

2020년 1차

**정보기술·사업관리
산업인적자원개발위원회(ISC)
이슈리포트
(ISSUE REPORT)**

빅데이터 분야 산업·인력 이슈 분석
- 뉴스 빅데이터 네트워크 분석을 중심으로 -

●●● 목 차 ●●●

빅데이터 분야 산업·인력 이슈 분석 - 뉴스 빅데이터 네트워크 분석을 중심으로 -

(요약)

I. 개 요	1
II. 이슈 분석 방법	3
III. 이슈 분석 결과	5
IV. 결론 및 시사점	16

비상업 목적으로 본 보고서에 있는 내용을 인용 또는 전재할 경우 내용의 출처를 명시하면 자유롭게 인용할 수 있으며, 보고서 내용에 대한 문의는 아래와 같이 하여 주시기 바랍니다.

정보기술·사업관리 산업인적자원개발위원회 사무국
- 송주호 책임 (02-2188-6948, juho@sw.or.kr)

※ 본 리포트는 한국기술교육대학교 융합학과 이호 교수와 공동으로 연구한 결과임

□ 빅데이터 분야 산업·인력 이슈 분석

□ 개 요

- 국내 기업의 빅데이터 이용률은 '17년 7.5%, '18년 9.5%(NIA자료) 수준에 미치며, 빅데이터 기술의 수요 역시 산업현장에서 여전히 높게 나타나고 있음
- 4차 산업혁명 도래에 따라 데이터 확보 및 활용이 개별 기업의 생존을 넘어 국가 간 데이터 주도권 경쟁으로 번지고 있는 '글로벌 데이터 전쟁' 확산
- 결국, 스스로 데이터를 확보하고 AI 기술로 분석, 예측하여 서비스를 창출할 수 있는 생태계를 구축하는데 필요한 역량을 가진 인재양성은 선택이 아닌 필수
- 이에 따라, 가장 최근의 미디어 빅데이터를 활용하여 빅데이터 분야의 이슈에 대해 탐색적으로 분석하고 인적자원개발 시사점을 도출하고자 함

□ 이슈 분석 방법

- 언론사 뉴스 최근 1년간('19.1.1부터 '19.12.31까지)의 '빅데이터' 단어 포함 기사 36,693건을 추출하고 데이터 전처리 작업을 수행
- 네트워크 분석 툴 'Gephi' 를 활용, 310,028개의 단어(노드, Node)와 3,055,574개의 단어 조합(엣지, Edge)으로 이루어진 네트워크를 구성
- 네트워크의 특성을 나타내는 Degree, Modularity, Eigenvector Centrality 등을 바탕으로 분석하고 네트워크 그래프를 시각화 함

□ 이슈 분석 결과

- 네트워크는 모듈성을 기준으로 총 11개의 군집으로 분류되었으며, 전체 네트워크상의 비중 10% 이상을 차지하고 있는 상위 4개의 군집을 대상으로 내용분석을 실시함

○ 군집 2. 'AI'

- AI 분석을 통해 결과를 이해할 수 있는, 설명 가능한 AI 분석 기술이 첨단 트렌드로 자리매김 하면서 AI 인력 수요가 증가함
- 그러나 AI 분석 기술을 개발하는 인원은 초고급의 인력이며, 오히려 AI 기술을 활용하여 분석이 가능한 빅데이터 분석 인력에 대한 수요가 훨씬 더 많기 때문에 이러한 점을 고려하여 인력양성 정책이 구분되어 수립·추진되어야 함
- 엔지니어와 분석가의 구분을 명확히 하고 이에 따라 각기 개별 특성을 고려한 인재양성 정책 수립이 필요함
- 특히, 빅데이터 산업 활성화를 위해서는 분야별로 분석이 가능한 융합적인 인재양성이 시급함

○ 군집 3. '지자체'

- 5G와 IoT, 인공지능 등 스마트시티 사업이 지자체 단위의 빅데이터 기술 확산에 매우 긍정적으로 작용함
- 특히 지자체별로 미세먼지 대응, 부동산 예측 등 지역문제 해결이나 전망을 위해 빅데이터 기술이 활용되고 있음
- 그러나 지자체 대부분은 분야별로 각기 다른 분야의 빅데이터 분석 전문가가 필요하다는 특성과 국내에 관련 전문가가 부족하다는 현실 때문에 전담 조직이 구성되거나 관련 인력양성이 되기 어려움
- 효과적인 빅데이터 사업 활성화를 위해서는 중앙정부뿐만 아니라 지자체에도 전문인력으로 구성된 전담조직 구성이 필요함

○ 군집 4. '금융 빅데이터'

- 스마트폰과 핀테크 확산으로 금융권의 빅데이터 활용 서비스와 더불어 카카오 등 비금융권 IT기반 기업이 제공하는 빅데이터 활용 금융서비스가 경쟁을 통한 긍정적 산업 발전에 기여하여 국내에 금융 빅데이터 활용이 활성화되었음
- 특히, 빅데이터를 활용한 소비패턴 분석, 신용등급 분석, 부정거래탐지 서비스 등 빅데이터의 금융 분야 활용이 빠르게 확산되고 있음

□ 결론 및 시사점

○ 빅데이터가 언급된 최근 1년간의 미디어 데이터 분석을 통해 산업현장의 주요 이슈는 다음과 같이 도출되었음

- 정부, 기업 모두 빅데이터는 혁신성장의 주요 동력이라는 공감대는 이미 형성되어 있으나, 여전히 개인정보 및 데이터 활용 관련 규제, 전문인력 부족으로 빅데이터 활용 및 활성화에 한계가 존재해왔음
- 그러나 최근 2020년 1월 9일 데이터3법 개정안이 국회 발의 약 14개월 만에 통과되면서 빅데이터 활용과 서비스 활성화 등 데이터 경제 발전에 대한 기대가 높아지고 있음
- 특히, 유전체, 바이오, 헬스케어, 반도체, 금융, 제조 분야 스마트팩토리와 스마트시티 등 분야에서 도메인 지식과 AI 기술을 활용한 빅데이터 분석 및 활용, 그리고 관련 융합인력 양성의 중요성이 높아지고 있음
- AI의 경우, 인력양성의 수요와 인력을 육성하는 교수자의 수요가 급증하고 있으나, AI 원천기술, 알고리즘 개발 등의 초고급 인력 수요는 매우 제한적임
- 오히려 AI 기술을 활용하여 빅데이터 분석, 빅데이터 엔지니어 등 도메인 분야의 활용을 목적으로 한 융합인재 양성이 효과적임

○ 빅데이터 산업 활성화를 위해서는 빅데이터가 활용되는 분야별로 도메인 지식을 이해하고 분석이 가능한 융합적 인재양성이 중요함

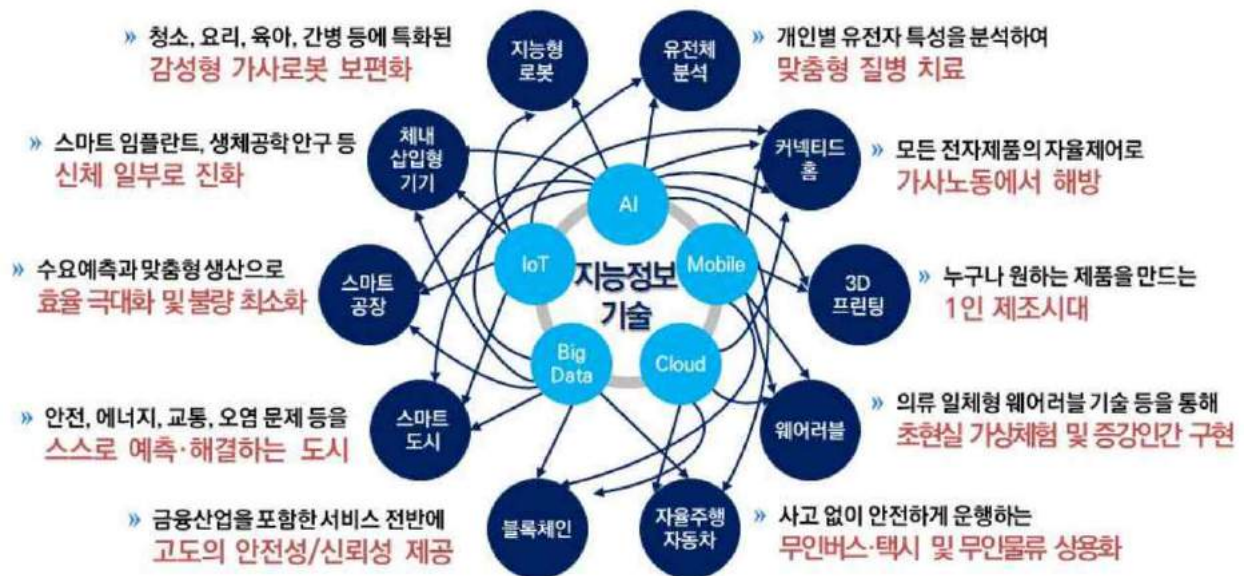
- 한국이 강점을 가지는 제조 분야 혁신을 위해서는 스마트팩토리 도입과 관련된 경험과 지식을 함양한 인재 양정으로 재도약의 발판을 마련해야 함
- 스마트시티 사업은 도시 특성을 고려한 문제 해결, 미래도시 계획, 첨단 기술의 활용 등 분야별 세부 전문성을 갖춘 빅데이터 분석이 가능한 전담 조직의 구성이 관건임
- 금융 빅데이터의 경우, 데이터분석 기술과 정보기술을 바탕으로 한 비금융권 IT기업들의 참여로 빅데이터를 활용한 금융서비스 산업에 긍정적 발전이 이루어지고 있음
- 관련하여, 핀테크 및 금융 빅데이터 융합 전문인력 양성이 활발하게 이루어진다면 관련 서비스 고도화와 산업의 발전을 통해 글로벌 경쟁력을 갖출 수 있을 것으로 기대됨

1. 개요

○ 4차 산업혁명 시대 빅데이터 인재 양성은 선택이 아닌 필수¹⁾

- 고도화된 정보통신기술 인프라(IoT, Cloud, Big Data, Mobile)를 통해 생성·수집·축적된 데이터와 인공지능(AI)이 결합한 지능정보기술이 경제·사회·삶 모든 분야에 보편적으로 활용됨으로써 새로운 가치가 창출되고 발전하는 사회도래

그림 1. 지능정보기술의 활용 확산



※ 출처: 정부 관계부처 합동, 지능정보사회 중장기 종합대책(안), 2016.12.

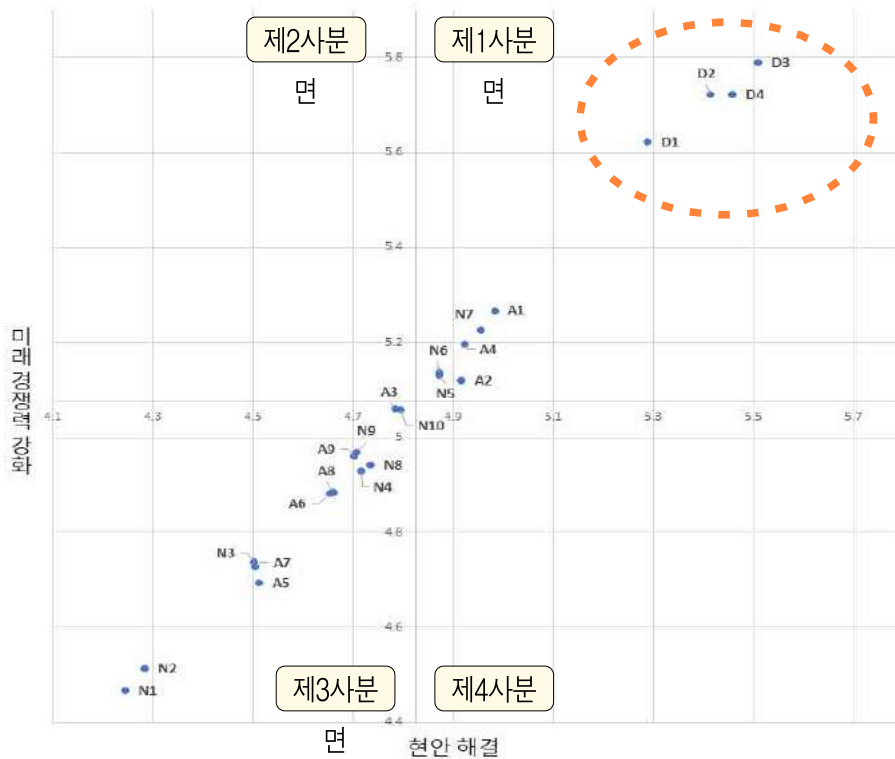
- 4차 산업혁명 도래에 따라 데이터 확보 및 활용이 개별 기업의 생존을 넘어 국가 간 데이터 주도권 경쟁으로 번지고 있는 ‘글로벌 데이터 전쟁’ 확산
- 데이터와 지식이 기존 생산요소(노동, 자본)보다 중요해지고 다양한 제품·서비스 융합으로 이종 산업간 경계가 붕괴되며, 지능화된 기계를 통한 자동화가 지적노동 영역까지 확장되는 등 경제·사회 전반에 혁신적인 변화가 발생
- 모든 사물에서 데이터가 끊임없이 생성되고 네트워크를 통해 실시간으로 전달(IoT, Mobile)되며 수집된 데이터를 효율적으로 저장하고 그 의미를 분석(Cloud, Big Data)할 수 있는 기업이 글로벌 시장을 견인
- 결국, 스스로 데이터를 확보하고 AI 기술로 분석, 예측하여 서비스를 창출할 수 있는 생태계를 구축하는데 필요한 역량을 가진 인재 양성은 선택이 아닌 필수

1) 지능정보사회 중장기 종합대책(안)(정부 관계부처 합동, 2016.12.)의 내용을 인용하여 작성

○ 이슈분석 배경 및 목적

- 빅데이터 활용은 4차산업과 데이터경제 활성화의 근간이며, 최근 데이터 3법의 개정안 의결('20.1.9)로 데이터 산업이 활성화, 관련 양질의 일자리 창출이 기대되고 있음
- 정보화진흥원에 따르면 국내 기업의 빅데이터 이용률은 '17년 7.5%, '18년 9.5% 수준에 미치며, 최근 연구보고서의 4차산업 핵심기술 수요조사 결과 아래 그림과 같이 빅데이터 분야의 기술 수요가 여전히 높게 나타나고 있음²⁾
- (목적) 최근 생산축적된 미디어 빅데이터를 활용, 빅데이터 관련 분야 산업 · 인력 현황과 이슈에 대해 탐색적으로 분석하고 인적자원개발 시사점을 도출함

그림 2. 미래경쟁력강화와 현안해결을 위한 빅데이터(D), 인공지능(A), 블록체인(N) 기술수요 조사결과



구분	D1	D2	D3	D4	A1	A2	A3	A4	A5	A6	A7	A8	A9	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
소분류	수집방법	저장처리/관리	분석/연계	활용/시간화	머신러닝	추론/지식표현	언어지능	시간지능	청각지능	행동/소셜지능	상황/상황이해	지능형 에이전트	표현형 인공지능(AI)	분산연산	고성능 분산연산	스마트계약	암호성	데이터 연계	IoT 연계	플랫폼 연계	관리혁신	사업연계	인프라
분류	빅데이터(D)				인공지능(A)									블록체인(N)									

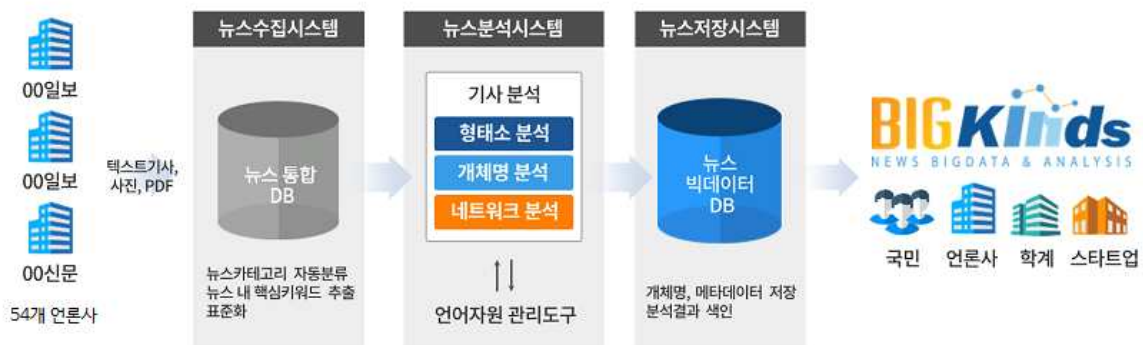
2) 과기정통부(2020), 중소SW기업 DNA(지능화기술) 내재화를 위한 기술·인력 운영전략 연구, 방송통신정책연구 2019-0-01443, 2020.2.2.

2. 이슈 분석 방법

□ 분석 데이터

- 54개 매체의 뉴스 데이터를 보유하고 있는 한국언론진흥재단의 ‘빅카인즈3)’ 서비스를 활용하여, ‘빅데이터’가 언급된 뉴스 36,693건을 분석 대상으로 함
 - 뉴스 추출 기간은 최근 1년, ‘19년 1월 01일부터 ‘19년 12월 31일까지로 한정

그림 3. 빅카인즈 서비스 개념도



※ 출처: 빅카인즈 서비스 홈페이지(<https://www.bigkinds.or.kr/>)

□ 분석 방법

- 각 뉴스 내에 추출된 키워드의 네트워크 분석을 위한 전처리 및 통계분석
 - 이음동의어, 유사단어처리 및 무의미한 단어 제외 등 데이터 전처리과정을 통해 총 310,028개의 단어(노드, Node)와 3,055,574개의 단어 조합(엣지, Edge)으로 이루어진 네트워크가 구성됨
 - 뉴스를 기반으로 한 네트워크의 경우, 단어간의 방향성이 존재하지 않기 때문에 무방향성(Undirected)의 특성을 나타내고 있어 이를 고려한 분석을 수행함
 - ※ 네트워크 분석 및 시각화를 위해 분석툴로 Gephi 0.9.2를 활용함
 - 뉴스의 특성상 매우 다양한 단어의 조합이 가능하고, 단어 연결정도(Degree)가 상대적으로 매우 낮은 연결은 네트워크상에서 큰 의미를 주지 않는다고 판단되어 추가적인 분석에서 제외하였음
 - ※ 단어 연결정도 3,000이상의 데이터를 기준으로 추가적인 분석을 수행함

3) 빅카인즈 서비스(BIGKinds)는 다양한 언론사로부터 수집한 뉴스로 구성된 통합 데이터베이스에 빅데이터 분석 기술을 접목하여 만든 새로운 뉴스 분석 서비스로, 1990년부터 현재까지 54개 매체의 약 6천만 건 뉴스 콘텐츠를 빅데이터화 하였음

○ 네트워크 구성

- 네트워크상의 주요 의미 파악을 위해 고유벡터중심성을 이용한 분석과 커뮤니티 구조 파악을 위해 모듈성을 기반으로 분석을 수행함
- 네트워크의 특성을 나타내는 연결정도(Degree), 모듈성(Modularity), 고유벡터 중심성(Eigenvector Centrality) 등을 바탕으로, 네트워크 그래프로 시각화 될 수 있음
 - ✓ **연결정도(Degree)** : 현재 노드가 얼마나 많이 다른 노드와 직접적으로 연결되어 있는지를 나타내는 수치
 - ✓ **모듈성(Modularity)** : 특정 노드가 일정 커뮤니티와는 가깝고 타 커뮤니티와는 거리가 먼 정도를 나타낸 지수로서 네트워크상의 커뮤니티를 구분하는 지표로 사용됨
 - ✓ **고유벡터 중심성(Eigenvector Centrality)** : 고유벡터 중심성(Eigenvector Centrality): 중요한 노드에 연결된 노드일수록 그 중요도가 높다는 전제로 수립된 중심성 지수 (0.0-1.0사이의 값으로서, 1.0에 가까울수록 높은 중요도를 나타냄)
 - ※ **중심성**: 네트워크상에서 어떠한 노드가 중요한지를 파악하기 위한 척도로서 연결 중심성(Degree Centrality), 매개 중심성(Betweenness Centrality), 고유벡터 중심성(Eigenvector Centrality), 페이지랭크(PageRank) 등 다양한 방법이 사용됨
 - ✓ **네트워크 그래프** : 네트워크의 특성을 시각화한 그래프로서 원(노드) 크기는 해당 네트워크상에서 상대적인 고유벡터 중심성의 값을 나타낸 것으로 원 크기가 클수록 중요한 단어임을 의미하며, 원을 연결하는 선(edge)은 동시에 등장하는 단어간의 관계를 표현한 것으로 굵을수록 높은 연결정도(Degree)를 가짐을 의미
- 모듈성에 기반에 구분된 군집은 포함된 키워드들을 살펴 분석가가 적절한 군집의 명칭을 선정하게 됨
- 구성된 네트워크에서 군집에 포함된 노드의 수가 많은 경우 한 번에 내용을 파악하기 어렵기 때문에, 구체적인 내용 분석을 위해 모듈성 기반 군집 알고리즘을 다시 한 번 활용하여 추가 군집분석을 수행하는 계층적 군집 분석을 수행함

3. 이슈 분석 결과

□ 네트워크 구성 특징

○ 주요 단어 : 네트워크상에서 어떠한 노드가 중요한지를 나타내주는 고유벡터중심성을 기준으로 다음의 순서로 단어들이 도출되었음

- ‘빅데이터(연결정도:23653)’, ‘AI(23,033)’, ‘기업(12,398)’, ‘4차산업혁명(14,072)’, ‘글로벌(12,500)’, ‘혁신성장(11,652)’, ‘정부(12,705)’, ‘한국(12,489)’, ‘데이터(11,107)’, ‘서울(13,064)’ 등

○ 군집 분류 : 네트워크상에서 커뮤니티를 구분하는 지표인 모듈성을 기준으로 총 11개의 군집으로 분류되었음

- 아래 표는 5% 비중 이상의 군집 8개의 포함 노드를 나타냄

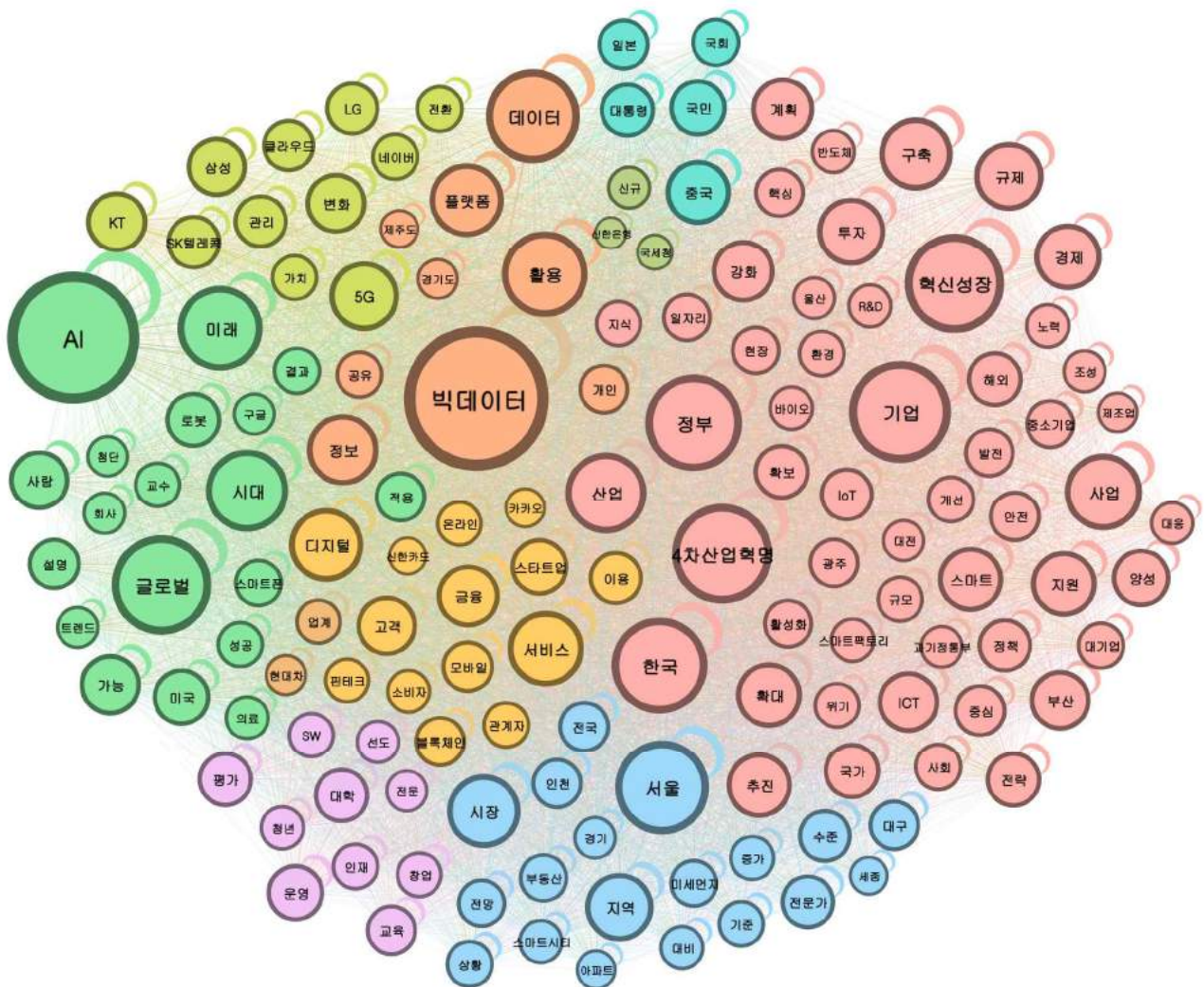
표 2. 군집별 포함 노드

구분	비중(%)	노드명
군집1	18.65	기업, 4차산업혁명, 혁신성장, 정부, 한국, 산업, 사업, 구축, 규제, 투자, 경제, 확대, 지원, 추진, 스마트, 계획, 강화, ICT, 부산, 양성, 국가, 확보, 해외, 전략, 정책, IoT, 중심, 안전, 광주, 발전, 핵심, 활성화, 중소기업, 현장, 일자리, 사회, 개선, 대기업, 환경, 규모, 지식, 위기, 스마트팩토리, 노력, 바이오, R&D, 반도체, 대응, 조성, 울산, 과기정통부, 대전, 제조업
군집2	15.59	AI, 글로벌, 미래, 시대, 가능, 사람, 미국, 로봇, 설명, 적용, 스마트폰, 결과, 성공, 교수, 의료, 회사, 트렌드, 구글, 첨단
군집3	13.78	서울, 시장, 지역, 전문가, 수준, 전국, 대구, 인천, 부동산, 전망, 미세먼지, 상황, 기준, 증가, 스마트시티, 대비, 경기, 아파트, 세종
군집4	12.27	서비스, 디지털, 금융, 고객, 스타트업, 이용, 모바일, 블록체인, 관계자, 온라인, 핀테크, 소비자, 카카오, 신한카드
군집5	7.76	운영, 평가, 대학, 교육, 인재, SW, 창업, 청년, 선도, 전문,
군집6	7.40	5G, KT, 변화, 삼성, SK텔레콤, 관리, LG, 클라우드, 네이버, 전환, 가치
군집7	7.18	중국, 국민, 대통령, 일본, 국회
군집8	6.97	빅데이터, 데이터, 활용, 플랫폼, 정보, 개인, 공유, 경기도, 제주도

○ 분석을 통해 도출된 최종 네트워크 그래프

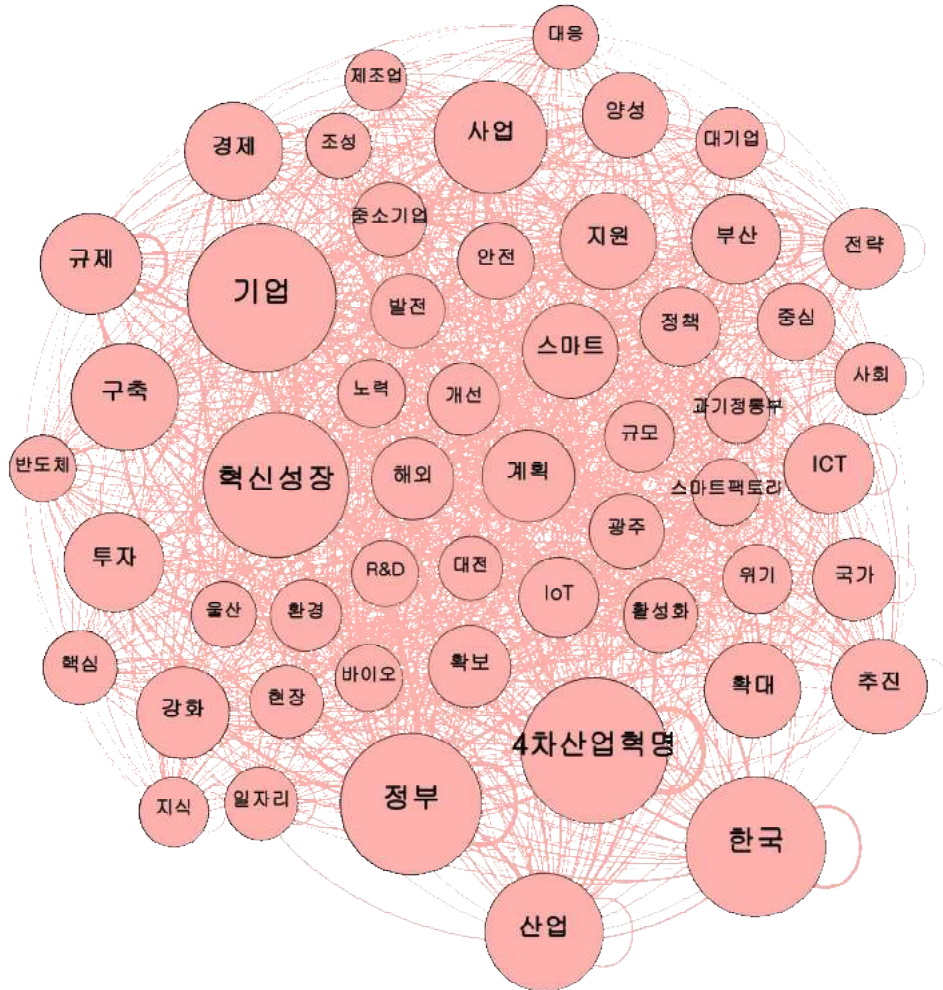
- 그래프상의 노드별 색깔은 모듈성을 기반으로 구분된 각각의 군집을 구분하기 용이하도록 임의로 부여된 것임
- 노드(단어)의 크기는 노드 간의 상대적인 중요도를 나타내는 것으로 크기가 클수록 높은 중요도를 의미함
- 네트워크 그래프의 총 11개의 군집 중, 네트워크상의 비중 10% 이상을 차지하는 상위 4개의 군집을 대상으로 추가적인 내용분석을 실시함

그림 4. 빅데이터 관련 네트워크 그래프



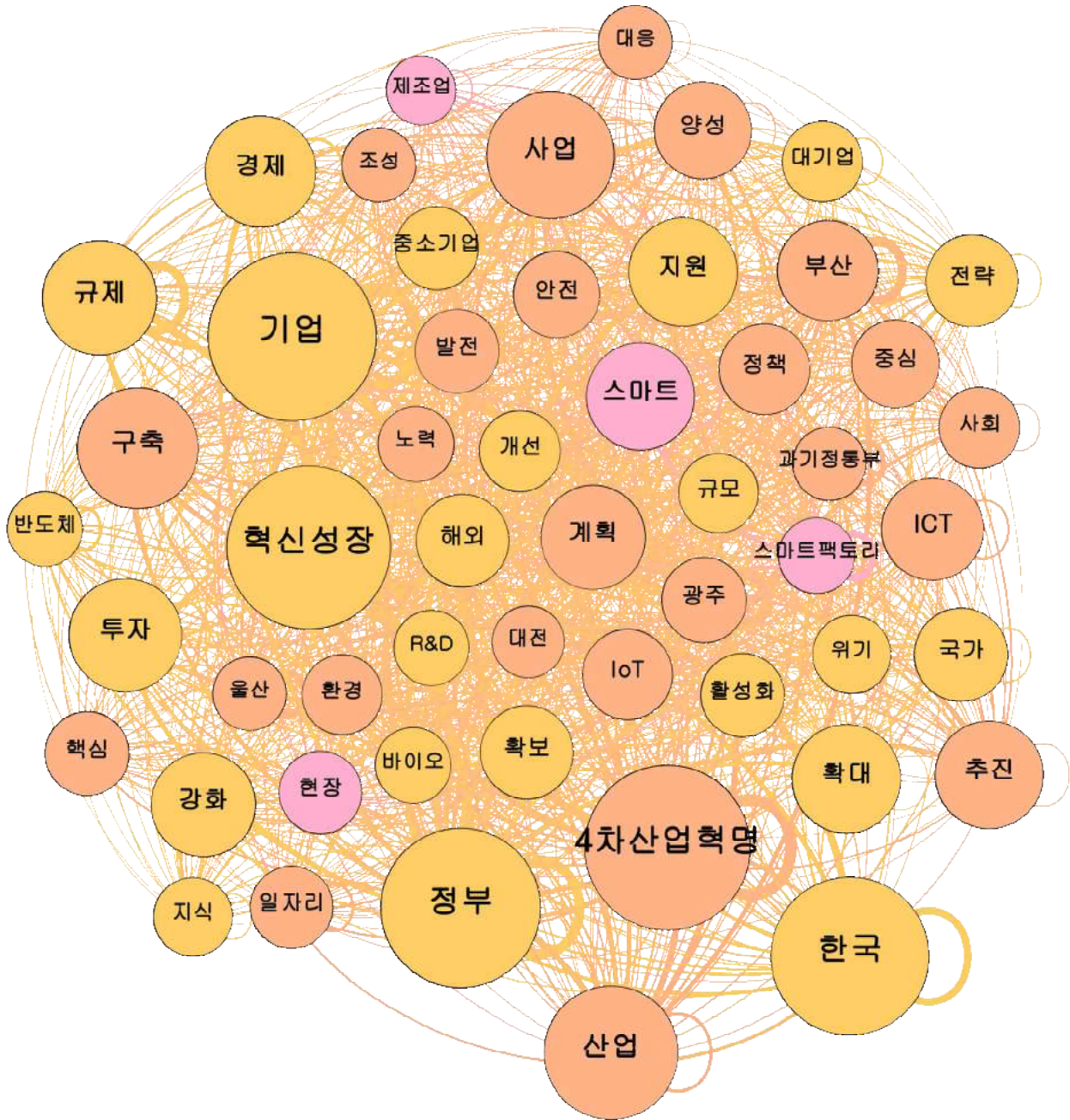
□ 1번째 군집: 4차산업혁명

그림 5. 4차산업혁명 군집 네트워크 그래프



- ‘기업(연결정도: 12,398)’ , ‘4차산업혁명(14,072)’ , ‘혁신성장(11,652)’ , ‘정부(12,705)’ , ‘한국(12,485)’ , ‘산업(8,036)’ , ‘사업(6,919)’ , ‘구축(6,665)’ , ‘규제(6,960)’ , ‘투자(5,778)’ 등의 순서로 높은 고유벡터중심성을 나타냄
- 해당 군집의 경우, 총 53개의 노드로 이루어진 군집의 내용을 한 번에 파악하기 어렵기 때문에 좀 더 구체적인 내용 분석을 위해 모듈성 기반 군집 알고리즘을 다시 한 번 활용하여 추가 군집분석을 수행하는 계층적 군집 분석을 수행하였음

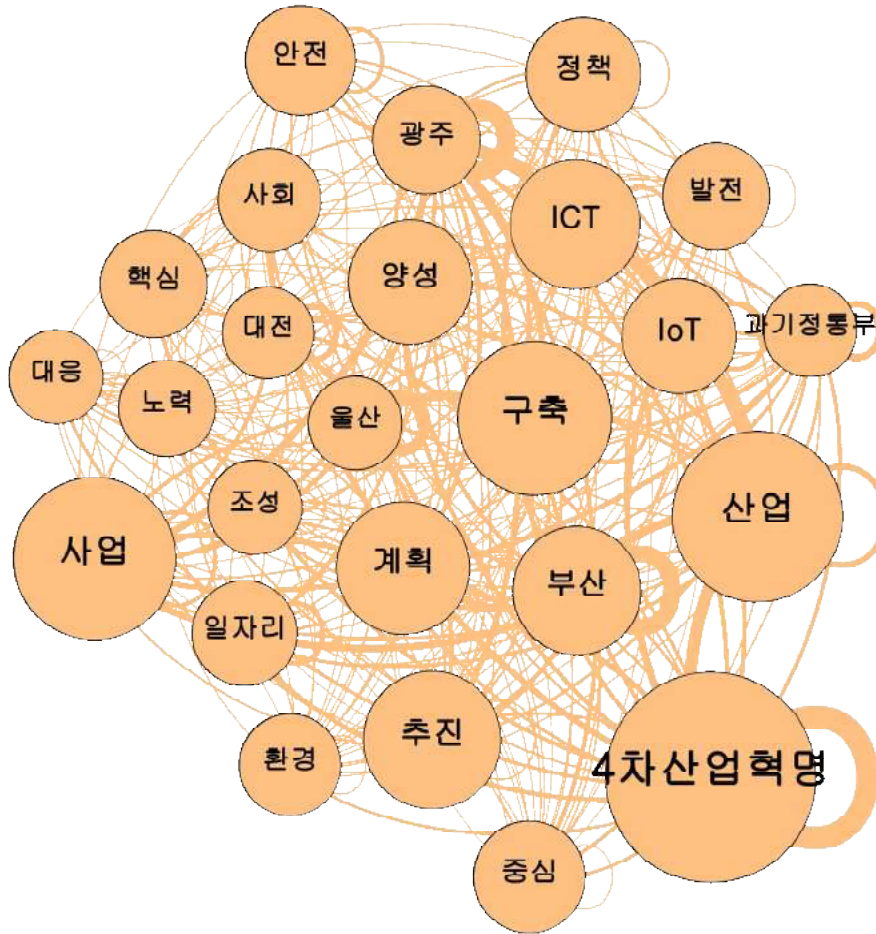
그림 6. 4차산업혁명 네트워크 계층 군집 분석 결과



- 계층 분석을 통해 총 3개의 군집으로 분류되었음. 군집은 위 그림처럼 노랑, 주황, 분홍의 3가지 색깔로 구분됨
- 4차산업혁명 군집의 하위군집 1~3에 대한 추가적인 분석은 다음과 같음

○ 4차산업혁명 하위군집 1

그림 7. 4차산업혁명 하위군집 1

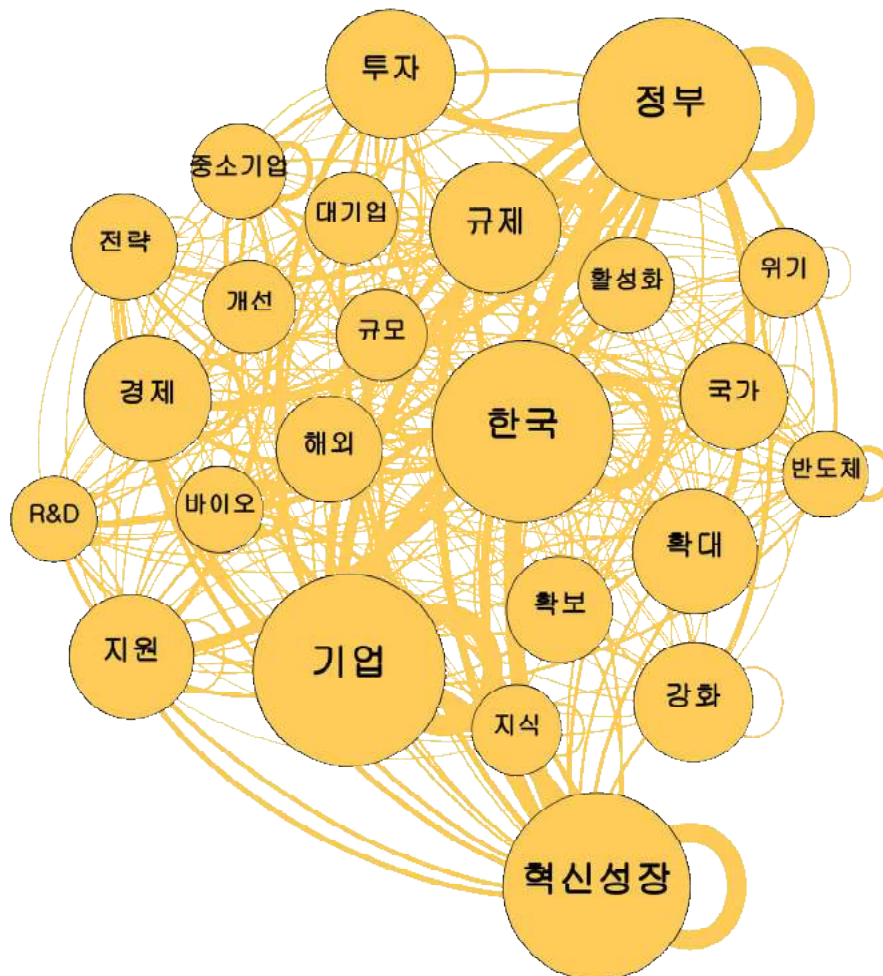


- 2020년에는 210억 개의 IoT 기기가 연결되어 각기 데이터를 전송할 것으로 예측('15년, 가트너)
- 이러한 빅데이터 산업 활성화를 위해 ‘과기정통부’ 를 필두로 각종 발전 정책을 펼치고 있음
- '19년 8월 과기정통부는 데이터 경제로의 전환을 선언하였고, 데이터 경제 활성화를 위해, 약 1,400억 원을 투입하여 빅데이터 플랫폼 및 센터 사업, 데이터 바우처 사업 등을 추진함
- 그러나 아직도 대전, 울산, 부산 등 일부 지자체에만 빅데이터 대응 인력이 운용되고 있는 한계가 존재함
 - * 빅데이터 전담조직을 갖춘 광역 지자체는 절반에도 못 미치고 있음 ('18년 기준, 한국 지역정보개발원)

- 또한, 개인정보 보호에 대한 우려는 기존 빅데이터 활성화에 주요 저해 요소가 되어 왔으며, 이에 대응하기 위해 익명 정보와 가명 정보를 구분하여 운영하는 안전한 빅데이터 이용 환경 구축에 많은 지원이 이루어지고 있음
- 결국, 빅데이터의 중요성에 대한 공감대는 형성되어 있으나, 지자체의 전담 인력 부재, 개인 정보 보호에 대한 우려는 몇 년간 빅데이터 활성화가 지지부진했던 주요 원인이라고 파악됨

○ 4차산업혁명 하위군집 2

그림 8. 4차산업혁명 하위군집 2

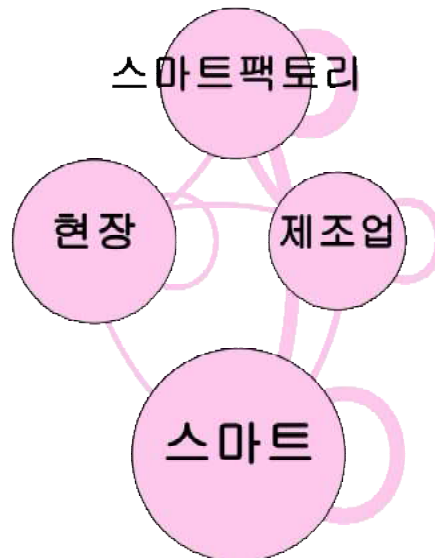


- 정부에서도 기업에서도 빅데이터는 혁신성장의 주요 동력이라고 생각하고 있음
- 해외의 경우, 유전정보 분석 등 바이오 분야를 중심으로 다양한 빅데이터 관련 R&D가 확대되고 있음
 - * 전 세계 생물정보학(Bioinformatics) 시장은 '25년 345.6 억불에 이를 것으로 전망

- 한국의 경우는 아직 많은 규제 속에 상업적인 분야의 빅데이터 활성화는 되고 있지 못한 것이 현실임
 - * 우리나라의 빅데이터 사용 및 활용능력 수준은 평가대상 63개 국가 중 56위('17년, 스위스 국제경영개발대학원(IMD) 국가경쟁력 평가)
- 반도체 분야에서도 상대적으로 데이터 저장에 목적인 메모리 반도체에만 편중되어, AI 기술을 활용한 빅데이터 분석에 많이 활용되는 시스템 반도체의 활성화 전략이 시급해 보임
 - * 한국은 전 세계 메모리 반도체 시장 점유율 1위 VS 비메모리 반도체 시장점유율 주요 선진국 중 최하위인 3%

○ 4차산업혁명 하위군집 3

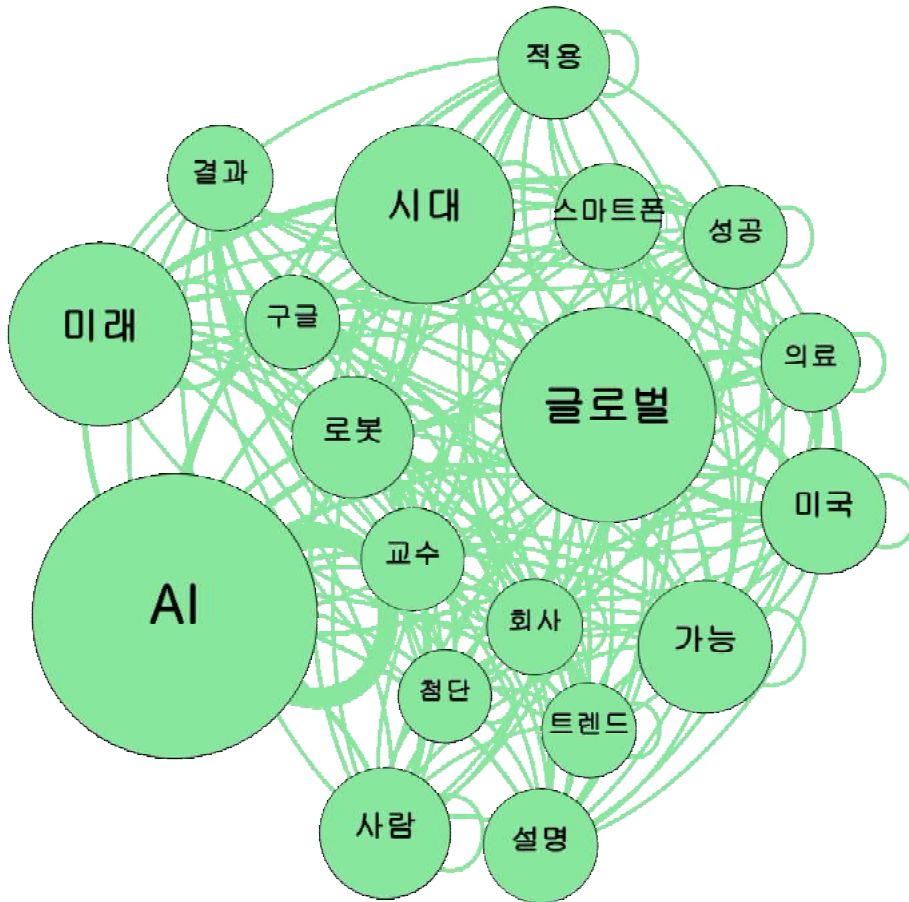
그림 9. 4차산업혁명 하위군집 3



- 제조 강국인 우리나라에서 관심을 가지고 있는 스마트팩토리 도입을 통한 제조업의 혁신 또한 빅데이터 분석이 핵심기술임
- 스마트팩토리는 공장 자동화를 넘어 제조 현장에서 발생하는 데이터를 취합하고 분석해 최적의 생산 환경을 만들어내는 지능형 공장으로 설비마다 데이터를 생산하여야 하고, 이를 취합 분석 할 수 있는 기기간 데이터의 연계 활용성이 핵심 성공 요인임
- 공장 자동화의 주요 성공 요인은 설비의 성능이었다면, 스마트팩토리의 성공은 빅데이터를 분석 할 수 있는 스마트팩토리 플랫폼이 좌우 할 것으로 예상됨

□ 2번째 군집: AI

그림 10. AI 군집 네트워크

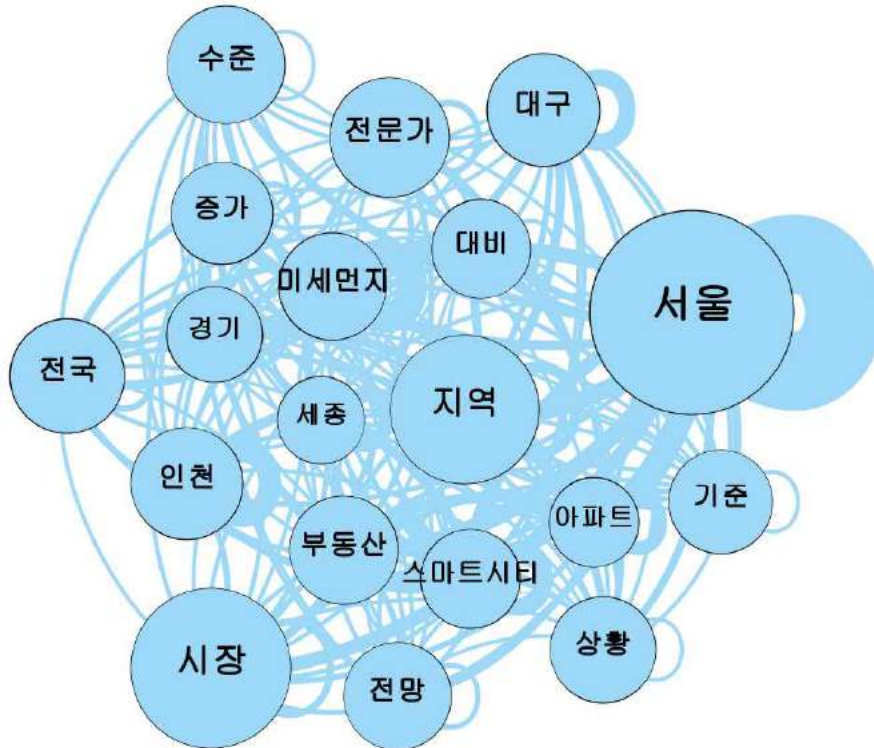


- ‘AI(연결정도: 23,033)’, ‘글로벌(12,500)’, ‘미래(9,465)’, ‘시대(9,158)’, ‘가능(4,818)’, ‘사람(5,114)’, ‘미국(4,567)’, ‘로봇(5,183)’, ‘설명(3,909)’, ‘적용(3,822)’ 등의 순서로 높은 고유벡터중심성이 나타났으며 총 19개의 노드로 구성됨
- AI 기술의 발전은 기존 추론 통계 방식 위주의 빅데이터 분석 방법을 탈피하는 주요 원인이 되었음
- 미국, 특히 구글을 중심으로 급속히 발전되고 있는 AI 기술은 의료 분야 등 더욱 복잡한 변수가 존재하는 분석에 혁신을 가져옴
- 현재는 사람이 AI 분석을 통한 결과를 이해할 수 있는 ‘설명 가능한 AI 분석 기술’ 이 첨단 트렌드로 자리매김하고 있음
- 이세돌과의 바둑대전으로 촉발된 AI에 대한 국민적 관심은 로봇과 스마트폰 등 폭넓은 분야에 적용되기 시작하면서, 지금은 2010년대에 관심이 집중되었던 빅데이터 분석보다 오히려 많은 관심을 받고 있음

- AI에 대한 학생들의 수요가 많아지면서 학교에서도 AI 관련 교수 채용이 급증하였고, 이에 따른 AI 분야 교수 부족 현상이 극심한 상황임
- 그러나 주의해야 할 것은 새로운 AI분석 기술을 개발하는 컴퓨터 공학 혹은 계산과학 쪽의 인원 수요는 사실 매우 제한적이며 초고급 인력이 필요함
- 오히려 적절한 AI 기술을 활용하여 분석이 가능한 빅데이터 분석 인력에 대한 수요가 훨씬 더 많기 때문에 해외의 경우는 통계, 경영, 머신러닝(AI) 등 융합적인 역량이 필요한 데이터 사이언티스트를 키우는데 더 집중하고 있는 추세임
- AI 기술 발전과 이에 따른 인력양성이 매우 시급하고 중요한 것은 맞지만, 이는 소수의 초고급 인재 양성에 집중해야 함
- 엔지니어와 분석가의 구분을 명확히 하고 이에 따라 각기 개별 특성을 고려한 인재양성 정책 수립이 필요함
- 특히 빅데이터 분석의 경우, 금융, 의료, 제조 등 분석 분야에 따라 도메인 지식이 필수적이기 때문에 각 분야별 빅데이터 분석 전문가 육성이 시급함
- 빅데이터 산업 활성화를 위해서는 분야별로 분석이 가능한 도메인 지식과 분석 역량을 모두 보유한 융합적인 인재 양성이 시급함

□ 3번째 군집: 지자체

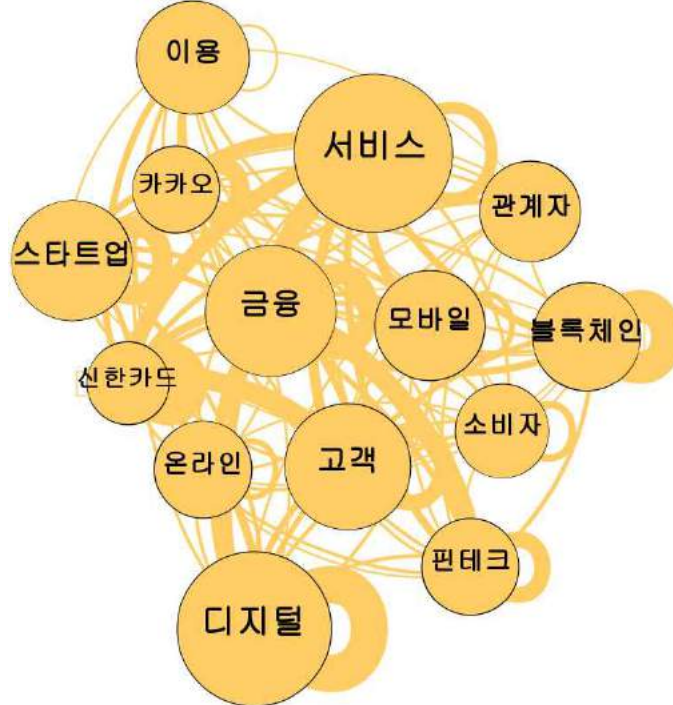
그림 11. 지자체 군집 네트워크



- ‘서울(연결정도: 13,064)’ , ‘시장(6,515)’ , ‘지역(6,159)’ , ‘전문가(4,489)’ , ‘수준(3,944)’ , ‘전국(4,298)’ , ‘대구(4,890)’ , ‘인천(4,914)’ , ‘부동산(4,372)’ , ‘전망(3,415)’ , ‘미세먼지(4,515)’ 등의 순서로 높은 고유벡터중심성이 나타났으며 총 19개의 노드로 구성됨
- 서울, 경기, 인천, 세종, 대구 등 일부 지자체를 중심으로 미세먼지 전망이나, 부동산 예측 등 다양한 분야에서 빅데이터 분석 기술을 활용하고 있음
- 특히 몇몇 지자체의 스마트시티 사업은 지자체 단위의 빅데이터 기술 확산에 매우 긍정적인 요소로 작용하고 있음
- 그러나 지자체 대부분은 분야별로 각기 다른 빅데이터 분석 전문가가 필요하다는 특성과 국내 전문가가 부족하다는 현실 때문에 전담 조직 자체가 없음
- 소수 전담조직이 있는 지자체도 전문인력으로 구성된 전담조직이 아닌 데이터 확보 및 인식개선 사업 등 지원업무를 수행하는 조직으로 구성되어 있음
- 효과적인 빅데이터 사업 활성화를 위해서는 중앙정부뿐만 아니라 지자체에도 전문인력으로 구성된 전담조직 구성이 필요함

□ 4번째 군집: 금융 빅데이터

그림 12. 금융 빅데이터 군집 네트워크



- ‘서비스(연결정도: 6,846)’ , ‘디지털(7,640)’ , ‘금융(5,779)’ , ‘고객(5,343)’ , ‘스타트업(4,998)’ 등의 순서로 높은 고유벡터중심성이 나타났으며 총 14개의 노드로 구성됨
- 네 번째 군집의 의미 분석 결과, 국내는 빅데이터와 관련한 금융서비스가 활성화되어 있음을 알 수 있음. 신한카드를 필두로 다양한 금융기관들이 빅데이터 활용 서비스를 제공하고 있음
- 소비패턴 데이터를 분석하여 적합한 카드를 추천해 주거나, 빅데이터를 활용에 신용등급을 분석하고 부정거래탐지시스템을 운영하는 등 특히 금융 분야에서의 빅데이터 활용이 빠르게 앞서 나아가고 있음
- 이는 세계적으로도 상당히 빠르게 발전하고 있는 상황으로 타 분야에서는 해외에 비해 상대적으로 빅데이터 활용도가 떨어지는 데 반해 매우 흥미로운 현상임
- 금융권의 빅데이터 활용 금융서비스와 더불어 카카오 등 비금융권의 핀테크, 블록체인 등 기술 기반의 다양한 스타트업들이 빅데이터 활용 금융서비스에 뛰어들어, 경쟁을 통한 관련 산업의 긍정적 발전에 기여하고 있음

4. 결론 및 시사점

- 빅데이터가 언급된 최근 1년간의 미디어 데이터 분석을 통해 산업현장의 주요 이슈는 다음과 같이 도출되었음
 - 정부, 기업 모두 빅데이터는 혁신성장의 주요 동력이라는 공감대는 이미 형성되어 있으나, 여전히 개인정보 및 데이터 활용 관련 규제, 전문인력 부족으로 빅데이터 활용 및 활성화에 한계가 존재해왔음
 - 최근 데이터3법 개정안이 국회 발의 약 14개월 만에 본회의에 통과('20.1.9)되면서 빅데이터 활용과 서비스 활성화 등 데이터 경제 발전에 대한 기대가 높아짐
 - 특히, 유전체, 바이오, 헬스케어, 반도체, 금융, 제조 분야 스마트팩토리와 스마트시티 등 분야에서 도메인 지식과 AI 기술을 활용한 빅데이터 분석 및 활용, 그리고 관련 융합인력 양성의 중요성이 높아지고 있음
 - AI의 경우, 인력수요와 더불어 인력을 육성하는 교수자의 수요가 급증하고 있으나, AI 원천기술, 알고리즘 개발과 같은 초고급 인력의 수요는 매우 제한적임
 - 오히려, AI 기술을 활용하여 빅데이터 분석, 빅데이터 엔지니어 등 도메인 분야의 활용을 목적으로 한 융합인재 양성이 효과적임
- 빅데이터 산업 활성화를 위해서는 빅데이터가 활용되는 분야별로 도메인 지식을 이해하고 분석이 가능한 융합적 인재양성이 중요함
 - 한국이 강점을 가지는 제조 분야 혁신을 위해서는 스마트팩토리 도입과 관련된 경험과 지식을 함양한 인재 양정으로 재도약의 발판을 마련해야 함
 - 정부와 지자체 주도에 기업, 시민이 참여하는 스마트시티 사업은 지자체별 특색을 고려하고 첨단 인프라와 다양한 기술을 활용하여 추진 중이나, 도시 특성을 고려한 문제 해결, 미래도시 계획, 첨단 기술의 활용 등 분야별 세부 전문성을 갖춘 빅데이터 분석이 가능한 전담 조직의 구성이 관건임
 - 금융 빅데이터의 경우, 데이터분석 기술과 정보기술을 바탕으로 한 비금융권 IT기업들의 참여로 빅데이터를 활용한 금융서비스 산업에 긍정적 발전이 이루어지고 있음
 - 관련하여, 핀테크 및 금융 빅데이터 인력의 융합 전문인력 양성이 활발하게 이루어진다면 관련 서비스 고도화와 산업의 발전을 통해 글로벌 경쟁력을 갖출 수 있을 것으로 기대됨