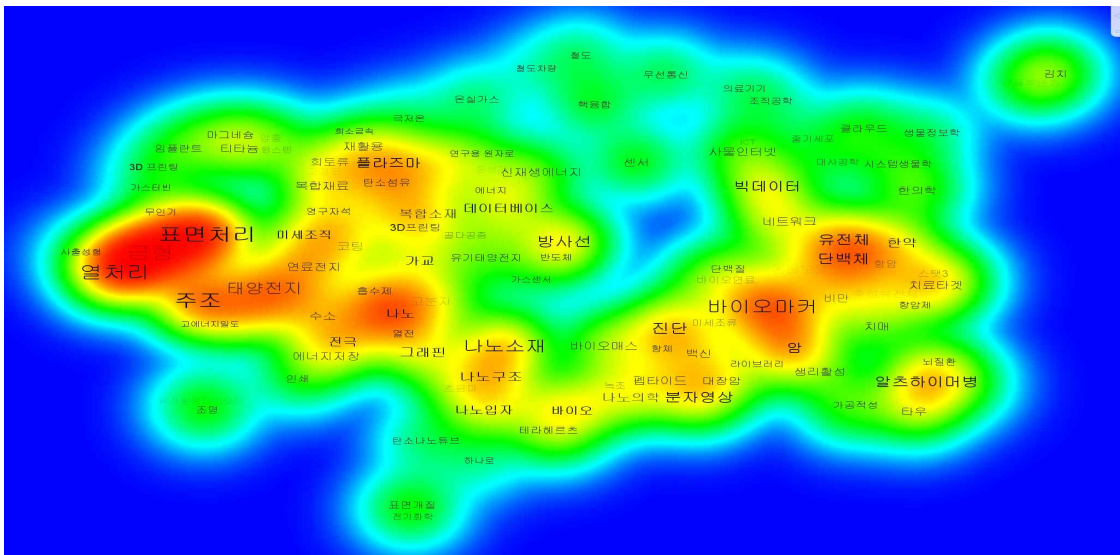
IV. 연구결과

3장에서 서술한 네트워크 분석결과 총 9개의 클러스터가 (그림 2)와 (그림 3)와 같이 도

출되었다.



(그림 2) 2015년도 25개 출연연 수행과제 대상 한글키워드 동시출현단어기반 네트워크 시각화



(그림 3) 2015년도 25개 출연연 수행과제 대상 한글키워드 동시출현단어기반 밀도 시각화

그 중, 첫 번째 클러스터는 ‘열처리’, ‘금형’, ‘주조’ 등 기계 및 제품 관련 키워드들이었으며, 나노소재를 포함하고 있었다. 그러나 금속분야 및 나노분야의 거리가 차이는 것으로 보아 ‘소재’ 라는 같은 주제를 공유하고 있지만 유사성 측면에서는 각각

별도의 주제에 소속된 것으로 해석할 수 있다. ‘소재’에 관한 클러스터링 사이에 위치한 두 번째 클러스터에서는 ‘태양전지’ 및 ‘그래핀’이 중심 단어인 것을 확인할 수 있다. 태양전지는 금속 및 나노소재가 주성분이 되며, 그래핀은 탄소나노소재로서 첫 번째 클러스터 사이의 합성 소재에 관한 기술분야인 것으로 나타났다.

한편, ‘나노소재’는 ‘바이오마커’를 중심으로 ‘유전체’, ‘암’, ‘알츠하이머’ 등이 주요 키워드로 이루어진 세 번째 클러스터를 연결하는 역할을 하는데, 이는 ‘바이오마커’의 주성분은 나노소재이며, 기술이 지닌 역할은 생명과학과 연결되는 점을 반영한다. 네 번째 클러스터는 ‘방사선’을 중심으로 ‘복합소재’, ‘신재생 에너지’, ‘연구용 원자로’ 등의 단어들로 구성된 군집을 형성하였으나, 앞서 살핀 클러스터보다는 특징적인 면이 상대적으로 낮은 것으로 나타났으며, 이후의 여러 클러스터들 또한 클러스터 고유의 특성이 상대적으로 불명확하였다. 그러나 마지막 클러스터인 ‘빅데이터’ 중심으로 연결되어 있는 키워드들의 경우, ‘사물인터넷’, ‘클라우드’, ‘센서’ 등을 포함하며, 향후 4차 산업혁명과 연결되는 최신 ICT 관련기술들을 설명하고 있는 것으로 분석된다. 이상의 군집 도출결과를 정리하면 다음의 <표 2>와 같다.

<표 2> 군집별 주요 키워드

군집 번호	군집의 주요 키워드
1	주조, 나노소재, 표면처리
2	태양전지, 연료전지, 그래핀
3	바이오마커, 유전체, 암
4	방사선, 데이터베이스, 복합소재
5	빅데이터, 사물인터넷, 클라우드, 센서
6	플라즈마, 재활용
7	네트워크, 단백질, 한의학, 김치
8	3D프린팅, 임플란트, 티타늄
9	펩타이드, 바이오, 바이오매스

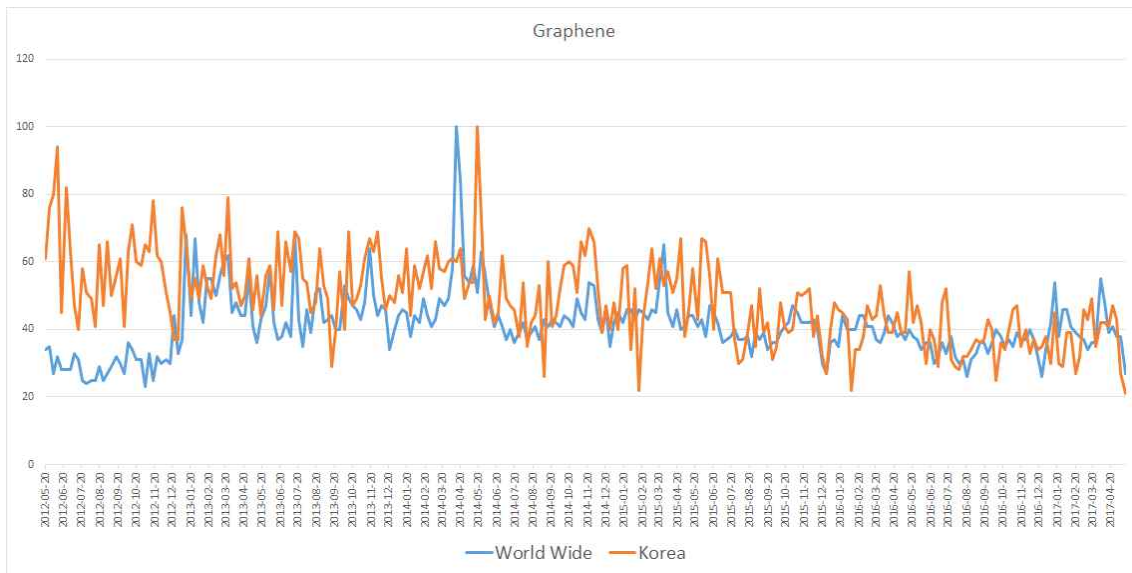
상기 클러스터링 분석으로 도출된 군집들의 군집별 핵심 키워드를 보면 ‘주조’, ‘태양전지’, ‘바이오마커’, ‘방사선’, ‘빅데이터’ 등으로 나타났고, 군집 간 연결성이 높은 키워드는 ‘방사선’, ‘그래핀’, ‘나노소재’ 등으로 나타났다. 그 중에서 다양한 과제수행기관을 보유한 분야를 확인하여 유사연구영역 분석 대상 주제를 도출한 결과 최종적으로 ‘태양전지’, ‘바이오마커’, ‘빅데이터’ 세 가지 분야가 도출되었고, 연결성이 강한 키워드에서는 상대적으로 영역구분이 용이한 ‘그래핀’을 분석대상으로 선정했다. 또한 향후 4차 산업혁명과 관련하여 발전가능성이 높은 ‘사물인터넷’은 연구개발주기 측면에서 추가로 선정했다. 이상 다섯 가지 연구 분야를 분석 대상 주제로 선정했는데, 상대적으로 큰 군집이었던 ‘방사선’과 ‘주조’를 제외한 것은 관련 과제는 다양하지만, 유관 과제 수행기관이 다양하지 않았기 때문에 제외했다.

<표 2>와 같이 도출된 군집별 주요 키워드를 <표 3>과 같이 클라우드 슈밥의 3가지 카테고리에 맞추어 각각의 기술을 선별하였다.

〈표 3〉 4차 산업혁명 기술의 카테고리화 선별 기술

4차 산업혁명 카테고리	선별 기술
물리적 시스템	그래핀
디지털 시스템	빅데이터
바이오 시스템	바이오마커

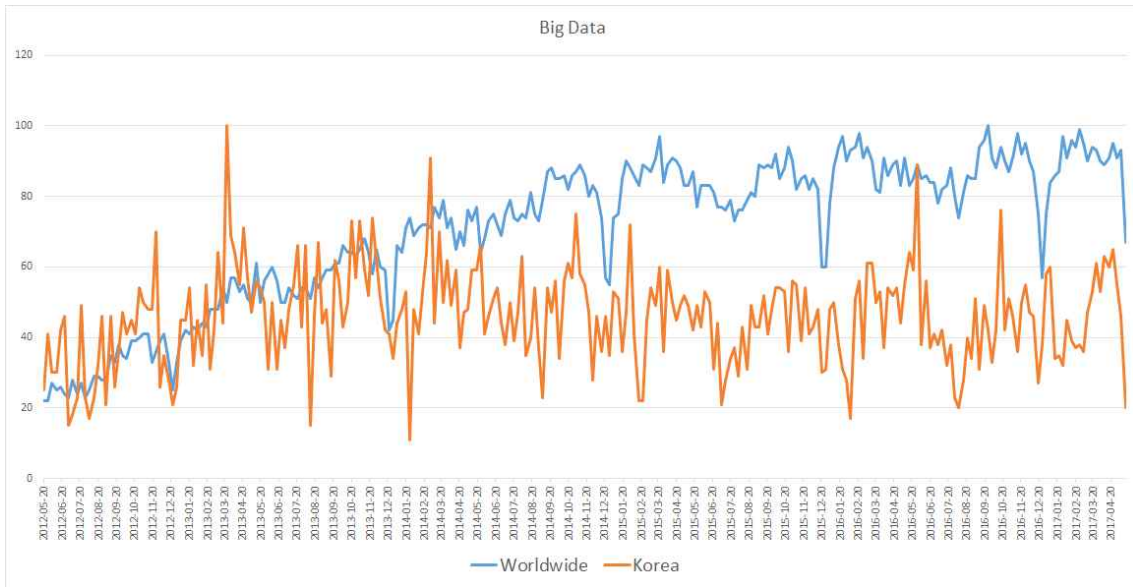
선별된 기술들을 키워드로 하여 5년간의 관심 트렌드를 비교한 결과는 다음 (그림4), (그림 5), (그림 6)과 같다. 먼저 (그림 4)의 그래핀(Graphene) 검색결과 국내의 경우 2012년 관심도가 해외와 비교하였을 때 2012년 5월부터 7월 사이에 전체 관심도 100기준으로 60부터 약90까지 상승하여 해당 기술에 대해 매우 관심이 높았던 것으로 추정할 수 있다. 이후 높았던 관심도는 등락을 반복하다 2014년 중순경에 최고 관심도인 100을 기록한다. 여기서 주목할 점은 해외의 관심도가 먼저 100을 기록하고 그 후 얼마 후 국내 관심도가 마찬가지로 100을 기록했다. 이후 국내와 해외의 관심도는 비슷하게 등락을 거듭하며 우하향 하였다.



(그림 4) Google Trends “Graphene”검색결과 국내·외 비교결과

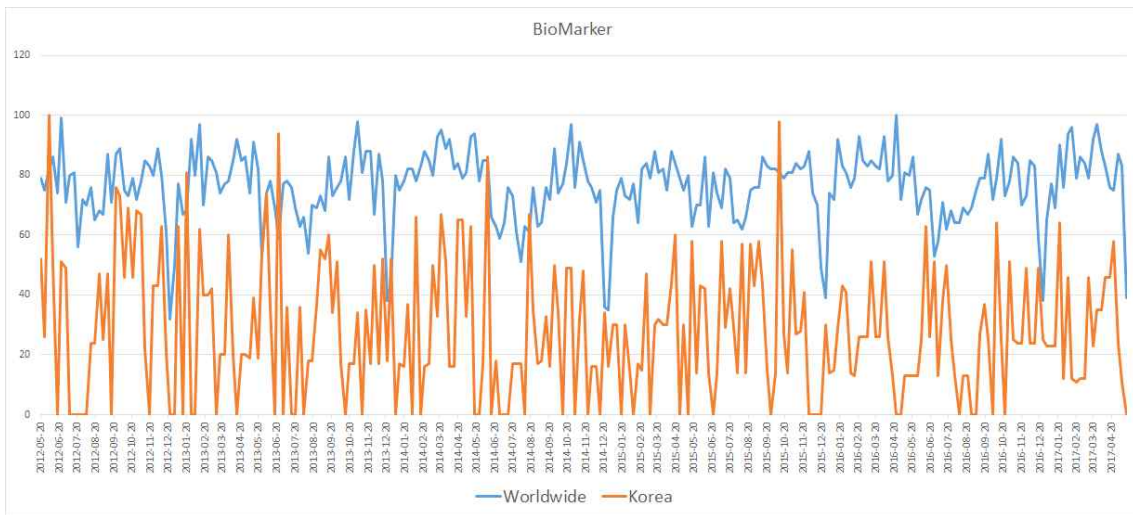
두 번째로 (그림 5)의 빅데이터(Big Data) 검색결과 2012년의 관심도는 국내와 해외 마찬가지로 20의 낮은 수준으로 시작하여 등락을 거듭지만, 점차적으로 우상향 하는 모습을 보여주었다. 특히 국내의 경우 2013년 3월경에 빅데이터에 대한 폭발적인 관심을 보였으며, 해외의 경우 등락을 거듭하며 꾸준히 우상향하여 2016년 9월경에 최대 관심치를 갱신하였다.

2017 한국기술혁신학회 춘계학술대회



(그림 5) Google Trends “Big Data”검색결과 국내·외 비교결과

세 번째로, 바이오마커(BioMarker)의 경우, 2012년부터 꾸준히 높은 관심도로 시작하여, 등락을 거듭하지만 상당히 높은 수준의 관심도를 유지하였지만, 국내의 경우 검색 자체가 매우 부족하여 0의 관심 수치를 나타내는 기간을 있을 정도로 해외와 비교하였을 때 저조한 관심도를 나타내었다.



(그림 6) Google Trends “BioMarker”검색결과 국내·외 비교결과

선별된 3개의 기술에 대한 트래픽 검색결과 우리나라는 그래핀의 경우 국내와 해외의 관심 수준과 패턴이 유사함을 확인하였다. 하지만, 빅데이터의 경우 전체적인 관심도 자체가 상대적으로 해외와 비교했을 때 낮은 것으로 나타났다. 마지막으로 바이오마커의 경우, 해외는 매우 높은 수준의 관심도를 유지하고 있지만 국내의 경우 분석을 위한 불

롭자체가 부족할 정도로 국내의 관심도가 매우 낮은 것으로 나타났다.

V. 토의 및 결론

본 논문은 4차 산업혁명과 관련된 기술에 대한 정보를 수집하고 이해하여 다가오는 미래에 보다 적극적으로 대처할 수 있도록 도움을 주고자 하였다. 이를 위해, 국내 R&D과제에 대한 네트워크 분석을 하여 전체적인 과제들의 현황을 보여주고 이중, 어떤 과제 키워드들이 어떤 형태로 밀집되어 지고 있는지 시각적으로 클러스터링하여 보여주었다. 이렇게 클러스터링된 정보를 통해, 어떤 기술들이 서로 교류하고 있는지, 또 어떤 키워드의 기술들을 중심으로 그룹으로 밀집하고 있는지 확인할 수 있었다. 총 9개의 클러스터가 (그림 2)와 같이 도출되었다. 이 중 3개의 클러스터에서 4차 산업혁명의 3가지 카테고리에 맞게 각각 1개의 주요 키워드를 선별하였다. 선별된 키워드는 <표 3>의 그래핀, 빅데이터, 바이오마커이며, 3개의 키워드에 대한 관심도를 알아보기 위해, 각각의 키워드를 구글 트렌드에 검색하여 5년간의 검색 트래픽을 받았다. 이렇게 받은 검색 트래픽을 서로 국내와 해외로 나누어 비교한 결과, 그래핀의 경우 국내와 해외의 관심 수준과 패턴이 유사함을 확인하였다. 하지만, 빅데이터의 경우 전체적인 관심도 자체가 상대적으로 해외와 비교했을 때 낮은 것으로 나타났다. 마지막으로 바이오마커의 경우, 해외는 매우 높은 수준의 관심도를 유지하고 있지만 국내의 경우 분석을 위한 볼륨자체가 부족할 정도로 국내의 관심도가 매우 낮은 것으로 나타났다. 이러한 결과는 그래핀과 빅데이터가 대중적으로 많이 소개되었기 때문에 상대적으로 바이오마커의 관심도가 다른 2가지 기술에 비해 많이 낮은 것으로 사료된다. 하지만, 해외의 관심도를 고려하였을 때, 그래핀을 제외한 나머지 기술들에 대한 국내 관심도는 상대적으로 떨어지는 것을 확인하였다. 본 논문에서 수행한 과제 현황의 네트워크 분석결과와 검색 트래픽 분석 결과는 향후 다년간의 과제 현황 변화 네트워크 분석 및 관련된 기술들에 대한 관심도 변화와 과제 현황 관련성 연구를 위해 활용될 수 있을 것으로 기대된다.

참고문헌

- 권오진 (2009) “과학계량학 기반 과학기술 지식 네트워크 구조 분석 모델 개발”, 서울 시립대 박사학위논문.
- 김종배 (2014) “특허 정보를 활용한 R&D 과제 유사도 측정 모델”, 「한국정보통신학회 논문지」, 18(5), pp 1013-1021.
- 김주호 외 (2012), “R&D과제의 기술분류를 이용한 사업간 유사도 분석 기법에 관한 연구”, 「Journal of Digital Contents Society」, 13(3), pp. 317-324.
- 박근철 (2008) “연구성과 활용 극대화를 위한 유사성과검색에 관한 연구”, 「한국인터넷정보학회 학술발표대회 논문집」, 9(2), pp. 655-658.
- 박동진 외 (2009), “유사과제과약을 위한 검색 알고리즘의 개발에 관한 연구”, 「한국콘텐츠학회논문지」, 9(11), pp. 54-62.
- 심준섭 (2011), “언어네트워크 분석 기법을 활용한 갈등프레임의 분석”, 「한국행정연구」, 20(2), pp 183-212.
- 이재윤·최상희 (2013), “계량정보학분야의 협력연구 네트워크 및 문헌네트워크 분석 : 국가, 기관, 문헌단위 분석”, 「정보관리학회지, 30(1), pp.179-191.
- 정옥남 외 (2011), “과제 유사도 측정 개선모형에 관한 실증적 연구”, 「디지털콘텐츠학회논문지」 v.12 no.4, pp. 457-465.
- 조재인 (2011), “네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향 분석”, 「정보관리학회지」. 28(4), pp 65-83.
- 최창우 (2004), “R&D proposal screening 시스템 설계: Text-Mining 접근”, 서울대학교 대학원 석사학위논문.
- 최영출·박수정 (2010), 지역교육청 기능 재검토를 위한 초등학교 공문서 분석: 사회적 네트워크 분석방법의 적용. 「지방정부연구」. 14(3), pp 165-188.
- 최영출·박수정 (2011), “한국행정학의 연구경향분석: 네트워크 텍스트 분석방법의 적용”, 「한국행정학보」, 45(1), pp 123-138.
- Hsieh, H.F., & Sarah, E. S., (2005), “Three Approaches to Qualitative Content Analysis”, *Qualitative Health Research*, 15(5), pp 1277-1288.
- Google Trends, <https://trends.google.com/trends>
- Nees, J. Eck., Ludo W., (2010), “Software survey: VOSviewer, a computer program for bibliometric mapping”, *Scientometrics*, 84(2), pp 523-538.