

Window of Statistics

통계의 창

2021
SUMMER
Vol.27

| ISSUE |

AI, 빅데이터 시대의 변화

경기도 판교 제로시티 자율주행 실증단지 운영 및 공공데이터 개방

AI시대, AI교육!

AI(인공지능) 산업통계 개발

| 통계광장 |

공공데이터 활용성을 선도하기 위한 통계청의 K-통계시스템 구축

문화관광분야에서의 빅데이터 활용

빅데이터 시대의 데이터 리터러시

| 통계탐방 |

뇌과학 인공지능의 융합을 추구하는 한양대학교 심리뇌과학과

| PEOPLE |

한국고용정보원 미래직업연구팀 연구위원 / 경영학 박사 김동규
인공지능 및 빅데이터 시대, 직업 세계는 어떻게 변화할까?



ISSUE



통계광장



통계집중탐구

통계탐방

004

경기도 판교제로시티 자율주행 실증단지 운영 및 공공데이터 개방
김형주 차세대융합기술연구원 경기도자율주행센터 선임연구원

012

AI시대, AI교육!
박찬 인천삼산초등학교 교사

020

AI(인공지능) 산업통계 개발
허정 한경대학교 법경영학부 교수
최혜리 소프트웨어산업정책연구소 연구원

028

공공데이터 활용성을 선도하기 위한 통계청의 K-통계시스템 구축
백승황 통계청 통계데이터기획과 사무관

040

문화관광분야에서의 빅데이터 활용
박근화 한국문화관광연구원 팀장

050

빅데이터 시대의 데이터 리터러시: 데이터를 다루는 습관
구자룡 벨류바인 대표

058

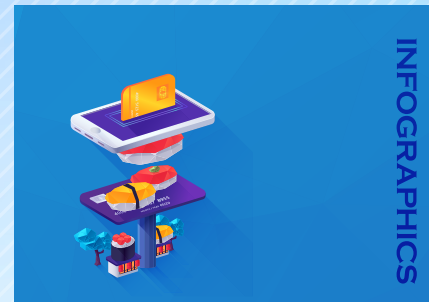
동형 암호화의 개념 및 현황
송용수 서울대학교 수리과학부 교수

064

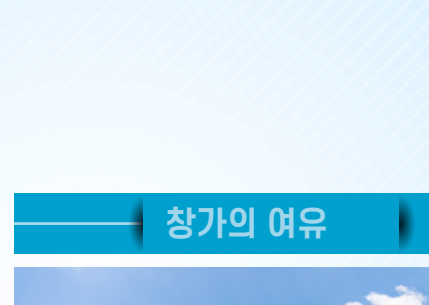
뇌과학 인공지능의 융합을 추구하는 한양대학교 심리뇌과학과
김성신 한양대학교 인텔리전스컴퓨팅학부 심리뇌과학과



교육



피플



창가의 여유



072

R에 도전하자... 따라가다 보면, 나도 R유저 ㉠
심송용 한림대학교 데이터과학스쿨 교수

084

해외 통계자료로 시장분석 보고서 만들기
이수동 (주)인포그래픽웍스 대표 겸(사)한국인포그래픽협회 대표

090

통계데이터대구센터 개소!
대구·경북지역 경제발전의 디지털 초석을 마련
김옥경 통계청 통계데이터기획과 사무관

094

인공지능 및 빅데이터 시대, 직업 세계는 어떻게 변화할까?
김동규 한국고용정보원 미래직업연구팀 연구위원 / 경영학 박사

100

**이것만 알면 나도 영상편집 전문가
태블릿PC로 영상편집을 해보자**
정영국 디자이너

106

100살까지 뇌를 지키는 8가지 방법
김여환 가정의학과전문의 / 의학박사

110

통계로 바라보는 세상이야기
신동헌 통계의 창 객원기자



발행일 | 2021년 5월 31일
발행인 | 서운주
발행처 | 통계교육원
기 획 | 김정란, 최희정

주 소 | 대전광역시 서구 한밭대로 713(월평동) 통계센터 통계교육원
전 화 | 042-366-6151
팩 스 | 042-366-6498
이메일 | ccn420@korea.kr

디자인 | Design MOTIFF
인쇄 | (주)중앙문화인쇄
ISSN 2005-1379
©2021. 통계교육원

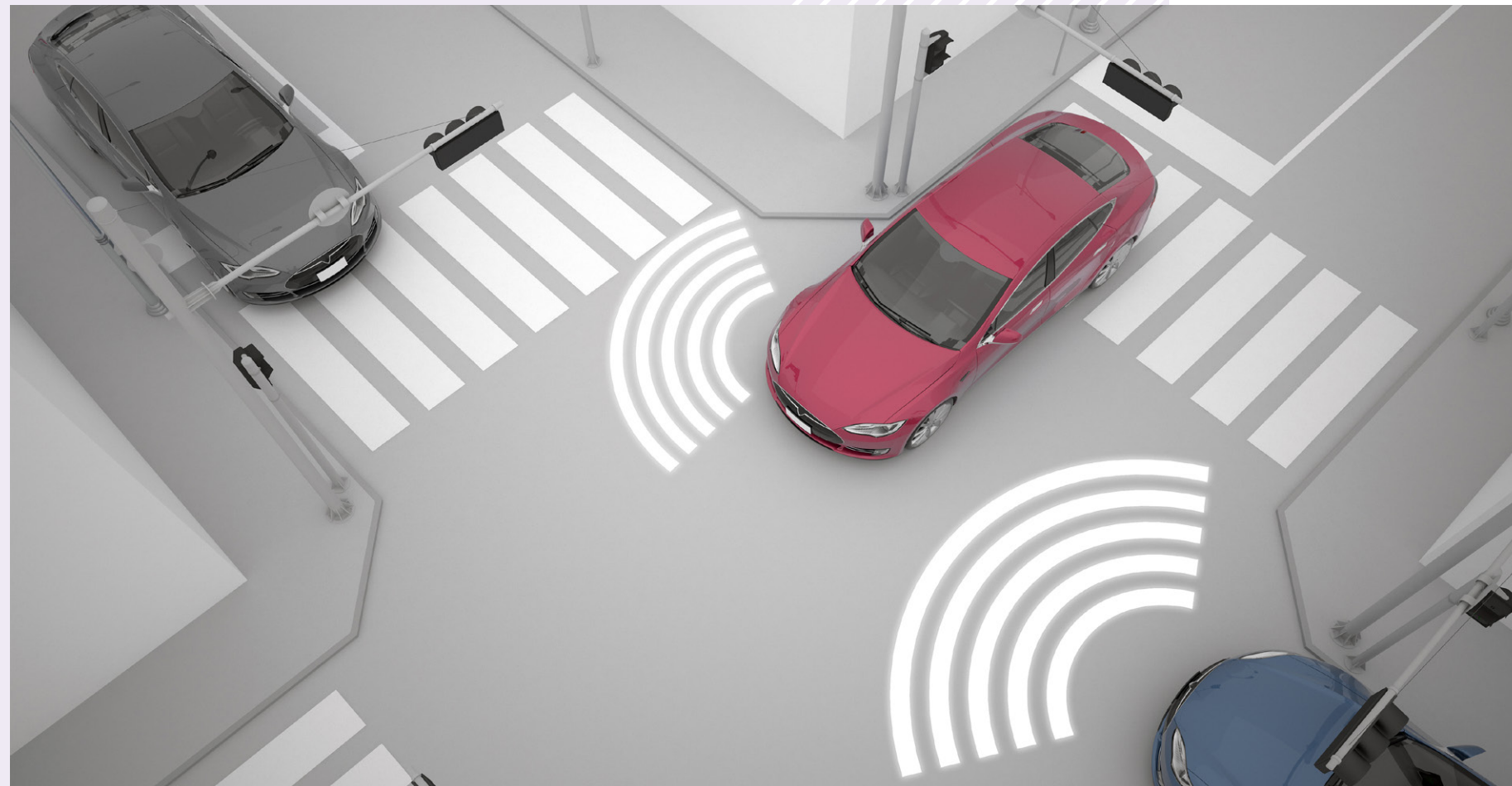
※ 『통계의 창』에 실린 내용은 필자 개인의 의견이므로 필자의 소속기관이나 본지의 공식적인 견해를 대변하는 것은 아닙니다.

ISSUE 01

김형주

차세대융합기술연구원 경기도자율주행센터 선임연구원 hyungjoo@snu.ac.kr

경기도 판교제로시티 자율주행 실증단지 운영 및 공공데이터 개방



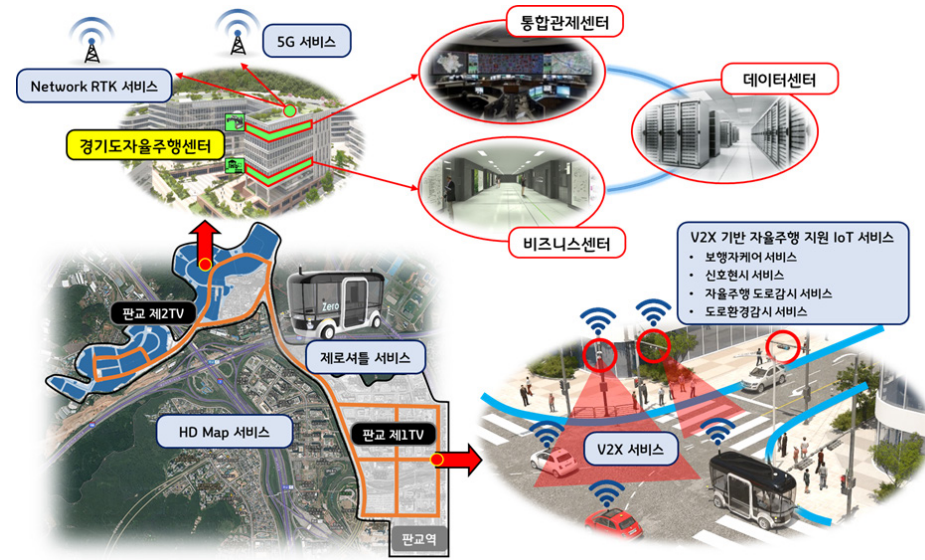
경기도 판교제로시티 자율주행 실증단지 개요

경기도 판교제로시티(PANGYO ZEROcity Autonomous Driving Pilot City)는 사고제로/비용제로/규제제로/탄소제로 등 건강한 자율주행 산업 생태계 조성을 목표로 구축된 자율주행 실증단지이다. 판교제로시티는 국내 최초 "실도로 기반 자율주행 실증 및 데이터 수집이 가능한 실증단지"로 일반차와 자율주행차가 공존하는 실제 환경의 자율주행 테스트베드이다. 판교 제1테크노밸리 및 제2테크노밸리 지역을 포함하는 구간으로 33만평의 공간적 규모 및 자율주행 실증노선 10.8km를 포함한다. 특히 판교제로시티는 경부고속도로, 수도권순환고속도로, 판교IC, 대왕판교IC, 판교역, 수서역 등을 포함하는 지역으로 향후 자율주행 모빌리티 서비스 제공을 위한 우수한 지리적 요건을 가지는 장점을 가진다. 또한 자율주행과 직·간접적인 관련이 있는 약 1,000개의 기업(IT 약 900개, CT 약 100개)이 입주해 있으며, 이들과의 자율주행 기술 협력과 새로운 서비스 수요를 도모하고 있다.



추진목표

- "사고" ZERO: 실세계 환경 자율주행 융합기술 테스트베드 구축
- "비용" ZERO: 자율주행 산업 생태계 지원 오픈플랫폼 구축
- "규제" ZERO: 신규 자율주행 서비스 실증 지원 민·관 협의체 구성
- "탄소" ZERO: 친환경 도심 자율주행 모빌리티 서비스 실증



판교제로시티 자율주행 실증단지 개요

판교제로시티는 국내 최초 실도로 기반 자율주행 실증 대상지로 자율협력주행을 위한 통합관제센터 및 IoT 서비스 인프라가 구축되어 있으며, 자율주행 관제플랫폼 연계를 통한 안전관리를 진행 중이다. 판교제로시티 내 141대의 CCTV 영상데이터, 지능형 영상분석 솔루션, 자율주행 IoT 인프라 서비스(보행자케어 서비스, 자율주행 도로 감시 서비스, 도로 환경 감시 서비스, 신호 현시 서비스)를 활용하여 자율주행차와 보행자에게 안전한 자율주행 환경을 제공 중이다.



판교제로시티 도로 IoT 인프라 서비스

공공 모빌리티 서비스를 위한 제로셔틀

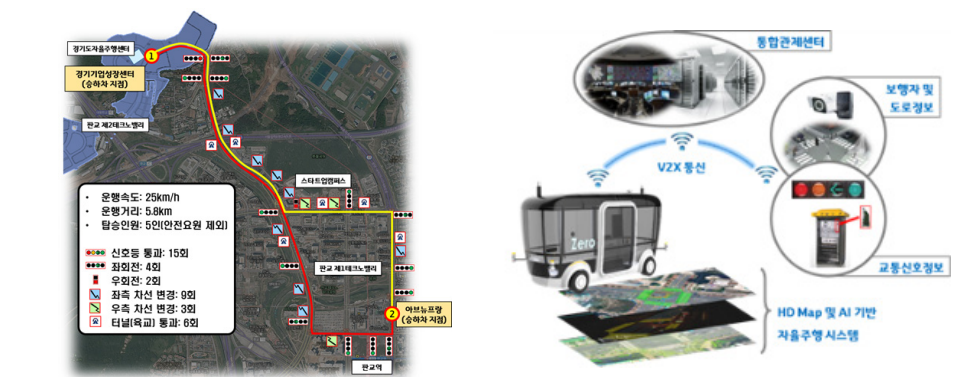
공공 모빌리티 서비스를 위한 자율주행차 제로셔틀은 11인승(입석 6인) 전기차 플랫폼으로 자차 다중센서, HD Map, 그리고 통합관제센터 V2X 연계 자율주행 등을 포함한다. 국토부 자기인증 및 임시운행허가 등이 완료된 차량으로 자율주행 Level 4(제한속도 25km/h) 기준 모빌리티 서비스 실증을 진행 중에 있다. 판교제로시티 내 5.8km의 일반도로를 일 4회(오전 2회/오후 2회) 운행 중에 있으며 2020년 10월 기준 총 운행시간 9,493분, 총 운행거리 1,424km, 총 탑승인원 685명으로 자율주행 생태계 조성 및 민간의 사회적 수용성 증진에 선도적 역할을 수행하고 있다.



공공 모빌리티 서비스를 위한 제로셔틀



판교제로시티 자율주행 관제플랫폼



판교제로시티 제로셔틀 모빌리티 서비스구간

제로셔틀 서비스 운영 시간

회차	타승시간	운행노선
1회차	타승시간	오전 10시 00분
	운행노선	①출발(경기기업성장센터) → ②도착(아브뉴프랑)
	타승위치(출발시간)	경기기업성장센터 앞 임시정거장(오전 10시 00분)
2회차	타승시간	오전 10시 30분
	운행노선	②출발(아브뉴프랑) → ①도착(경기기업성장센터)
	타승위치(출발시간)	아브뉴프랑 앞 임시정거장(오전 10시 30분)
3회차	타승시간	오전 11시 00분
	운행노선	①출발(경기기업성장센터) → ②도착(아브뉴프랑)
	타승위치(출발시간)	경기기업성장센터 앞 임시정거장(오전 11시 00분)
4회차	타승시간	오전 11시 30분
	운행노선	②출발(아브뉴프랑) → ①도착(경기기업성장센터)
	타승위치(출발시간)	아브뉴프랑 앞 임시정거장(오전 11시 30분)
5회차	타승시간	오전 11시 50분
	운행노선	경기기업성장센터 앞 임시정거장(오전 11시 50분)
	타승위치(출발시간)	아브뉴프랑 앞 임시정거장(오전 11시 20분)
6회차	타승시간	오전 11시 20분
	운행노선	①출발(경기기업성장센터) → ②도착(아브뉴프랑)
	타승위치(출발시간)	경기기업성장센터 앞 임시정거장(오전 11시 00분)
7회차	타승시간	오전 11시 50분
	운행노선	②출발(아브뉴프랑) → ①도착(경기기업성장센터)
	타승위치(출발시간)	아브뉴프랑 앞 임시정거장(오전 11시 30분)
8회차	타승시간	오전 11시 30분
	운행노선	경기기업성장센터 앞 임시정거장(오전 11시 50분)
	타승위치(출발시간)	아브뉴프랑 앞 임시정거장(오전 11시 20분)



공공 목적의 자율주행차 제로셔틀 운영은 향후 교통소외지역 해소, 수송분담률 제고, 교통약자의 이동성 증진, 사회문제 해결을 위한 모빌리티 대안 탐색 등에 수범사례로 발전이 가능하다. 특히 자율주행기술과 플랫폼 서비스를 기반으로 거점 연계, 배차의 불균형 해소, 대중교통 사각지대 해소 등의 지역균형 효과를 기대할 수 있으며, 맞춤형 모빌리티, 교통약자 우선, 친환경 등의 공공 모빌리티로서의 가능성을 탐색할 수 있다.



판교제로시티 제로셔틀 활용 모빌리티 서비스

판교제로시티 활용 지속가능한 자율주행 산업 생태계 조성

경기도는 판교제로시티 운영을 통한 “지속가능한 자율주행 산업 생태계 조성”을 위해서 실증테스트 지원/데이터수집 및 서비스/협력 및 글로벌 도약지원/협력 거버넌스 강화/자율주행 공공 모빌리티 개발 및 서비스 실증 등을 진행 중에 있다(하단 그림 참조). 판교제로시티에는 SOS랩, 비트센싱, 웨이티즈, 에이모, 서울로보틱스, 모빌테크, 컨텍모빌리티, 모라이 등 자율주행 스타트업이 입주하여 실증테스트를 진행 중이며, 자율주행 R&D 및 사업수행을 위한 공동연구를 수행하고 있다. 판교제로시티 자율주행 인프라 제공은 자율주행 기술을 개발하는 신생 기업들의 혁신 성장에 기여하고 있으며, 자율주행차 및 인프라에서 수집되는 데이터를 빅데이터 서비스 플랫폼을 통해서 제공하고 있다. 자율주행 스타트업의 CES(Consumer Electronics Show) 참가 등 글로벌 도약 지원, 협의체 구성을 통한 협력 거버넌스를 강화하고 있다. 무엇보다 경기도가 자체 개발한 자율주행차 제로셔틀의 일반인 탑승 실증을 안전하게 진행함으로써 자율주행기술에 대한 국민들의 사회적 수용성을 높이고 있으며, 향후 제로 모빌리티 서비스 기반 지속가능한 자율주행 산업 생태계 조성에 힘쓰고 있다.



판교제로시티 활용 지속가능한 자율주행 산업 생태계 조성

판교제로시티 활용 자율주행 공공데이터 수집 및 개방

경기도는 판교제로시티 활용 자율주행 공공데이터 수집 및 개방을 목적으로 국가중점데이터 개방사업(차세대융합기술연구원 주관, 한국지능정보사회진흥원 전문기관)을 진행 중이다. 판교제로시티 자율주행 공공데이터는 자율주행 통합관제 데이터 및 제로셔틀(자율주행차) 데이터로 구분된다. 자율주행 통합관제 데이터는 총 8종 이상으로 실시간 V2X 데이터(테스트자율주행차), 실시간 IoT 데이터(횡단보도 보행자 감지, 신호현시, 도로노면), 실시간CCTV데이터, 통합데이터셋(시간별, 날씨별, 상황별 등 가공한 데이터셋) 등을 포함한다. 제로셔틀(자율주행차) 데이터는 총 13종 이상으로 차량 센서 데이터(카메라, 라이다, 레이더, GPS/INS, V2X메시지(교통신호등)), 운행로깅 데이터(차량상태, 객체/위치 인식, 주행판단, 차량제어정보, 고정밀지도), 가공데이터셋(시간별, 날씨별, 상황별 등 가공한 데이터셋) 등을 포함한다.



판교제로시티 활용 자율주행 공공데이터 수집 및 개방 추진내용

판교제로시티 자율주행 실증단지에서 스마트시티로의 발전

경기도 판교제로시티 자율주행 실증단지는 향후 스마트시티로 발전하기 위해 판교제로시티 내외부 환경분석, 정보화 현황 분석, 목표모델 및 실행계획 수립 등을 진행 중이며, 스마트시티 발전을 위한 Smart-서비스 설계, 현장 시설물 설계, 정보통신망 설계, 통합 운영센터 설계 등을 진행 중이다. 판교제로시티 자율주행 실증단지 사업은 스마트시티로 발전하기 위한 밑거름이 될 것으로 판단되며, 스마트시티 사업과의 시너지 효과가 기대된다. 또한 판교제로시티 자율주행 실증단지는 스마트시티 7대 혁신요소 중 공유 모빌리티 및 자율주행서비스를 포함하는 것으로 향후 판교테크노밸리를 중심으로 자율주행 모빌리티 수범사례로 활용되어질 것으로 판단된다.



시사점

경기도 판교제로시티 자율주행 실증단지 운영을 통한 시사점은 다음과 같다. 첫째, 4차산업혁명 모빌리티 산업 생태계 변화에 따른 중앙-지방 정부 간 역할 분담 논의가 필요하다. 4차산업혁명 시대에 맞는 지방자치단체의 역할 변화와 이에 따른 지방자치단체의 독립적인 플랫폼 및 서비스가 중시사업으로 확대될 수 있는 방안 수립이 필수적이라 판단된다. 기존의 중앙정부 중심에서 벗어난 지방정부의 권한 확대, 수평적/유기적 행정조직의 연결을 통해서 4차산업혁명 시대에 맞는 중앙-지방 정부 간의 역할 분담이 필수적이다. 이에 경기도 판교제로시티가 자율주행 산업 생태계 조성의 중앙-지방 정부 간 역할 분담 수범사례로 활용될 것으로 판단된다.

둘째, 자율주행 관련 스타트업 및 민간기업과의 협력 방안에 대한 노력이 필요하다. 자율주행 관련 스타트업의 경우 법적규제 및 비용적 한계 등이 높은 진입장벽으로 인해 자율주행 기술개발에 어려움이 존재한다. 또한 자율주행 관련 산업 발전을 위해서는 민간기업과 정부기관의 협력적 거버넌스 강화가 필수적이다. 이에 경기도 판교제로시티가 실도로 기반 자율주행 실증테스트 지원, 판교제로시티 입주 스타트업 협력, 자율주행 리더기업 및 중소기업 간 네트워킹 및 기술공유 지원 등을 수행함으로써 수범사례로 활용될 것으로 판단된다.

셋째, 자율주행 관련 산업발전의 마중물 역할을 위한 공공데이터 개방이 필수적이다. 공공데이터는 정부 및 공공기관이 보유한 데이터를 민간에서 활용할 수 있도록 개방하는 것으로 향후 자율주행 산업발전을 위해서는 공공데이터 개방이 필수적으로 요구된다. 이에 경기도 판교제로시티는 자율주행 통합관제 및 자율주행차 데이터 수집을 바탕으로 국가중점데이터사업과 연계를 통한 자율주행 공공데이터 개방, 나아가 자율주행 빅데이터 표준 마련에 기여할 것으로 판단된다.



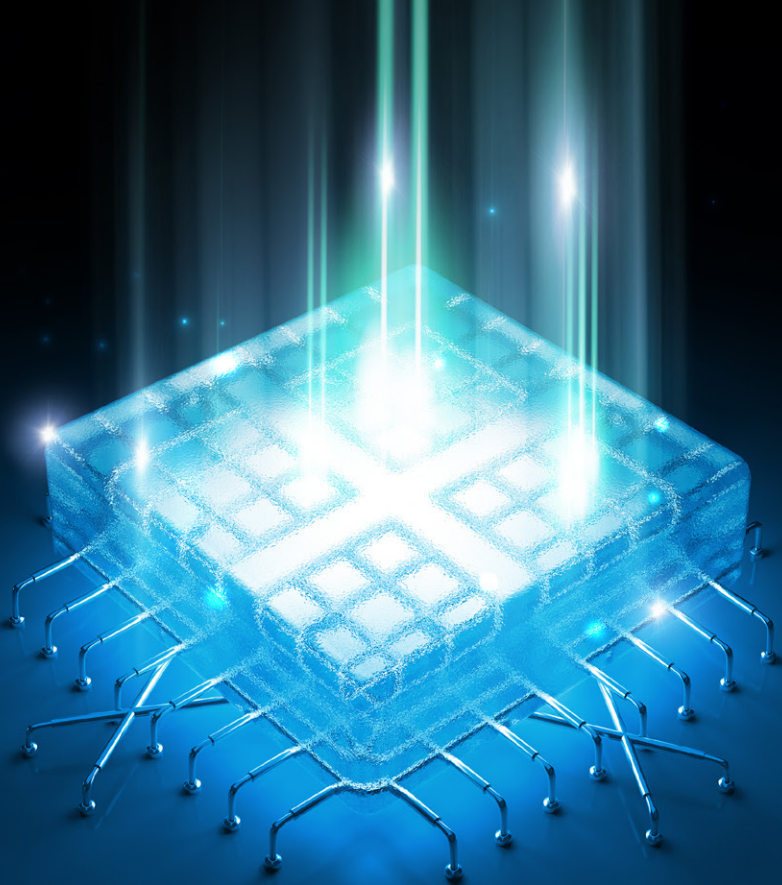
ISSUE 02



박찬
인천삼산초등학교 교사

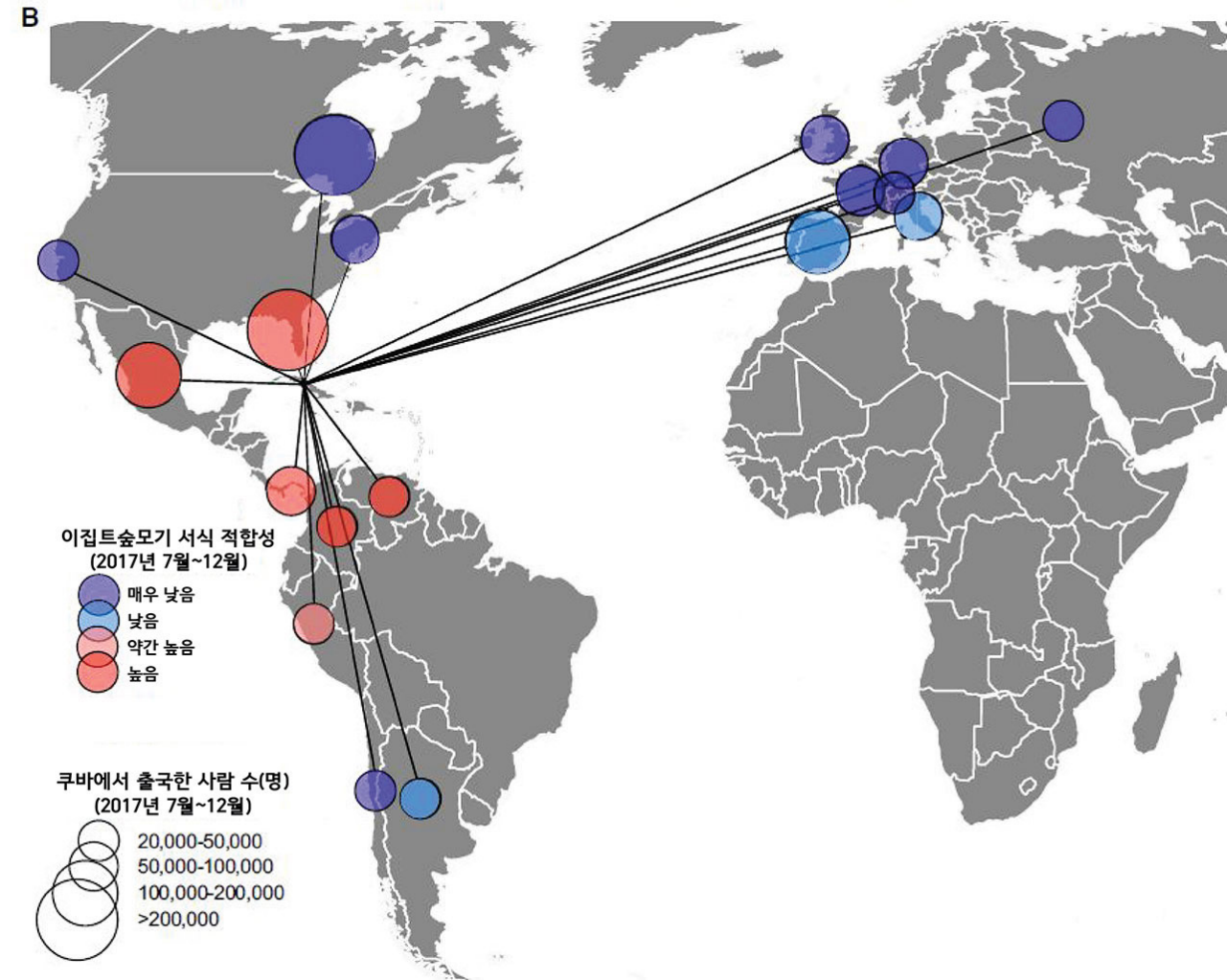
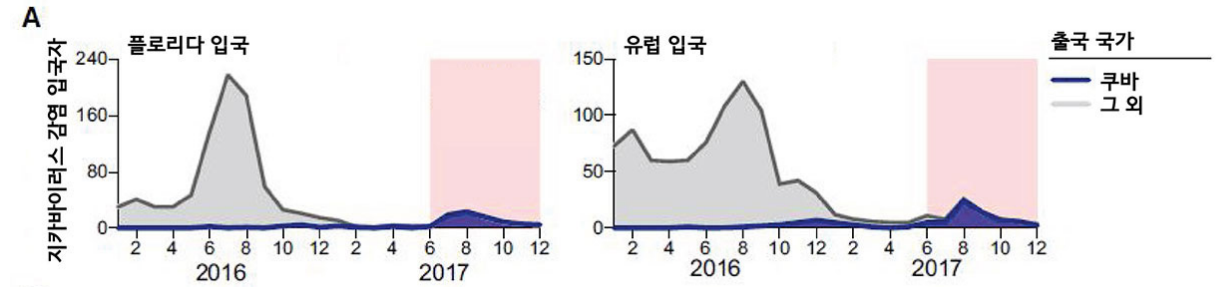
AI 시대

AI 교육 !!!!!



인공지능시대의 도래

2020년 세계보건기구(WHO)는 전 세계를 전염병의 공포로 몰고 간 코로나바이러스감염증-19(COVID-19)에 대해 세계적 대유행을 뜻하는 '팬데믹'을 선언하였다. 하지만 놀랍게도 신종코로나바이러스로 인한 코로나바이러스감염증-19(COVID-19)의 확산을 처음으로 경고한 곳은 세계보건기구(WHO)가 아니라 캐나다의 인공지능인 블루닷(BlueDot)이었다. 인공지능 블루닷이 보건 전문가들보다 먼저 코로나바이러스감염증-19(COVID-19)의 위험성을 인식하고 경고한 것이다.



블루닷은 항공으로 미국 플로리다와 유럽에 입국한 지카바이러스 감염자의 출발지를 분석
출처: N.D.Grubbaugh et al.(2019). Travel Surveillance and Genomics Uncover a Hidden Zika Outbreak during the Waning Epidemic.

최근 인공지능 스피커를 가지고 있는 가정들도 많고 인공지능 챗봇으로 상담받는 일도 많아졌다. 스마트폰으로 사진을 찍을 때 인공지능이 사람의 얼굴이나 신체를 인식하여 스티커를 붙이거나 얼굴 모양을 바꾸어서 사진을 찍을 수 있도록 돕기도 한다. 또한 쇼핑몰에서 나의 취향을 알아서 새로운 상품을 추천하기도 한다. 이렇듯 우리 생활에도 이미 인공지능이 깊숙이 들어와 있다.



인공지능의 역사

인공지능(AI, Artificial Intelligence)은 인간의 학습능력, 추론능력, 지각능력과 같은 인지적 능력을 인공적으로 구현한 컴퓨터 시스템이다. 즉, 기계가 인간처럼 생각하고 학습하여 판단할 수 있는 능력을 갖춘 것을 의미한다.

영국의 수학자 앨런 튜링은 2차세계대전 당시 독일군 암호를 해독해 영국을 구한 암호학자였다. 그는 “기계도 인간처럼 생각할 수 있을까?”라는 상상을 하게 되었다. 튜링은 1950년에 철학 저널 Mind 에 발표한 「Computing Machinery and Intelligence」에서, 기계가 지능적이라고 간주할 수 있는 조건을 언급했다. “기계가 생각할 수 있는가?, 기계가 인간과 비슷하게 대화할 수 있는가?”라는 질문에 대해, “컴퓨터로부터의 반응을 인간과 구별할 수 없다면 컴퓨터는 생각(thinking)할 수 있는 것”이라고 주장하면서 기계의 지능을 테스트하는 튜링테스트를 제안하였다. 이것이 인공지능의 시발점이었다. 1956년에는 미국의 컴퓨터 과학자 존 매카시(John Macathy)가 다트머스 대학에서 10명의 과학자가 모여 두 달 동안 인공지능을 연구하고자 제안하면서 인공지능이라는 용어를 처음으로 사용하게 되었다. 지능을 가진 기계를 일컫는 역사상 첫 번째 표현으로 '인공지능'이라는 말을 사용한 것이다. 이후 인공지능은 두 번의 붐과 같은 부흥의 시대와 두 번의 겨울과 같은 한계와 실망의 시기를 보냈다. 지금은 인공지능이 딥러닝과 빅데이터, 클라우드 등의 컴퓨팅 파워로 세 번째 붐의 전성기를 맞이하고 있다. 2016년에는 구글의 알파고와 우리나라 프로 바둑기사였던 이세돌과의 바둑대결에서 알파고의 승리로 인공지능에 더욱 큰 관심을 불러일으키게 되었다.



인공지능의 활용

인공지능은 인간의 여가, 산업, 의료, 예술, 금융, 복지 등의 분야에 활용되어 인간 이상의 능력을 발휘하며 많은 문제를 해결하고 있다. 예를 들어 암을 진단하는데 활용하는 인공지능 왓슨, 공장의 생산에 투입되어 각종 센서로 데이터를 받아들여 스스로 판단하고 행동하면서 단순하고 반복적인 업무들과 힘들고 위험한 일을 맡아 처리하는 인공지능 로봇, 그림을 그리고 작곡하며 악기를 연주하는 인공지능들이 이미 상용화되어 사용되고 있다. 또한 스마트폰 비서와 인공지능 스피커가 생활의 편리함을 증대시켜 줄 뿐만 아니라 노인의 말동무가 되어주고 아이들의 선생님이 되어 주기도 한다. 넷플릭스나 페이스북에서 영화나 친구를 추천해주는 추천 시스템, 인터넷에서 자주 볼 수 있는 챗봇 등 우리 생활에 가깝게 활용되는 인공지능들도 있다.

미래기술로 기대되는 자율주행자동차도 인공지능의 집약체이다. 스마트폰이나 인공지능 스피커가 사용자의 음성을 인식해 집 안의 모든 사물인터넷(IoT) 기기를 연결하고 사용자의 요구사항에 따라 자동으로 작동하거나 원격으로 조종하는 스마트홈도 인공지능으로 가능할 수 있다.

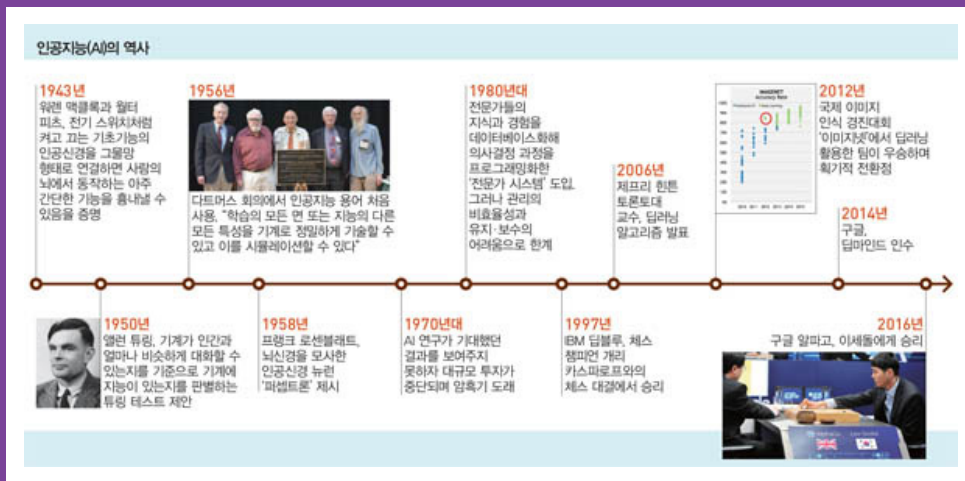
지금 우리의 아이들이 성장하여 직업을 가질 때가 되면 어떤 일을 하든 인공지능이 적용되지 않은 것이 없을 것이다. 우리 아이들에게 인공지능에 대한 개념을 알고 활용할 수 있는 능력이 없다면 미래 사회에서도 태될 가능성이 커질 것이다.



시대적 요구로서 등장한 인공지능교육

손정의 소프트뱅크 회장은 인공지능이 인류 역사상 가장 큰 혁명을 초래할 것이라고 말하며 이에 우리나라가 해야 할 것은 “첫째도 인공지능, 둘째도 인공지능, 셋째도 인공지능”이라고 하면서 전폭적 인공지능 육성 정책의 필요성을 강조하였다.

인공지능 교육에 대한 세계 동향을 살펴보면 다음과 같다. 인공지능 분야에서 세계 최고 경쟁력을 보유하고 있는 미국은 2016년 10월 국가과학기술협의회에서 인공지능 국가 연구 개발 전략을 발표하고, 인공지능의 기술적 수요 확인, 공공 정책 수립 및 장기적 연구 개발 투자 효과 극대화를 위한 국가 차원의 프레임워크를 제안하였다. 유럽연합(EU)의 국가들도 초·중등학교에서 인공지능 교육의 필요성을 인지하고, 학교급과 나이에 상관없이 모든 시민이 인공지능 문해력을 습득할 수 있도록 노력하고 있다. 중국에서는 국가 주도로 인공지능 교과서를 개발하고 2019년부터 초·중등



출처 : 주간경향(<http://weekly.khan.co.kr/>)



출처 : 청와대(<https://www1.president.go.kr/articles/6786>)

교육에 도입하고 있으며 일본, 인도 등에서도 미래인재 양성 차원에서 초·중등 교육에 인공지능 교육을 준비하고 있다.

우리나라도 시대적 요구에 부응하고자 2019년 12월에 인공지능 국가전략을 발표하였다. 'IT 강국을 넘어 AI 강국으로'라는 비전을 세우고 세계를 선도하는 AI 생태계를 구축하며 인공지능을 가장 잘 활용하는 나라, 사람 중심의 인공지능 구현이라는 핵심전략을 세웠다.



인공지능 국가전략을 통한 미래상

인공지능이 국가의 주요 의제가 됨에 따라 교육 분야에서도 인공지능 교육에 대한 요구가 커지고 있다. 이에 정부는 2020년 11월에 인공지능 시대 교육정책 방향을 발표하면서 대한민국의 미래 교육이 나아가야 할 길을 제시하였다. '인간다움과 미래다움이 공존하는 교육 패러다임 실현'을 비전으로 제시하였으며 3대 정책 방향을 내세웠다. 첫째, 미래인재 양성의 방향은 「감성적 창조 인재」로 설정하였다. 둘째, 학습환경은 「초개인화 학습환경」으로 변화할 것을 예고했다. 셋째, 정책 형성과정은 「따뜻한 지능화」를 추구할 것이라고 밝혔다.

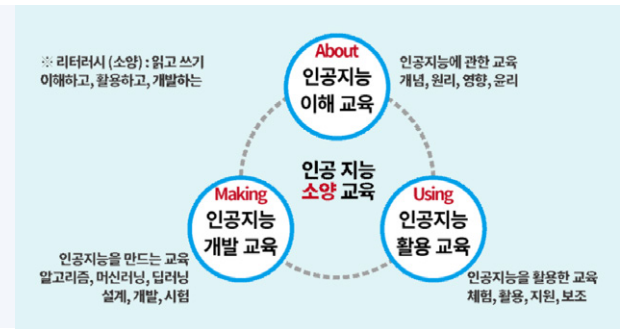


인공지능교육의 방향

인공지능교육이란 인공지능의 혜택을 누리기 위해 필요한 지식과 기능을 배우고, 인공지능과 함께 살아가기 위해 필요한 가치와 삶의 방식을 배우는 교육을 의미한다. 초·중등 인공지능 교육에서는, 인공지능 리터러시(AI literacy)를 갖추고 현재의 컴퓨팅 사고력과 소프트웨어 교육 역량을 기반으로 인공지능의 기능을 적용하여 창의적인 산출물을 만들어 낼 수 있는 역량을 주요 목표로 한다.

인공지능교육의 유형은 크게 인공지능 이해교육, 인공지능 활용교육, 인공지능 개발교육으로 나눌 수 있다. 인공지능 이해교육은 인공지능 용어, 지식, 개념, 원리, 법칙, 알고리즘 등에 대해 학습하여 인공지능의 원리와 인공지능 자체를 이해하는 학습이다. 이론과 실습을 통해 인공지능의 지식과 기능을 갖추는 데 중점을 둔다. 인공지능 활용교육은 자신과 생활에서 주어진 문제를 해결하기 위해 인공지능을 활용하는 내

용과 방법을 다룬다. 인공지능의 작동원리에 대해 잘 몰라도 인공지능 툴을 적용하여 학습에 도움을 받는 것에 관심을 둔다. 이론적인 부분보다는 기능적이고 실용적인 내용을 중점적으로 다룬다. 이미 개발된 다양한 인공지능 기술을 활용하여 교과의 문제를 해결하면서 교과에 흥미를 느끼고 학습자의 역량을 강화한다. 인공지능 개발교육은 인공지능을 만드는 교육이다. 인공지능의 알고리즘, 머신러닝 및 딥러닝 등을 적용한 인공지능을 만들 수 있도록 설계, 개발, 시험해보는 분야라 할 수 있다.



인공지능 교육의 요소(출처 : 서울시 교육청)



인공지능교육의 실제

학교 현장에서 학생들을 지도하며 느낀점은 다양한 인공지능교육 유형이 학생들의 단계에 맞추어 모두 필요하다는 것이다. 그러나 현재 학교의 학생들과 교사들의 역량들을 고려해볼 때 인공지능 자체에 대한 이해와 깊이 있는 개발교육에 앞서, 이미 개발되어 있고 쉽게 사용할 수 있는 인공지능을 교과와 학습에 활용하는 인공지능 활용 교육이 접근하기 쉽다.

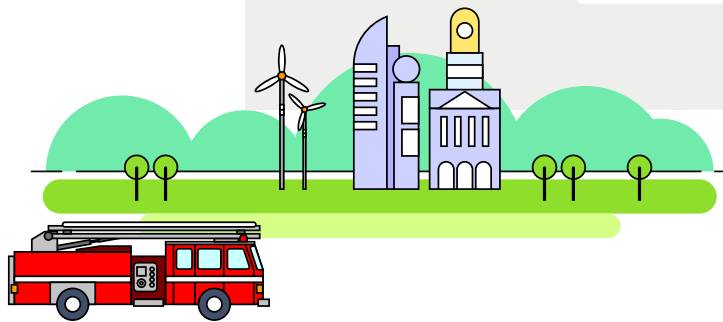
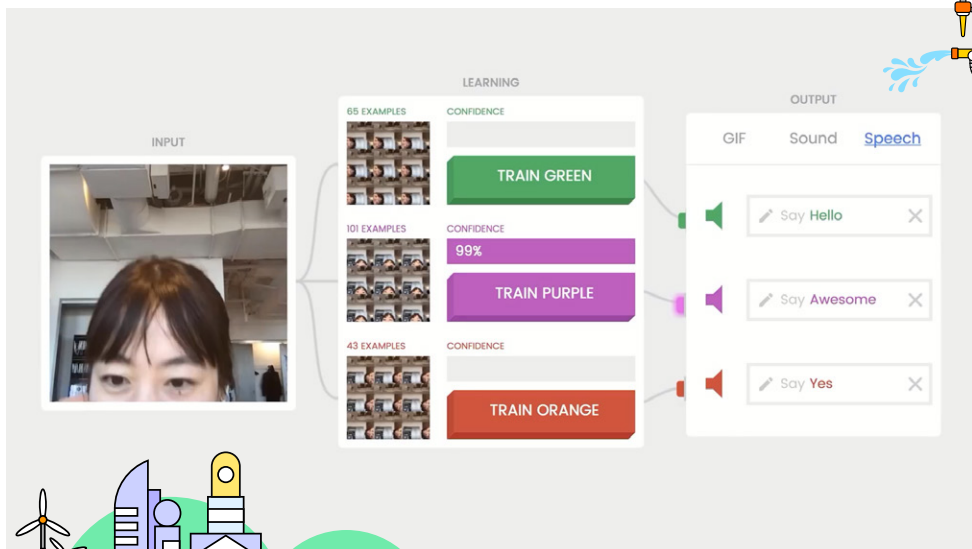
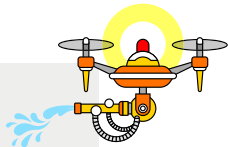
예를 들어 구글 오토드로우(AutoDraw)라는 인공지능 도구는 아이들의 낙서와 같은 그림을 인공지능이 인지하고 추측하여 추천 그림을 제안하는 도구이다. 이런 도구를 이용하면 학생들은 그리기 능력에 제한받지 않고 생각을 자유롭게 표현할 수 있으며 이러한 그림들을 자기소개하기, 마을지도 그리기, 놀이터 안전수칙 만들기, 환경 포스터 제작하기 등 이미지를 활용하는 학습에 다양하게 사용할 수 있다.



네컷만화나 잇셀프와 같은 인공지능웹툰 제작도구를 이용하여 캐릭터와 배경을 선택하고 텍스트를 넣어 주면 텍스트에 따라 캐릭터의 표정, 동작 등을 인공지능이 추천하여 누구나 쉽게 웹툰을 제작할 수 있다. 이 인공지능 웹툰에는 페르소나 인공지능이 적용되어 있는데 이것은 챗봇 인공지능 기반으로 음성인식, 자연어인식, 감정인식을 하는 인공지능이다. 이런 도구를 이용하면 어린 학생들도 만화로 자신의 생각을 자유롭게 표현할 수 있다. 인공지능 만화 그리기 도구가 읽은 책 소개하기, 감상문 쓰기, 생활에서 경험한 것 표현하기, 생태환경, 생활의 불편함을 개선하는 아이디어 내기 등 만화로 생각을 표현하는 학습에 활용될 수 있다.



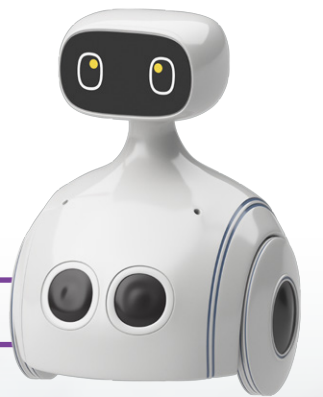
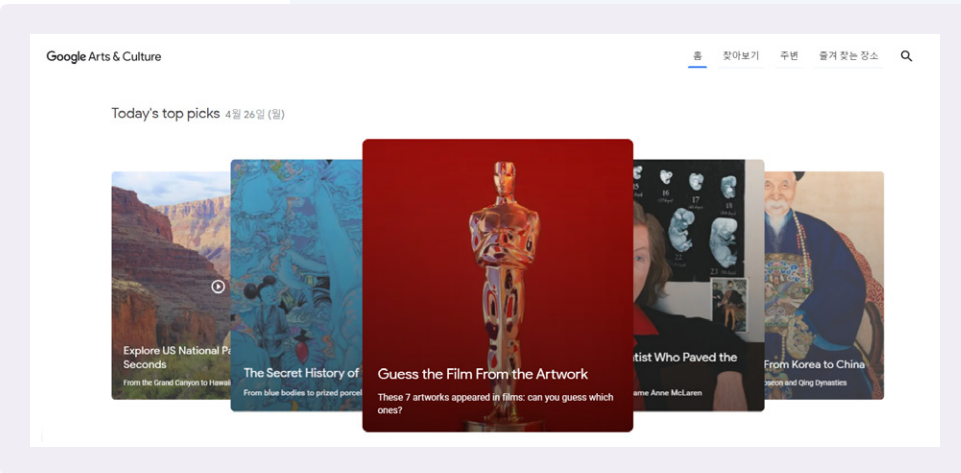
인공지능의 머신러닝을 경험할 수 있도록 서비스하고 있는 구글의 티처블 머신(Teachable Machine)을 활용하면 코딩 없이 머신러닝의 과정을 경험할 수 있다. 티처블 머신은 이미지와 소리, 자세를 분류하는 머신러닝을 통해서 다양한 분류 인공지능을 제작할 수 있다. 티처블 머신은 사용자가 학습에 필요한 사진을 촬영하거나 소리를 녹음해서 이를 기반으로 분류 작업을 한다. 하지만 여기서 사용자가 제공하는 데이터만으로 머신러닝을 진행한다면 높은 성능을 보이기 어렵다. 따라서 티처블 머신은 이미 사전에 훈련된 모델에 전이 학습을 진행한다. 이런 서비스 때문에 다양한 모양 분류, 동식물 분류, 암석분류, 다양한 소리와 자세를 분류하는 활동을 간단한 클릭 몇 번으로 가능하게 할 수 있다.



딥드림 제너레이터(Deep Dream Generator)라는 인공지능은 딥러닝 기술을 시각 이미지에 적용한 기술로, 다양한 사진을 고흐나 고갱이 그린 그림처럼 바꾸거나 다른 질감을 입힐 수 있고 이미지를 조작·왜곡하여 새롭고 추상적인 이미지로 바꿀 수도 있다. 이처럼 인공지능을 활용하여 새로운 영감을 얻거나 시각적인 콘텐츠를 제작하는데 도움을 받을 수 있다.

이외에도 구글렌즈, 구글 아트앤컬처(Google Arts & Culture), 구글번역, 파파고번역, 구글포토, 두들 바흐, 구글 어시스턴트, AI Duet 등 셀 수 없이 많은 인공지능 활용 도구들이 이미 개발되어 있다. 물론 이 도구들이 어떤 과정을 거쳐서 결과물을 내는지 그 과정을 알 수 없어서 인공지능을 블랙박스로 부르기도 하지만 인공지능에 대해 친숙해지고 인공지능을 활용하는 것만으로도 교수학습에 도움을 줄 수 있다. 이렇게 다양

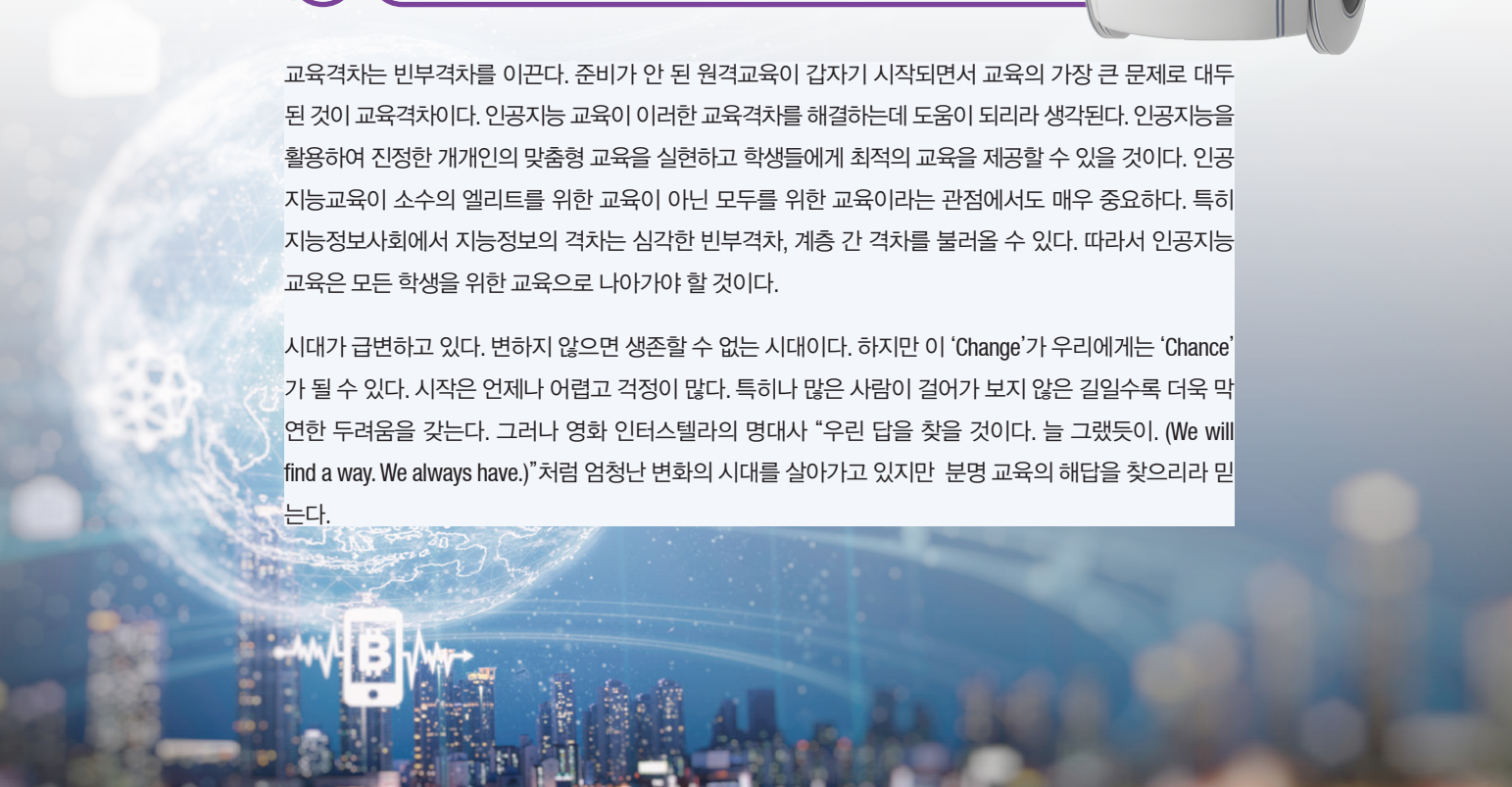
한 활동에 인공지능을 활용하는 경험을 통해 학생들과 교사들도 인공지능에 보다 가까이 다가가고 깊이 있게 학습하고자 하는 시도들이 생길 것이다.



모두를 위한 인공지능교육

교육격차는 빈부격차를 이끈다. 준비가 안 된 원격교육이 갑자기 시작되면서 교육의 가장 큰 문제로 대두된 것이 교육격차이다. 인공지능 교육이 이러한 교육격차를 해결하는데 도움이 되리라 생각된다. 인공지능을 활용하여 진정한 개인 맞춤형 교육을 실현하고 학생들에게 최적의 교육을 제공할 수 있을 것이다. 인공지능교육이 소수의 엘리트층을 위한 교육이 아닌 모두를 위한 교육이라는 관점에서 매우 중요하다. 특히 지능정보사회에서 지능정보의 격차는 심각한 빈부격차, 계층 간 격차를 불러올 수 있다. 따라서 인공지능교육은 모든 학생을 위한 교육으로 나아가야 할 것이다.

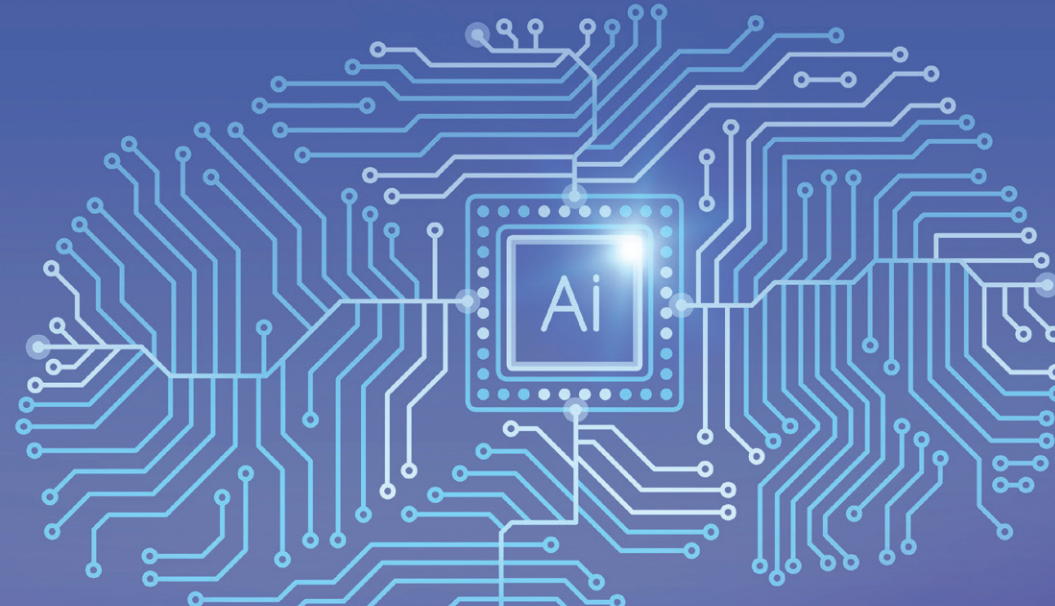
시대가 급변하고 있다. 변하지 않으면 생존할 수 없는 시대이다. 하지만 이 'Change'가 우리에게 'Chance'가 될 수 있다. 시작은 언제나 어렵고 걱정이 많다. 특히나 많은 사람이 걸어가 보지 않은 길일수록 더욱 막연한 두려움을 갖는다. 그러나 영화 인터스텔라의 명대사 "우린 답을 찾을 것이다. 늘 그랬듯이. (We will find a way. We always have.)" 처럼 엄청난 변화의 시대를 살아가고 있지만 분명 교육의 해답을 찾으리라 믿는다.



ISSUE 03

허정 한경대학교 법경영학부 교수 | 최혜리 소프트웨어산업정책연구소 연구원 |

AI(인공지능) 산업통계 개발



인공지능 산업 동향

인공지능의 역사는 반세기 이상을 거슬러 올라가지만, 일반 사람들에게 인공지능이라는 용어가 익숙해진 것은 생각보다 오래되지 않았다. 그러나 스마트폰의 등장이 단기간에 전 세계를 바꿔놓은 것처럼, 인공지능이 가져올 변화 또한 매우 급격하고 훨씬 큰 변혁일 것으로 예상된다. 4차 산업혁명과 함께 인공지능을 포함한 디지털 신기술에 쏟았던 전 세계적 관심은 포스트 코로나 시대와 맞물려 더욱 증폭되고 있다. 실제로 인공지능 시장 규모는 가파르게 성장하고 있다. IDC에서는 2021년 전 세계 인공지능 시장 규모를 전년 대비 16.4% 성장한 3천 275억 달러로 전망하였고, 2024년까지 5천 543억 달러에 이를 것이라는 예측을 발표했다. PwC는 인공지능이 글로벌 경제에 미치는 경제적 가치가 2030년 15조 7,000억 달러에 달할 것으로 추산하며 오늘날 가장 큰 시장 기회를 창출할 것이라고 기대했다.

해외 주요 선진국에서는 민간 주도의 인공지능 시장 활성화가 진행 중이다. 인공지능 분야는 인간 수준의 지능을 구현하기 위해 많은 데이터와 빠른 처리를 위한 하드웨어와 소프트웨어를 필요로 하는데, 특히 전통적인 소프트웨어 강국인 미국은 Google, MS, Facebook, Amazon 등의 선도 기업들을 중심으로 인력·데이터·하드웨어·투자 등 모든 면에서 앞서 있다. 실제로 미국은 인공지능 전 분야에서 대학, 기업들의 기술 수준이 압도적인 우위에 있다. Element AI의 글로벌 AI 인재 보고서에서는 인공지능 관련 21개의 국제학회에 실린 논문을 바탕으로 AI 학술분야 인재를 조사하였는데, 미국이 10,295명, 중국 2,535명, 한국 405명으로 나타났다.

국내 및 해외 주요국 인공지능 기술 수준 비교

국가	상대수준(100%)			기술격차(년)
	기초	응용	사업화	
미국	100.0	100.0	100.0	0.0
한국	86.2	88.3	87.6	1.5
일본	87.2	89.0	88.3	1.4
중국	90.8	92.7	92.0	1.0
유럽	93.3	92.8	89.3	1.0

*자료 : IITP(2019), ICT 기술수준 조사보고서

영국 및 캐나다는 인공지능 최고 권위자와 핵심 인력을 보유하고 있어 머신러닝의 허브 역할을 하고 있으며, 유럽의 경우 사회 현안, 식량문제, 기후 문제처럼 문제 해결 중심으로 기술 개발을 진행하고 있다. 중국, 일본에서는 경쟁우위가 있는 기존산업 및 유망산업을 바탕으로 한 시장 경쟁력 확보를 위해 노력하고 있다. 중국은 특히 딥러닝 등의 활동이 활발해지면서 기술력이 상승하고 있으며, 자국 기업을 활용한 산업별 플랫폼을 구축, 막대한 데이터를 축적함으로써 인공지능 경쟁력 확보에 나서고 있다. 실제로 중국은 슈퍼컴 Top 500 랭킹 1위이며, 인공지능 학술연구(논문게재) 실적이 세계 1위이다. 일본은 인공지능 자체에 대한 기술 개발이나 직접적인 관심·투자보다는 로봇, IoT, 빅데이터 분야 등 응용 측면에서 접근하고 있는 것으로 보인다. 한국도 주요 기업들이 인공지능을 전략 사업 부문으로 인식하고 본격적인 투자 확대에 나서고 있다. 삼성전자는 세계 5개국에 7개 인공지능 센터 설립, 네이버는 인공지능 R&D를 위한 네이버랩스 설립, 이동통신사들은 인공지능 스피커, 스마트홈 투자 확대, 현대자동차는 인공지능을 전략투자 분야로 선정하고 투자를 확대하고 있다.



인공지능 산업 범위

인공지능은 기반 기술의 일종으로 다양한 분야에 적용됨에 따라 개념 정의 또한 여러 관점에서 이루어지고 있으나, 공통적으로는 인간의 지적 활동을 모방하거나 대체하는 기술로 정의할 수 있다¹⁾. 산업의 가치 사슬 측면에서 본다면 인공지능 산업은 1. 인공지능 생산 도구 제공, 2. 인공지능을 활용한 제품 및 서비스 제공, 3. 인공지능 시스템 구축 및 지원을 위한 서비스 제공으로 구성된다²⁾. 따라서 인공지능 산업은 인공지능 관련 기술을 개발하거나 인공지능을 활용한 제품 및 서비스를 생산·유통·활용하는 등의 과정에서 가치를 창출하는 산업이라고 할 수 있다.



보다 세부적으로 살펴보면, 협의의 의미에서 인공지능산업은 인공지능 소프트웨어, 하드웨어, 서비스를 생산·제공하는 산업이다. 인공지능을 활용하여 제품·서비스를 생산·제공하는 기업들은 인공지능 전방 산업으로 분류할 수 있다. 전방 산업에는 인공지능 수요기업과 개발기업이 모두 해당할 수 있는데, 전방 산업의 일부 기업은 인공지능 산업의 개발 영역을 포괄하기 때문이다. 따라서 타 산업에서 인공지능을 적용하는 산업 분야로 자체 개발 업무(인공지능 전문 조직 존재)가 있는 '타 산업-개발기업'과 자체 개발이 아닌 개발된 솔루션을 구매하는 '타 산업-수요기업'으로 구분된다. 또한, 데이터 수집·구매 및 구축 컨설팅, 데이터 거래, 분석 등 인공지능 산업과 연계된 생태계는 인공지능 후방 산업으로 볼 수 있다.



해외 인공지능 정책 및 통계 현황

해외 주요 선진국에서는 국가별 특성에 맞게 지원 중심의 정책을 실행 중이다. 미국은 'AI 이니셔티브 행정 명령(2019)' 등 연구개발과 인력에 대한 정부의 장기적·선제적 투자를 통해 민간의 자생적 경쟁력을 높이고, 인공지능 투자에 우선순위를 부여하고 있다. 특히, 민간이 추진하기 어려운 차세대 연구개발 및 군사·안보 분야 활용에 중점을 두었다. 유럽연합은 정부 주도의 호라이즌 2020(Horizon 2020) 프로그램을 진행하였으며, 후속프로그램인 호라이즌 유럽(Horizon Europe)을 통해 2027년까지의 구체적 실행 계획을 발표하였다. 독일은 인공지능을 통한 중소기업 분야 산업경쟁력을 확보하고 노동시장 변화에 대응한 직업훈련 등을 포괄하는 인공지능 육성전략(2018)을 실행하였으며, 영국은 민간과의 협력을 기반으로 인공지능

인재 양성 및 비즈니스 환경조성에 투자를 집중하는 AI Sector Deal(2018)을 제안하였다. 프랑스는 산업경쟁력 확보, 사회문제 해결과 더불어 일자리 변화 대응을 포함하는 인공지능 권고안(2018)을 발표하였다. 중국, 일본 등에서도 국가 주도의 인공지능 산업 육성 및 인재 양성 정책 수립을 시행 중이다. 중국은 2017년 7월 '차세대 AI 발전 계획'을 발표하고, 2020년까지 인공지능 산업 규모를 1,500억 위안 이상, 인공지능 관련 산업 규모는 1조 위안 이상 달성을 목표로 제시하였다. 인터넷 플러스와 중국 중흥 2050의 일환으로 국가 주도로 인공지능 관련 분야를 이끌고 있으며, 해외에 있던 인공지능 관련 인재들이 중국으로 복귀 중이다. 일본은 인공지능 전략 2019 및 Society 5.0 등 산업 활력 제고 및 저성장·고령화 등 사회문제 해결을 위한 수단으로 인공지능 기술혁신을 가속화하고, 연간 인공지능 응용인재 25만 명, 고급인재 2,000명, 최고급인재 100명 등을 양성할 예정이다.

이렇듯 다양한 인공지능 정책, 전략 추진과 더불어 해외 각국은 인공지능 관련 통계 수집·작성에서도 적극적으로 임하고 있다. 미국, 캐나다 등 북미 지역에서는 미국이 특히 여러 기초통계 및 가공통계를 통하여 인공지능 활용 현황과 직업 수요 전망 등을 조사하고 있다. 북미 지역의 경우 대부분 Gartner, McKinsey, Element AI 등 민간 컨설팅 기업이 작성한 통계가 대부분인 것이 특징이다. 유럽은 유럽연합 국가들을 대상으로 유럽연합 집행위원회에서 실시한 인공지능 관련 조사를 다수 확인할 수 있었는데, 유럽 여러 국가를 대상으로 하기 때문에 국가 간 비교 또한 가능하여 높은 활용도가 기대된다. 아시아 지역에서는 중국과 일본을 주요 검토 대상으로 봤을 때 다른 지역에 비해 활용 현황조사뿐만 아니라 인식조사를 활발하게 진행하고 있다는 특징이 있다.

주요 국가별 인공지능 통계 현황

구분	국가	통계명	수행기관
기초통계	미국	2017년 딜로이트 인지기능 현황조사	Deloitte
		2019년 CIO 설문조사	Gartner
		IoT 프로젝트의 발판을 마련하는 인공지능	Gartner
		AI 도입을 시도하고 있지만, 기술격차로 인해 어려움을 겪는 기업	Gartner
		인공지능 시대의 삶: 5개 산업 분야에서의 인공지능의 성과 및 과제	KPMG
		2020년 글로벌 AI 실태조사	McKinsey
	캐나다	미 연방정부의 AI 도입 실태 결과	Stanford University
		글로벌 AI 인제보고서 2020	Element AI
	유럽연합	유럽 기업의 인공지능 기술 도입 및 활용, 인식 현황 설문조사	European Commission
		유럽의 인공지능	Microsoft/EY
		인공지능 백서에 대한 협의: 유럽의 접근	European Commission
	프랑스	유럽연합과 인공지능	European Commission
		인공지능으로 가는 프랑스 기업	Tata CS
	중국	인공지능 국민인식 조사	ByteDance
인공지능 교육부문 교육현황에 대한 조사		중국인공지능발전전략연구원	
중국의 과학자 및 프로젝트 관리자 설문조사		CSET	
인공지능이 직장에 미칠 영향에 대한 조사		일본노동조합총연합회	
일본	과학기술에 관한 국민인식조사 - Society 5.0	문부과학성	
	2018 ICT 도입·활용 및 대처 상황에 대한 국제기업 설문조사	총무성	
	지자체에서의 AI 활용에 관한 조사 연구	사카시 지방자치연구기구	
	AI/로봇에 대한 초등학교생의 인식조사	야마가타 국립대학	
가공통계	미국	2020년 AI 예측	PwC Japan
		AI 프런티어의 메모: 수백 개의 활용 사례로부터의 시사점	McKinsey
		AI 프런티어의 메모: 인공지능이 세계 경제에 미치는 효과의 모형화	McKinsey
		인공지능 투자 분석	CSET
	캐나다	미국의 인공지능 직업 관련 시장 수요 분석	Gartner
		AI관련 기업 순위 지표 개발	CSET
	유럽연합	2019 캐나다 AI 생태계	Element AI
		AI 세계 구축: 국가적·지역적 AI 전략 보고서	CIFAR
중국	AI 프런티어의 메모: 유럽의 디지털 및 인공지능 격차 대응	McKinsey	
	중국기업2020: AI 순의 보급 및 확산 백서	Alibaba/Accenture	

해외 통계 사례에서 인공지능과 관련하여 포함되는 주요 조사항목들을 살펴보면, 인공지능 활용 현황, 인공지능 활용이 미치는 영향, 인공지능 관련 인식, 인공지능 관련 정부 차원의 노력 등으로 구분할 수 있다.

1) 백승의, 임구진, 여동숙(2016), 인공지능과 사회의 변화: 정보화정책 23(4), 3-23.
 2) 허정, 최혜리, 권이슬(2020), 인공지능산업실태조사: 소프트웨어정책연구소.

해외 국가별 인공지능 통계 주요 조사항목

분류	조사항목	
인공지능 활용 현황	AI 기술 도입 장벽 활용 중/계획인 AI 기반 기술/분야 AI 기술 도입 현황/계획 AI 관련 자금조달/투자 규모 AI 기술 도입 방식 AI 기술 수준 AI 기술 도입 단계(성숙도)	AI 기술 도입/활용 주체 AI 관련 수요 AI 활용 비즈니스 사례 AI 기술 도입 경로 AI 기술 도입 성과 AI 스타트업 기업 현황 AI 시장 규모
인공지능 활용이 미치는 영향	AI 활용 이유/기대효과 업무환경 변화 일자리 변화	경제적 변화 산업 변화 비즈니스 변화
인공지능 관련 인식	AI 활용을 위한 필요 역량 AI 발전에 대한 인식 AI 활용에 따른 우려사항 국제 협력 필요성 AI 발전 동력 AI 인지 여부	AI 중요성 인식 AI 특성 AI 활용 기술에 대한 정보 욕구 AI 활용 시 주의사항 AI 활용을 신뢰할 수 있는 분야
인공지능 관련 정부의 노력	AI 규제 필요성/중요성	AI 전략별 중요성

인공지능과 관련된 해외 통계에서 주로 생산되는 통계 유형을 기초 조사통계 기반으로 살펴보면, 인공지능 관련 기술, 재무/투자, 인력/고용, 경영관리/환경, 법률/정책, 인식, 교육 등으로 구분할 수 있으며, 각 통계 사례를 통해 통계생산 현황의 파악이 가능하다.



국내 인공지능 정책 및 통계 현황

한국 또한 "IT 강국을 넘어 AI 강국으로!"라는 슬로건으로 범정부 역량을 결집, 인공지능 시대의 미래 비전과 전략을 담은 '인공지능 R&D 전략'과 '인공지능 국가 전략'을 2018년, 2019년에 잇따라 발표하며 인공지능 산업경쟁력 강화와 더불어 사람 중심의 인공지능 실현을 위한 국가 비전 실현을 노력하고 있다. 인공지능 국가 전략에서는 1. 세계적 수준의 인공지능 경쟁력을 확보, 2. 경제·사회 전반의 활력을 제고, 3. 국민 행복과 삶의 질 향상 등의 내용을 담고 있다. 한국은 전통산업의 글로벌 경쟁력 저하로 인한 인공지능 기반 새로운 성장 모멘텀 확보가 필요한 시점이다. 국내 제조업 부가가치율은 25%를 상회하는 수준으로 주요 선진국보다 낮다(Global Insight, 2018). 또한, 해외 인공지능 선도국과의 국가 경쟁력 격차가 크데, 인공지능 활용 역량이 있어서는 어느 정도 수준에 도달해 있으나 인공지능 기술 자체의 경쟁력은 낮은 편이기 때문이다. 국내 인공지능 시장의 협소한 규모(글로벌 대비 국내 인공지능 시장 비중은 1.3%; IDC, 2018)와 낮은 투자 수준(인공지능 투자 비중은 0.7%; 중기부 2019)은 자생적 인공지능 생태계 형성을 어렵게 하는 원인으로 작용하고 있다.

이러한 국가 차원의 비전 실현을 위해 정부에서는 인공지능 산업에 대한 특성을 보다 정확히 파악하고자 '인공지능산업실태조사'를 2019년에 국가 승인 통계로 지정하여 수행하고 있다. 실제로 변화하는 미래 국가 혁신성장을 견인하고 복잡한 디지털 산업 생태계 육성을 위해서는 인공지능산업 또한 증거기반 정책(Evidence-based policy)의 수립이 강조되는 시점이다. 최근 증거기반 정책은 객관적이고 실증적인 자료에 기반하여 정책적인 결정을 하려는 시도로써, 과학적 증거가 정책적 의사결정의 기반이 될 때 사회적 목표가 더 잘 수행될 수 있다는 이론에 근거하여 정책적 수요가 증가하고 있다. 특히 인공지능과 같은 신산업 육성을 위한 정책 수립은 시장의 불확실성이 높아 산업 수요에 기반한 정보 수집이 필수적이다. 현재 국내에서는 인

공지산업과 관련된 기초통계로 과학기술정보통신부에서 주관하는 인공지능산업실태조사가 유일하다.

인공지능산업 생태계 조성·활성화 정책 수립을 위한 현황 파악 및 기초자료 확보를 위해 실시 중인 본 조사는 인공지능 소프트웨어, 하드웨어, 서비스를 생산·제공하는 인공지능 기업들의 사업 현황 및 실태 파악이 목적이다. 조사 대상은 '1인 이상의 인공지능 관련 사업을 영위하는 사업체: 인공지능 기술 개발 및 인공지능 적용 제품·서비스·플랫폼의 생산, 유통, 활용, 부가서비스(조사/분석, 컨설팅, 중개) 과정에서 가치를 창출하는 사업체'로, 일반 현황, 기술 및 시업, 매출, 인력, 투자 및 개발, 애로사항 등의 조사항목으로 구성되어 조사를 진행하고 있다.

이외에도 국내에서 진행되고 있는 SW·ICT 관련 실태조사 및 기타 유관 실태조사 등 총 23개의 통계를 검토하여 인공지능 관련 설문 문항 유무를 확인해 봤을 때 최종적으로 8개의 기초통계에 대해 세부적인 분석이 필요하다는 결론에 이르게 되었다.

국내 인공지능 관련 조사항목 유형 비교

구분	유형		
	사업	기술	인력
소프트웨어산업실태조사	[신SW사업 진출 현황] ① 세부 분야 ② 적용 사례 [신SW사업 성과] ① 매출 비중 ② 매출 발생 예상시기 [신SW사업 추진계획] ① 추진 계획 ② 연구개발 투자 비중	[공개SW 도입 분야 및 활용 수준(범위)] [신SW사업 기술역량] ① 연구개발 비중 ②(세계/국내) 최고 기업 대비 수준	[신SW사업 인력] ① 현황 - 학력별 ② 부족 인원 ③ 채용계획 인원- 신임/경력별
소프트웨어융합실태조사	-	[SW신기술 현황] ① 도입 현황 ② 실행 단계 ③ 향후 1~2년 내 도입 계획	[SW신기술 인력 현황] ① 현황 - 학력별 ② 부족인원 ③ 채용계획 인원
ICT중소기업실태조사	[4차 산업혁명 관련 제품 개발/출시 현황]	[4차 산업혁명 관련] ① 개발/활용 중인 기술 분야 ② 해당기술 확보(습득) 방식 [4차 산업혁명 대응을 위한 정책 지원]	-
ICT기업경기조사	-	[국내 ICT산업 발전을 위한 주력 분야]	-
전자문서산업실태조사	-	[적용 중(예정)인 R&D 신기술 분야]	-
정보보호산업실태조사	-	[4차 산업혁명 관련] ① 개발/활용 중인 기술 분야별 제품명 ② 매출액/수정매출액	-
중소기업정보화수준조사	-	[ICT 신기술 필요도] [ICT 신기술 도입 계획 및 의사] [ICT 신기술 도입 현황] ① 도입 시기 ② 활용 수준 ③ 투자 비용 ④ 담당 인원	-
정보화통계조사	-	[AI 기술/서비스 관련] ① 이용 여부 ② 향후 이용 의향 및 이유(이용 목적, 미이용 이유)	-

국내 통계에서는 주로 신소프트웨어 기술, 4차 산업혁명 관련 기술 현황을 파악하는 문항에 인공지능이 일부 들어가 있는 형태를 보이며, 전반적인 문항 구성을 보면 기술 도입 여부 및 계획, 투자 규모, 인력 현황, 애로사항 등 일반적인 산업실태조사 항목으로 구성되어 있음을 확인하였다. 다만, 유럽의 통계에서 공통적으로 확인할 수 있는 인공지능의 영향력 관련 내용 및 산업 지원 정책 관련 인식에 대해서는 거의 다루고 있지 않은 것으로 보인다. ICT 국가 경쟁력 확보를 위해 신기술 분야에 대한 국가적 차원의 지원 필요성이 강조되는 시점인 만큼, 해당 문항들에 대한 검토도 필요할 것으로 생각된다.



국내 인공지능 산업 통계 체계 개선 방향

인공지능은 단일 기술이 아닌 범용기술(General Purpose Technology)로써 전 산업 및 사회 전반에 큰 영향을 끼치게 된다. 인공지능 기술 발전의 속도와 활용 범위 확산이 매우 빠르게 전 산업을 혁신시키고 있어 새로운 산업 환경에 대한 정확하고 신속한 정보의 파악이 필요하다. 특히, 우리나라 전통 주력 산업의 경쟁력이 약화 되는 상황에서 게임 체인저(Game Changer)가 될 수 있는 인공지능에 기반한 디지털 전환 및 부가가치 창출은 국가 경제를 위해 매우 중요한 일이다. 따라서 전 산업에서 일어나고 있는 인공지능 활용에 대한 조사는 실제 산업 현장에서 체감하는 변화들에 대해 정확히 파악할 수 있고, 정책 지원의 기초자료로 활용할 수 있다.

해외 사례를 종합하여 볼 때 국내 인공지능산업실태조사는 인공지능 기술을 개발, 즉 공급하는 기업으로 조사의 범위를 한정하고, 조사 내용 또한 사업, 인력, 기술 등 다방면으로 구성되어 있어 차별성과 경쟁력이 있는 조사라고 할 수 있다. 또한, 생산 측면에서의 효과만을 확인할 수 있다는 점은 체계적이고 정확한 현황 파악에 중요한 역할을 하게 된다. 더불어 해외 통계 사례들의 대다수는 민간 영역에서 글로벌 시장조사기관 등을 중심으로 산업 기초 현황 및 효과 추정을 위한 필수자료에 한정하여 한시적으로 실시된다는 한계가 있다. 이에 반해 국내 인공지능산업실태조사는 국가 승인통계로 자료의 연속성 확보가 가능하며, 향후 시계열 분석을 통한 보다 심도 있는 시사점 발굴이 가능하므로 지속적인 조사와 환경 변화에 맞게 수정, 발전시켜 나갈 필요가 있다.

그러나 국내의 인공지능 통계는 인공지능산업실태조사가 유일하다는 것이 아쉬운 점이다. 앞서 정리된 해외의 인공지능 관련 통계 사례들을 살펴보면, 대부분 기업 혹은 기관의 활용이나 인공지능의 인식 등에 대한 통계들이 다수를 차지하고 있다. 특히, 유럽과 미국, 중국, 일본 등 우리의 주요 경쟁국들은 인공지능에 관한 여러 종류의 조사를 시행하며 다양한 측면에서 인공지능 활용 현황과 인식 수준에 대해 파악하고자 노력하고 있다. 이는 인공지능 관련 정책 개발 및 추진에 있어 실제 각 분야에서 얼마나 활용되고 있으며, 수요기업의 인식이 어떠한지를 중요하게 생각하고 있음을 알 수 있다. 이러한 상황에 비해 한국은 국가 차원에서 조사하고 있는 인공지능 관련 통계로 인공지능 개발기업을 대상으로 하는 '인공지능산업실태조사'가 유일하다. 해당 조사는 인공지능 하드웨어, 소프트웨어, 서비스 등 인공지능 산업에서 공급 측면의 현황을 파악하고자 하는 목적에 있어서는 다른 나라에서 유사한 사례를 발견할 수 없을 정도로 인력, 기술, 자본 등 다양한 내용을 상세하게 조사하고 있어 그 가치가 충분한 조사이다. 그러나 위에서 언급한 것처럼 인공지능이 가지고 있는 기술적, 경제적, 국가적 중요성을 고려한다면 인공지능을 둘러싼 산업 생태계를 보다 폭넓은 관점에서 이해하는 것이 매우 중요한 일이다.

이처럼 범용기술로써 인공지능의 특성과 중요성, 해외 통계 사례들을 참고한다면 국내 인공지능 산업 통계 체계의 개선 및 확장이 필요하다. 먼저 인공지능 산업의 범위를 포괄적으로 반영하면서도 체계적인 통계 생산을 위해 인공지능 산업 주체를 크게 공급기업/수요기업으로 나누고, 각각을 대상으로 하는 독립된 통계조사가 필요하다.

국가 인공지능산업 통계 조사체계 예시

구분	인공지능산업실태조사	인공지능융합실태조사
조사대상	인공지능 하드웨어, 소프트웨어, 서비스 공급 기업 * 정보통신업 내 인공지능 소프트웨어·서비스를 생산·제공하는 사업체 와 제조업에서 인공지능 하드웨어를 생산하는 사업체	인공지능 공급 기업(인공지능산업실태조사)을 제외한 인공지능 하드웨어, 소프트웨어, 서비스 수요 기업 * 정보통신업을 제외한 전산업(통신, 전기전자, 자동차, 의료기기, 조선, 국방/항공, 기계로봇, 섬유제조 등 제조업 및 금융업, 통신/미디어, 유통/물류, 헬스케어, 전문과학 및 기술서비스, 예술, 여가업 관련 서비스업 분야) 내 인공지능을 활용하는 사업체
조사규모	개발기업 사업체(약 930여개)	활용기업 사업체
조사방법	전수조사	표본조사



먼저 현행 '인공지능산업실태조사'는 국내외 사례를 참고하여 일반 현황, 기술 현황, 매출 현황, 인력 현황, R&D/투자 현황 등을 세분화하고, 향후 수출 현황, 사업화 현황 등을 추가할 것을 고려해 볼 수 있다. 향후 인공지능 산업의 특성에 기반한 정책을 수립하기 위해서 해당 산업 내 세부 지표 및 2차 분석 등이 절실하다. 현재까지 인공지능 분야는 기술 개발 단계에서 사업화 단계로 전환되고 있는 바 새롭게 인공지능 산업에 편입된 기업의 특성과 환경들을 면밀하게 파악할 필요가 있다. 과거 전

체 매출액 중 사업 관련 매출액으로 추정하는 방식에서 벗어나 사업체 기반 매출액, 생산액, 수출액, 순이익, 영업이익, 시장가치 등 다양하고 객관적인 성과 지표들이 필요하다. 또한 이를 바탕으로 기업 성장 및 시장 규모에 영향을 미칠 수 있는 인공지능 부문만의 다양한 특성 요인들을 파악하는 것이 필요하다.

둘째, 인공지능 수요기업을 대상으로 주요 산업별 인공지능 활용 현황과 도입 환경을 조사할 수 있는 독립적인 조사체계가 필요한데, 과학기술정보통신부는 2021년부터 '인공지능융합실태조사'를 추진할 계획에 있다. 이와 더불어 인공지능 산업 관련 다양한 가공통계가 생산될 수 있도록 전문통계가 향후 마련되어야 할 것이다. 인공지능 부문 인력 동향 조사, 인공지능 고용 실태조사, 대국민 인공지능 기술 파급효과 인식 조사 등 다양한 측면에서의 조사들을 기획·추진할 필요가 있다.

셋째, 글로벌 차원에서 비교 가능한 전문통계의 개발을 제안하고자 한다. 앞서 해외 주요국의 인공지능 관련 통계를 검토해 본 결과 아직 우리나라를 제외하고는 정부 또는 공공기관 중심으로 기초통계 및 전문통계를 생산하는 국가는 거의 없음을 확인할 수 있었다. 미국, 캐나다 등은 단년도 현황 파악을 목적으로 글로벌 컨설팅 기관에 의뢰하여 분석·활용할 뿐 연속성 있는 통계를 생산하고 있지 않아 글로벌 비교가 불가능한 것이 큰 제약으로 작용하고 있다. 이러한 원인에는 첫째, 인공지능산업은 아직 신산업이란 특성으로 인해 산업의 범위 선정 및 주요 변수 파악이 어렵다는 점이다. 둘째, 우리나라의 국가 주도 정책 수립 특성상 다른 나라에 비해 현황 파악이 빠르다는 측면이 있다. 특히 ICT 부문에서의 통계 체계는 타 국가와 비교해 상대적으로 고도화되고 정합성이 높은 특성을 보인다. 이러한 우수성에도 불구하고 우리나라의 수출 및 국내총생산에서 IT 관련 제품 및 신산업이 차지하는 비중이 높고 향후 전체 산업 대비 인공지능 산업의 비중이 확대될 가능성이 큰 만큼 더욱 선제적인 준비가 필요하다. 국가 차원에서 인공지능 경쟁력을 균형 있게 갖추기 위해서는 현행 제한된 상황에서도 국가 간 비교 가능한 통계 및 지수 개발이 반드시 필요한 시점이다.

마지막으로 인공지능이 국가산업 및 우리 사회에 미치게 될 파급력을 판단하여 해당 수준에 맞는 정책 범위 설정 및 실행전략들을 세우는 것이 필요할 것이다. 이에 인공지능이 미치는 경제적 파급효과를 측정하는 부문도 매우 시급하다고 할 수 있다. 인공지능 기술이 범용기술로서 전 산업 및 사회에 파급력을 가지며, 인공지능 산업은 이를 기반으로 국가 혁신성장을 이끌고 있기 때문에 경제적 효과 추정과 더불어 사회적으로 미치는 간접적 효과까지 추정하는 것이 중장기적으로 합리적인 투자계획을 마련하기 위해 꼭 필요하다.

인공지능이 주도하는 디지털 전환 및 4차 산업혁명의 완성은 'AI+X'에 대한 세부적이고 효과적인 정책 수립이 필수적이다. 이러한 측면에서 인공지능융합실태조사, 인공지능 국가 경쟁력 지수, 경제적 파급효과 추정 등은 전 산업 분야에 인공지능을 기반으로 한 디지털 전환 및 인공지능 활용 수준을 파악함으로써 '인공지능 국가 전략'을 실현하기 위한 증거기반의 세부 정책을 효과적으로 지원하게 될 것이다.



공공데이터 활용성을 선도하기 위한 통계청의 K-통계시스템 구축

통계청 통계데이터기획과 백승환 사무관

01

데이터를 둘러싼 법·제도·인식 등 급격한 환경 변화

4차 산업혁명 시대의 핵심 자원인 데이터의 이용 활성화를 통한 신산업 육성이 범국가적 과제로 대두되고 있다. 이를 위해 2019년 12월 데이터 3법(개인정보보호법, 신용정보법, 정보통신망법)이 개정되었다. 상업적 목적을 포함한 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있도록 허용한 것이다. 아울러, 코로나19를 극복하기 위해 D.N.A(Data, Network, AI) 생태계 강화, 데이터담 구축 등 디지털 혁신을 통한 지능형 정부로의 변화를 추진 중이다. 또한, 인터넷, SNS, IoT 등 다양한 경로로 개인과 기업의 다양한 데이터가 수집되고 있고, 공공기관이 보유한 데이터를 활용하여 신산업을 창출하

거나 데이터에 기반한 과학적 행정을 활성화하기 위한 법적 제도가 마련되고 있다.

이런 제도와 환경의 변화는 대량의 데이터가 저장되고 활용되는 분위기를 확산시킬 것이다. 동시에 데이터 연계, 결합, 활용 과정에서 개인정보가 재식별되거나 오남용될 가능성 등에 대한 불안감도 확대되고 있다. 개인정보를 가명화하더라도 다양한 원인에 의한 재식별화의 가능성은 열려 있어 프라이버시 침해 및 법적 책임 문제가 대두될 수 있을 뿐만 아니라 데이터 관리 차원에서 외부의 해킹이나, 관리 실수, 내부자에 의한 고의 유출 위험성이 있기 때문이다.

❖ 데이터 프라이버시 침해 사례

· 원본자료 (microdata) 공유에 의한 프라이버시 침해(예시)

ID	지역	연령	성별	직업	자산	부채	소득
1	서울	83	남	무직	28,874	1,784	570
...
3,571	제주	50	여	군인	34,475	2,547	4,785
...

유일한 특성을 가진 개인을 식별 개인의 민감한 정보 노출

· 분석결과 (macrodata) 제공에 의한 프라이버시 침해(예시)

지도 상에 분석결과 서비스	(이용자 공격) 분석 조건 반복 선택	인원	비밀 정보
남 + 50~65세 + 박사 + 기혼	5	아파트 1개동 남, 50세, 박사 1명 : 식별	
남 + 50~65세 + 박사 + 기혼, 이혼	6	민감정보 이혼 : 노출	
남 + 51~65세 + 박사 + 기혼, 이혼	5		
남 + 50세 + 박사 + 이혼	NA		

(기존 보호 조치) 5미만 NA 표기

❖ 전통적인 데이터 프라이버시 보호 기법

· 비식별조치 가이드라인('16.6., '20.9)에 따라 비식별화(자료삭제, 라운딩, 범주화, 국소통합 등) 처리 후 데이터 공유할 경우 데이터 질 저하, 재식별화 우려 등의 문제 발생

왜 새로운 데이터 보호 기술이 필요한가?

데이터의 활용성과 정보보호 간에는 상충관계(trade-off)가 있다고 한다. 데이터 활용 및 결합 시 정보를 많이 남기게 되면 개인정보 침해 또는 재식별 위험이 증대하는 반면, 비식별화를 위해 가명화, 라운딩 처리, 범주화, 마스킹 등의 기법을 활용할 경우 데이터 가치의 훼손이 발생한다. 아울러, 가명 내지 비식별화 처리로 인해 현 시점에서는 재식별 위험이 없어도 추후 다른 정보와 결합할 경우 미래 시점에서 재식별될 위험도 상존한다. 이는 비모수 통계 분석에서 변수의 개수에 비례해서 추정의 정확도가 떨어지는 '차원의 저주(curse of dimensionality)'와 유사한 문제로 볼 수 있다. 따라서, 데이터 활용성과 정보보호 간 상충성을 해결할 수 있는 새로운 기술이 필요하다.



암호화하지 않은 평문(plain text) 형태의 자료를 이용할 경우에는 첫째, 해킹 시 심각한 개인정보 침해가 발생될 수 있고, 둘째, 데이터 사용에 대한 포괄적 동의에 따라 제공되는 범위, 종류, 용도 등을 특정하기 어렵다. 반면, 정보노출위험이 최소화된 기법이 적용된 자료나 암호문(cipher text) 형태의 자료를 이용할 경우에는 정보주체인 국민의 데이터 권한 강화 및 보안성이 개선되고, 장기적으로 데이터 이용에 대한 국민 신뢰 회복 및 공공 데이터 수집, 축적 및 공공 목적 활용성 강화에 기여할 수 있다. 정보노출위험 최소화 기법으로는 차등정보보호(differential privacy), 암호문 체계에서는 4세대 암호기술인 동형암호(homomorphic encryption)가 주목받고 있다.

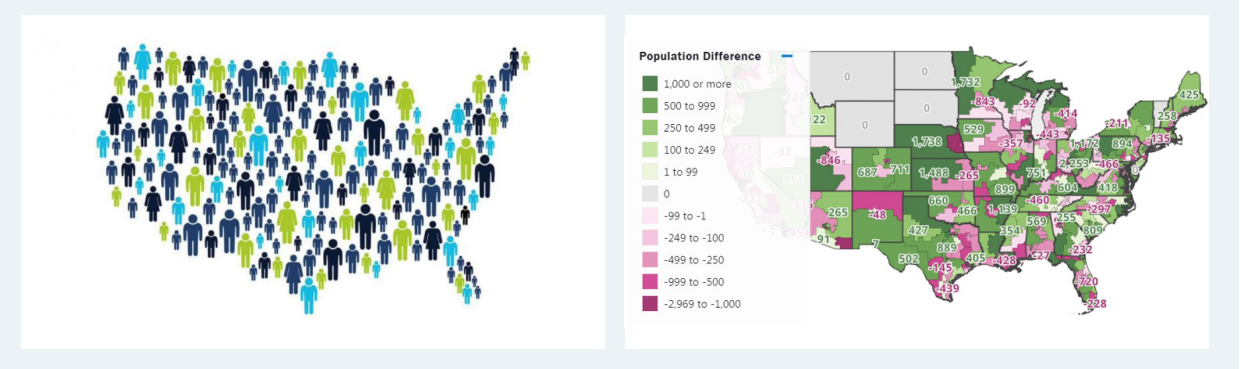
새로운 기술인 차등정보보호 기술

차등정보보호란 사용자 행동패턴을 파악하여 정보가 노출되는 데이터의 특정한 값 대신 근사치로 데이터를 모아 집계 방식으로 보정하는 방법을 의미한다. 정보가 노출되더라도 원래 값이 아닌 근사값이기 때문에 가장 위험한 상황에서도 정보 노출위험을 제어할 수 있다.

예를 들어 표본의 평균을 구할 때 n명의 평균과 n-1명의 평균을 비교하면 제외된 한 개인의 사적 정보를 알 수 있게 되는데, 표본 평균의 결과값에 약간의 잡음(ε)을 추가하여 발표하면 결과값에는 큰 영향이 없는 반면 한 개인의 값을 복원하려고 할 경우 매우 큰 잡음(n×ε)에 직면하게 되므로 개인 정보를 보호할 수 있다.

구글 크롬(chrome) 브라우저는 RAPPOR(Randomized Aggregatable Privacy-Preserving Ordinal Response) 기법 개발을 통해 차등정보보호 방식을 구체화하고 있다. 특정 웹사이트 접근 기록에 대해 차등정보보호 기술을 적용하여 정보를 수집하고 악성 소프트웨어나 웹페이지의 특성을 파악하여 악의적 사이트를 차단하는 기술을 제공하고 있다. 정보를 수집하는 서버는 웹사이트 접근 기록에 잡음(ε)을 추가하여 저장함으로써 프라이버시를 보호하면서도 통계값의 유용성은 유지할 수 있다. 또한, 미국 인구조사(census)도 차등정보보호 방법론이 적용되었다. 차등정보보호가 실제 적용된 사례는 대부분 데이터의 규모가 매우 크고 개인정보보호의 유인이 상당한 경우라고 볼 수 있다.

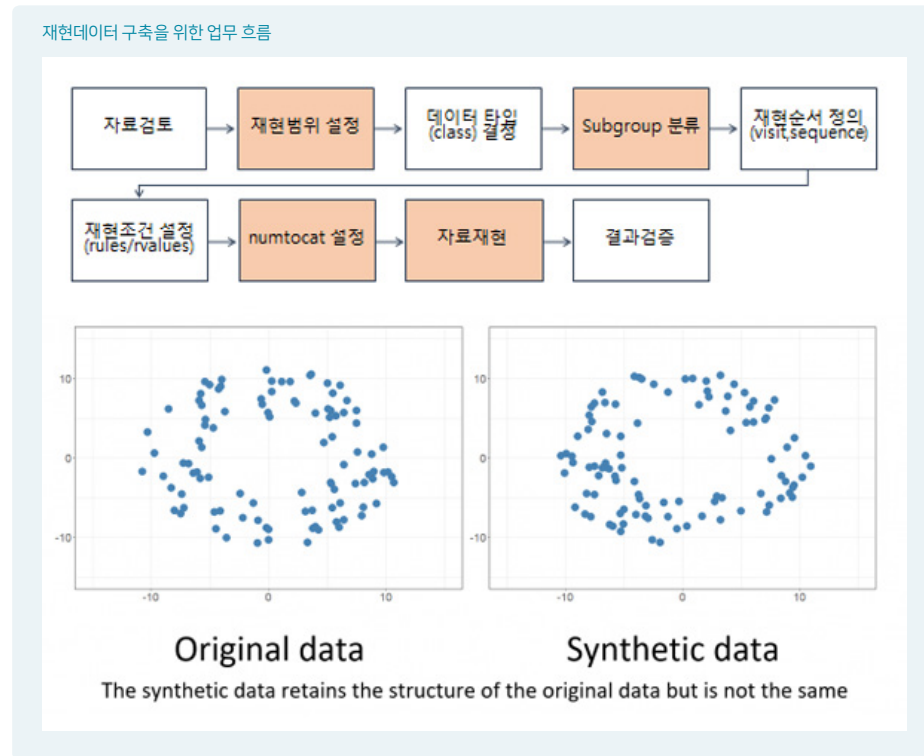
차등정보보호를 도입하여 발표한 2020 US Census



다만, 차등정보보호 방식으로는 복잡한 계산 수행이 어려울뿐만 아니라, 분석 정확도가 손실되어 안전한 결과값 도출이 어려운 한계가 있다. 또한, 하나의 연산에서 개인정보를 보호하는 차등정보보호가 다른 연산에서도 개인정보를 보호할 수 있는 것은 아니라는 점도 유념해야 한다. 이론적인 한계와 별도로 개인정보를 가공한 결과보다는 개인정보가 포함된 로데이터(raw data)를 필요로 하는 경우가 많다는 한계도 존재한다.

차등정보보호 기술을 이용하면 전통적인 노출제한 방법의 정보 손실 문제를 극복하기 위해 구축된 원자료와 유사한 분포를 가지는 가상의 데이터인 재현데이터 구현이 가능하다. 재현데이터는 원본과 통계적으로는 유사하나 가상으로 다시 만들어진 합성데이터(synthetic data)로서 실제 측정된 원자료

와는 다른 자료이지만 원자료를 적절하게 설명하는 가상의 모형을 통해 생성되기 때문에 모형의 특성이 유지되는 장점이 있다. 즉 원자료를 그대로 이용하면서 발생하는 보안 위협을 낮추면서도 원하는 통계값을 도출할 수 있음을 의미한다.

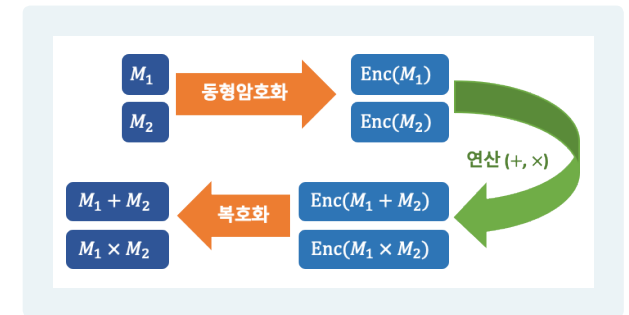
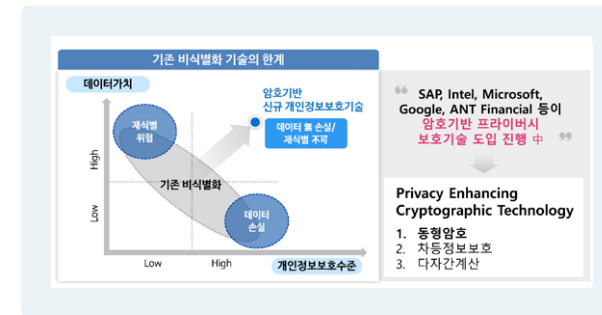


※ Synthetic Data: Unlocking the power of data and skills for machine learning(dataingovernment.blog.gov.uk) 중 발췌

재현데이터는 마이크로데이터 형태로 데이터를 이용할 수 있으므로 데이터 처리에 대한 자유도가 증가하며, 원래 데이터의 분포와 변수 간 관계 등이 유지된다. 이러한 재현데이터는 학습용 데이터 또는 연구용 데이터로서의 활용성이 높다. 현재, 센서스나 통계청 중심의 학술연구를 넘어 학습용 데이터 생산이나 타 분야 서비스 개발을 위한 솔루션과 특화된 재현데이터 전문 스타트업이 등장하고 있다.

또 하나의 새로운 기술인 동형암호기술

동형암호는 암호화된 상태에서 복호화 과정 없이 연산이 가능한 기법으로 평문의 연산 결과와 암호문의 연산 결과를 복호화한 값이 동일하다는 특징이 있다. 즉, 동형암호는 암호화와 연산 간 교환법칙이 성립하는 암호체계를 일컫는다. 연산에 앞서 복호화를 하지 않아도 되고, 최종 연산 후 필요한 경우에만 복호화하면 되기 때문에 개인정보 노출 위험이 자연스럽게 최소화될 수 있어 정보보호를 획기적으로 개선할 수 있다. 동형암호기술은 4세대 암호화 기술로 알려져 있고, 미국 MIT에서는 10대 차세대 기술로 분류하기도 하였다(One of 10 Emerging Technologies, MIT Technical Review, 2011).



※ 평문 M_1, M_2 에 대한 암호문 C_1, C_2 가 주어졌을 때, 이 두 암호문으로부터 $M_1 + M_2$ 에 대한 암호문 $C_{M_1+M_2}$ 혹은 $M_1 \times M_2$ 에 대한 암호문 $C_{M_1 \times M_2}$ 등을 얻을 수 있음

통상적으로 동형암호 개념을 이해하기 위해 들고 있는 덧셈 연산 예시를 통해 동형암호를 설명해보자. 8과 13을 더하는 연산을 동형암호를 통해 실행하면 다음과 같다. 평문은 8과 13이고, 사용하는 키는 3과 5이며 암호화 방식은 모듈러 연산(mod3, mod5)이다. 8과 13은 3과 5로 각각 나눠서 나온(2, 3)과(1, 3)으로 암호화되며, 암호화된 결과를 덧셈 연산하면 같은 수로 나눈 나머지끼리 더해야 하므로 (3, 6)이 된다. 암호화된(3, 6)을 복호화하기 위해 암호화 당시 사용한 키를 활용하게 되면서(3, 6)은(0, 1)으로 전환되고 결국 3으로 나누었을 때 나머지가 0이고, 5로 나누었을 때 나머지가 1인 값은 21로 8+13과 같은 값이 나오게 된다. 이런 원리를 이용하게 될 경우 암호화 상태에서 통계적인 분석과 연산이 가능하다.

동형암호기술은 1970년대에 연구가 시작된 이후 2009년 IBM 연구위원인 Gentry에 의해 기술적 가능성이 증명되었다. 최초의 동형암호는 1비트(bit) 연산처리에 30분 이상이 소요될 정도의 느린 처리 속도가 단점이었다. 그러나, 10년 사이에 처리속도가 약 10만배 빨라졌고, 대용량 데이터 처리 지원과 GPU 연계를 통한 연산 자체의 병렬처리 연구 등을 통해 산학은 원본 데이터 처리 속도와 비슷한 속도를 보유하는 것을 목표로 하고 있다. 동형암호기술은 원본 데이터를 암호화할 때 암호문의 크기가 최소 수십배 이상 커져서 연산 시 처리 속도가 느리다는 단점이 있다. 현재는 여전히 평문 연산에 비해 상당히 처리 속도가 느린 상황이지만, 각종 연구가 진행 중이어서 향후 처리 속도가 빨라질 것을 기대한다. 동형암호기술 연구는 현재 개인정보보호 이슈가 민감한 금융/보험, 유통/마케팅, 의료 분야 등에서 활발히 진행되고 있다. 데이터 결합 활용 시 상호 신뢰를 구조적으로 확보할 수 있어 데이터 오남용에 대한 우려가 해소되고, 개인정보의 유출 없이 필요한 데이터만 추출할 수 있기 때문이다.

국가별 동형암호기술 개발 현황

개발시기(년)	2012	2016	2017
국가			
개발기업/대학	미국	미국	한국, 프랑스, 스위스
동형암호 프로그램	IBM	마이크로소프트(MS)	서울대, 베르사유데, 쟈말토
동형암호 프로그램	HeLib	SEAL	헤안(HEAAN), TFHE
특징	최초의 동형암호 프로그램	사용이 편리한 동형암호 프로그램	연산 속도, 효율, 활용 분야 측면에서 가장 앞서 있음

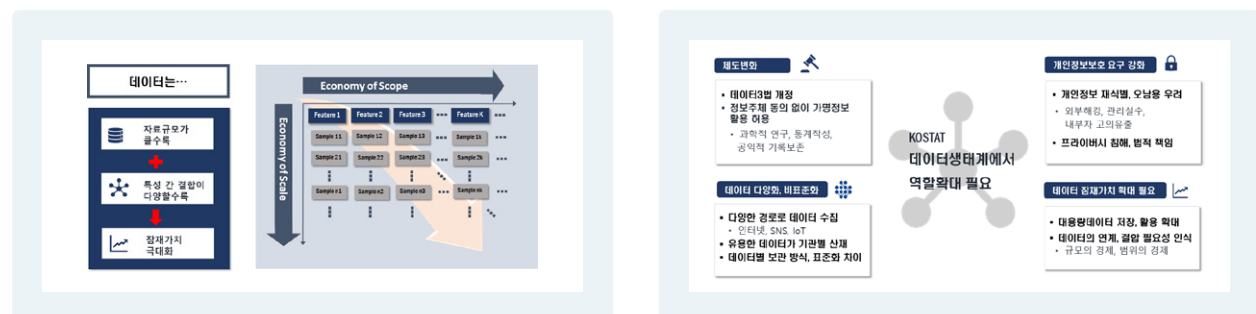
자료: 수학기반 산업데이터에 세 연구센터

동형암호기술은 공공 데이터 영역에도 적용이 가능하다. 개별 공공기관에 산재된 각종 마이크로데이터를 암호화한 채 결합하여 활용할 수 있기 때문이다. 해당 공공기관의 서버를 떠나지 않은 상태에서 기관 간 자료의 결합 활용이 가능해지므로 데이터 연계·결합의 장애물을 극복할 수 있는 유용한 수단이 될 수 있다.

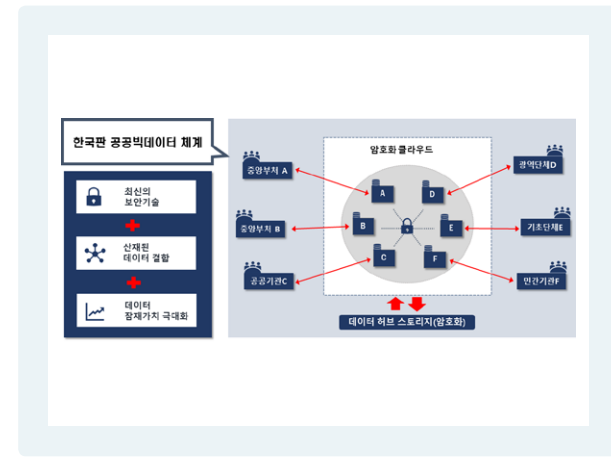
공공데이터 생태계 강화를 위한 K-통계시스템 도입 추진

데이터는 데이터 규모가 커질수록 가치가 증가하는 규모의 경제(economy of scale)는 물론 데이터 특성 간 결합이 다양할수록 가치가 배가되는 범위의 경제(economy of scope)라는 특성을 갖고 있는 재화이다. 공공데이터 역시 일반적인 데이터의 특성을 그대로 가지고 있다. 일반적으로 공공데이터는 공공기관이 행정을 수행하는 과정에서 확보하게 되는 공공재로서 분야별 행정의 대상 전체에 대한 대규모 전수자료라는 특성을 가지고 있다. 또한, 분야별로 생성된 공공데이터의 항목을 타 공공기관의 데이터 항목과 결합할 경우, 그 결합의 횟수가 많아질수록 연계된 공공데이터의 정보가 다양해지면서 정책을 보다 정교하게 설계할 수 있게 된다. 공공데이터 연계를 통해 과학적인 근거에 기반한 효과적인 정책이 수립·시행되면서 국민 삶의 질이 개선될 것이다. 공공데이터 연계·결합을 원활하게 수행할 수 있도록 데이터3법 개정, 데이터기반행정활성화법 제정 등 법적·제도적 기반도 강화되고 있다.

그러나, 이런 법·제도적 기반이 강화되고 있음에도 여전히 양질의 데이터가 각 중앙부처와 지방정부의 독자적인 서버에 분리된 채 쌓여 있는 실정이다. 이는 데이터 원천별 보관 방식이나 표준화 등의 차이, 데이터 보유 기관 간 정보보안과 관련된 상호 신뢰 부족 등에 기인할 뿐만 아니라 데이터의 활용성과 정보보호 간 상충관계가 극복되지 못하고 있기 때문이기도 하다. 이런 문제점을 극복하고, 공공데이터의 부가가치를 충분히 창출하기 위해 새로운 데이터 보호기술을 바탕으로 데이터를 안전하게 연계·결합할 수 있는 데이터 공유·활용 체계를 도입할 필요가 있다.



K-통계시스템은 중앙정부와 지방자치단체 등이 보유한 각종 공공 데이터를 안전하게 연계하여 누구나 활용할 수 있게 만드는 데이터 체계를 의미한다. 즉, 통계청이 보유한 통계등록부 등 데이터베이스를 가교(架橋)로 삼아 산재되어 있는 데이터를 최신의 암호기술 기반으로 연계할 수 있는 클라우드 시스템의 일종이라고 이해할 수 있다. 이때 데이터는 암호화된 상태로 클라우드 공간에 각각 분리해 저장하고, 필요할 때 데이터를 결합하거나 연산할 수 있다. 각 기관의 데이터는 기관별로 소유한 박스에 안전하게 보관되므로 다른 기관의 데이터와 섞이지 않고, 박스와 박스가 충돌해서 깨질 일도 없다.



공공 데이터를 결합하여 사용하게 되면 근거기반정책(evidence-based policy)을 보다 효과적으로 수행할 수 있다. 다만, 이 경우에도 개개 국민의 사적 개인정보는 우선적으로 보호되어야 할 것이다. 이를 위해 K-통계시스템을 활용하면 암호화 기법으로 활용하여 민감한 정보를 암호화한 뒤 암호화된 상태로 연계하거나 분석을 위한 각종 연산이 가능하다. 각 공공기관은 연계, 분석 시 기관별 자료를 하나의 기관으로 이관하거나 하나의 메모에 모두 통합할 필요 없이 필요에 따라 암호화된 고유키를 활용하여 연계하면 되므로 보안 위험에 대한 불안감을 해소할 수 있다. 암호화된 상태로 클라우드 공간에서만 데이터를 활용할 수 있기 때문에 데이터 집적에 따른 '빅브라더'의 발생 우려도 사라진다.

K-통계시스템 도입의 기대효과

K-통계시스템이 구축될 경우 예상할 수 있는 잠재적 응용 사례를 생각해 보면 다음과 같다.

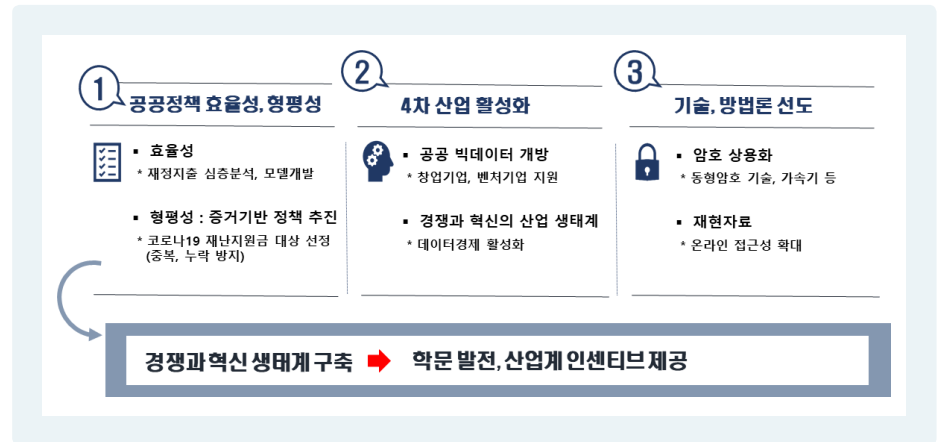
첫 번째는 의료정보 간 결합을 통한 국가 의료 체계 강화 및 기술 향상을 들 수 있다. 국내 의료 데이터의 양과 질은 세계 최고 수준인데 반해 환자의 프라이버시 보호 문제로 데이터 공유가 어려운 상황이며 의료법은 타 기관 간 진료데이터 공유를 엄격히 제한하고 있다. 건강검진 내역, 진료 내역, 처방 내역, 치료 효과 등 의료데이터는 모일수록 개별 병종에 대한 예방·치료방법 개발이나 개인별 맞춤형 의료서비스가 가능해지는 등 가치가 급증하는 특성이 있다. K-통계시스템을 활용하여 암호화된 상태에서 전국민에 대한 의료 데이터를 공유하고 활용할 수 있게 되면서 개인별 의료 정보는 최고 수준으로 보호하면서도 학계, 의료계 연구자의 감염병, 유전병, 난치 질환 연구가 가능해질 수 있다.

두 번째는 공공정책의 효율성과 형평성을 높이는 데 기여할 수 있다는 점이다. 각종 재정지출 자료와 개인·가구 단위 재분배 효과를 연계·결합함으로써 상세한 재정지출 효과 분석을 할 수 있을 뿐만 아니라 맞춤형 정책 모델 개발도 가능해진다. 개인과 가구의 소득과 자산이 정책 시행 전에 비해 정책 시행 후에 어떻게 변화되었는지를 기관 간 데이터를 K-통계시스템 내에서 암호화된 상태로 연계·분석할 수 있기 때문이다. 국가 전반적인 상황 외에 기초자치단체의 읍·면·동 혹은 그 이하의 집계구 형태 분석이 가능해지기 때문에 과학적 근거에 기반한 지방자치 실현을 위한 기초자료 제공도 가능해진다. 아울러, 각종 정책을 설계함에 있어 형평성을 고려한 정책 대상자나 지역 선정 등이 가능해지므로 근거기반(evidence-based) 맞춤형 정책수립을 통해 국가 전반의 형평성을 제고할 수 있고, 정책 대상자 및 지역 누락이나 중복확인을 통해서도 형평성을 높이는 데 기여할 수 있다. 예를 들어 경제·사회 위기 전후의 사업자 매출과 소득을 파악하여 맞춤형 소상공인 지원에 활용한다든지 112신고와 신용카드 데이터를 결합·분석하여 범죄 위험도에 따른 경찰력을 배치하는 등 K-통계시스템을 통한 데이터 결합 가능성은 무궁무진하다.

세 번째는 빅데이터, 인공지능, 사물인터넷 등으로 대표되는 4차산업혁명에 기여할 수 있다는 점이다. 기존 빅데이터를 보유한 디지털 플랫폼 기업과 데이터 접근이 어려운 신생 창업 기업 간 기술어린 운동장을 평평하게 함으로써 데이터 격차(data divide) 극복에 기여한다. 경제 전반에 신규 진입을 촉진

하여 경쟁과 혁신의 데이터 생태계 구축에 기여할 수 있다. 금융데이터, 신용데이터, 보험데이터 간 결합을 통한 새로운 금융상품 개발이 가능해지고, 통신데이터와 교통데이터 결합을 통해 유통체계의 혁신을 가져올 수도 있다.

네 번째는 빅데이터와 암호 상용화 부문에 대한 세계적 선도 가능성이다. K-통계시스템 개발을 통해 암호 상용화를 위한 학문적 발전과 산업계 동형암호 가속기 개발 등에 적극적인 유인을 제공할 수 있다. 또한 ICT 강국을 넘어 빅데이터와 암호 상용화 분야에서도 세계적으로 선도적 위치를 차지할 계기로 작용할 수도 있다. K-통계시스템을 통해 가장 완벽한 암호체계로 공공 빅데이터를 운영하는 최초의 국가가 될 수 있다.



K-통계시스템에서 활용가능한 통계 데이터의 확충 필요

K-통계시스템에서 기관별로 산재된 데이터를 연계·결합하기 위해서는 개인, 사업체 등 단위별 모집단 구축이 필요하다. 즉, 단위별 모집단을 축(backbone)으로 다양한 데이터를 연계·결합하여 필요한 데이터셋을 구축함으로써 데이터 활용이 극대화될 수 있다. 이 축에 해당하는 고품질의 표준화된 모집단으로 기능할 수 있는 것이 바로 통계등록부(statistical register)라고 볼 수 있다.

통계청이 구축하고 있는 통계등록부는 인구, 사업체 등에 대해 조사자료와 각 부처의 행정자료를 연계하여 작성한 통계단위별 모집단을 일컫는다. 현재, 인구·가구통계등록부, 주택·건물통계등록부, 기업(사업체)통계등록부, 경제활동통계등록부가 구축되어 있다. 즉, 통상적으로 데이터를 연계하여 분석할 개인, 가구, 주택, 일자리 등에 대한 모집단을 확보하고 있음을 의미한다. 향후 통계청은 소득, 교육 정보 등 추가 자료를 입수하여 핵심적인 속성정보에 대한 통계등록부 구축을 강화함으로써 K-통계시스템 구축 이후 활용을 활성화하기 위한 노력을 지속적으로 추진할 예정이다.

통계 데이터 서비스 패러다임을 이용자 시각으로 획기적 전환

통계청은 다양한 데이터 간 연계·분석을 지원하고 맞춤형 정책 대응을 위해 보다 촘촘한 정보를 생산하는 등 통계이용자 수요 변화를 기반으로 국가통계관리도 탄력적으로 개선하고자 한다.

우선 실험통계 제도를 마련하고 다양한 데이터 활용을 통한 통계작성을 촉진할 필요성이 있다. 실험 통계란 기존의 승인통계와 다른 빅데이터, 비정형 데이터 등 다른 데이터나 방법론을 적용하여 작성하는 통계로 작성 이후 일정 기간 품질 점검을 위해 관찰이 필요한 준(準)통계를 의미한다. 데이터 환경의 변화에 따라 탄력적으로 국가통계를 관리하기 위한 것으로 실험통계제도 도입을 위해 국가통계 관리체계 개편안을 마련하기 위해 관계 부처와의 협의, 전문가 의견 수렴 등을 진행 중이다. 실험통계가 도입될 경우 실험통계와 국가승인통계와의 차이나 실험통계 운용방식 등을 구체화하여 통계이용자의 혼란을 방지하기 위한 노력들이 병행될 필요가 있다.



또한, 이용자의 지속적인 통계자료 활용 확대 요구에 대응하여 통계자료의 이용범위를 넓혀주는 통계자료 프리존 제도 도입을 검토 중이다. 통계자료 프리존 제도를 도입을 위해 객관적이고 일관성 있는 통계자료 제공 기준을 정립하고, 서비스 플랫폼 일원화를 준비 중이다. 통계청이 보유하고 있는 자료의 개방을 추가적으로 확대하고, 제도 설계와 더불어서 시범과제 운영을 통해 요건의 적절성과 절차의 보완점 및 이용자 요구사항을 발굴하여 통계자료 프리존 제도운영에 반영할 예정이다. 제도 설계와 시범 운영을 병행함으로써 제도의 안정적인 도입· 정착에 기여할 것이고, 공공 빅데이터에 대한 접근성 확대를 위한 토대를 마련함으로써 K-통계시스템을 보다 제도화할 수 있을 것으로 기대한다. 통계 데이터 서비스 혁신을 통해 이용자의 체감 만족도를 높임으로써 데이터 공유·개방·활용에 기여할 수 있게 되기를 바란다.

K-통계시스템 도입·구축·활용을 위한 선결조건

K-통계시스템의 성공적 구축 및 활용성을 높이기 위해서는 아래의 조건들이 충족될 필요가 있다.

첫째, 동형암호기술 상용화 등 K-통계시스템이 안전하고 효과적으로 작동할 수 있는 기술과 장비의 개발이 함께 이루어져야 한다. 통계청은 2020년에 통계데이터 활용 활성화를 위한 동형암호 적용방안 연구를 용역사업으로 진행한 바 있고, 2021년에는 과기정통부 연구개발 사업에 동형암호 국가통계분석시스템 개발 과제가 선정되어 3개년 사업으로 추진될 예정이다. 아울러, 과기정통부 등 관계 부처와 협업하여 새로운 프라이버시 보호기술의 개발을 지속적으로 추진할 계획이다.

둘째, 암호화 상태에서는 연계·결합될 데이터의 구조, 품질 등에 대해 확인할 수가 없으므로 기관 간 데이터의 표준화와 품질이 적절하게 관리될 필요가 있다. 행정안전부 주도로 공공 데이터의 표준화를 추진 중이지만 여전히 기관별 서버에 적재된 데이터의 용어, 도메인 등에 있어 표준화 수준이 미흡한 실정이다. 연구자가 데이터를 직접 눈으로 확인하고 연계·결합의 장애요인을 제거하기 위해 에디팅, 결측값 대체 등 전(前) 처리과정을 거쳐야 데이터 간 연계율이 높아지고 분석 시 과다, 과소해석할 수 있는 여지가 줄어든다. 이런 사전적인 전처리를 직접 수행하지 못하므로 기관별 데이터의 표준화와 품질을 일정 수준 유지하기 위한 노력들이 필요하다.

표준화와 품질 관리를 위한 또 다른 방법으로는 육안으로 수행하지 못하는 전처리 과정을 암호화 상태에서 진행할 수 있는 기술 개발을 제시할 수 있다. 재현데이터를 활용해서 원본 데이터의 통계적 성

격은 유지하면서 새로운 형태의 데이터를 구현해내는 방법이다. 연구자는 재현데이터를 활용하여 데이터 전처리에 필요한 방법론을 구상하고 프로그래밍하여 K-통계시스템을 통해 데이터 연계·결합에 필요한 전처리 과정을 시행하게 된다. 결국, 재현데이터는 K-통계시스템의 작동을 원활화하기 위한 또 하나의 중요한 데이터 보호기술이다.



보다 실무적으로는 국가승인통계 작성에 활용될 수 있을만큼 충분한 데이터 전처리 기술을 보유한 통계청의 기술, 경험 및 노하우를 타부처와 인사 교류 등을 통해 공유할 수 있는 장이 필요하다. 실제 데이터를 정제된 형태로 가장 많이 보유하고 가장 활발히 활용하는 통계청의 각종 경험들이 다른 기관에게 전파되어 데이터 연계·결합을 통한 이용이 활발해지기를 기대한다.

이외에도 기관 간 데이터를 개방·공유·활용할 수 있는 문화가 더 튼튼히 자리잡을 필요가 있다. 대부분의 기관이 데이터 공유를 꺼리는 이유는 보유 데이터가 다른 기관 데이터와 연계·결합되는 과정에서 발생할 수 있는 보안의 문제와 특정 기관에 데이터가 집중됨에 따라 발생할 수 있는 빅브라더 문제 때문이다. K-통계시스템은 새로운 암호기술을 적용하여 보안상 문제를 해결하고, 필요에 따라 데이터를 이관하지 않고 연계·결합하는 한국판 공공 빅데이터 체계이기 때문에 데이터 공유의 저해 요인을 해소할 수 있다. 즉, 데이터 개방·공유·활용의 문화와 K-통계시스템의 구축 및 활용은 정(正)의 상관성을 가지고 있으며, 각 데이터 보유 기관들이 타 기관과 협조하여 데이터를 공동활용할 수 있는 유인체계가 K-통계시스템에 내재되어 있다.

K-통계시스템 구축 및 활용의 전제조건

	통계청	타 기관
새로운 암호기술 상용화	<ul style="list-style-type: none"> 동형암호기술 상용화 차등정보보호 및 재현데이터 	<ul style="list-style-type: none"> 예산, 인력의 적극적 지원
데이터 표준화, 품질 강화	<ul style="list-style-type: none"> 데이터 전처리 기술 전파·공유 	<ul style="list-style-type: none"> 유관 기관 협력을 통한 표준화 수준 및 데이터 품질 향상
제공데이터 확충	<ul style="list-style-type: none"> 소득, 자산, 교육 데이터 추가 입수를 통한 통계등록부 강화 	<ul style="list-style-type: none"> 데이터 공유·활용 문화 촉진

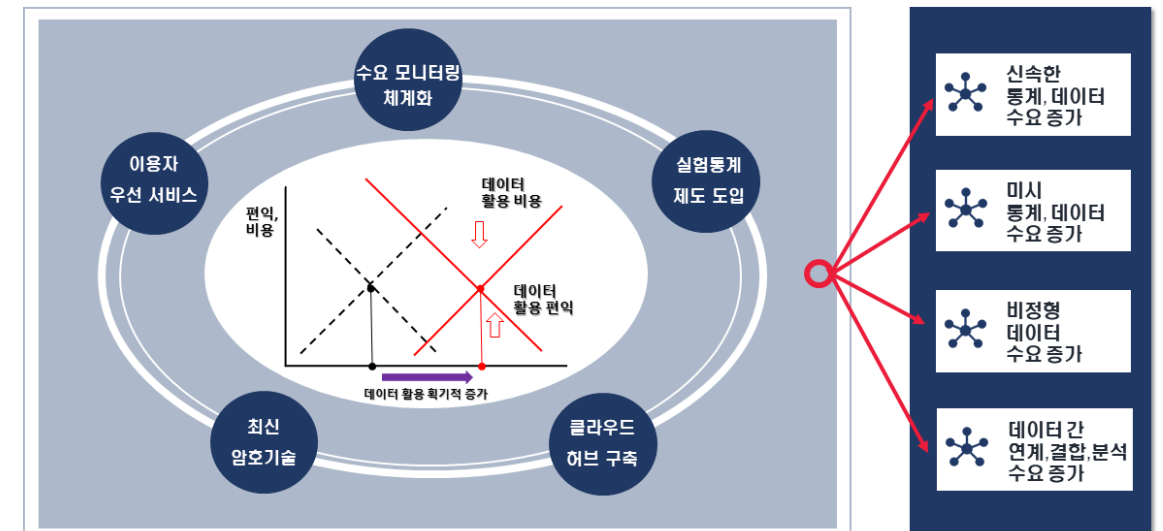
나가면서

우리나라는 K-통계시스템 구축을 위한 최적의 여건을 가지고 있다. 우선 다양한 공공기관 및 민간기업들이 ICT 기술을 기반으로 데이터 간 연계·결합이 가능한 양질의 데이터를 보유하고 있다. 아울러, 데이터 3법이 통과되고, 혁신적 금융서비스를 위한 규제샌드박스가 도입되는 등 데이터 활용을 활성화하기 위한 법·제도적 발판이 마련되어 있다. ICBM(IoT, Cloud, Bigdata, Mobile)으로 대표되는 혁신적인 기술 인프라도 보유하고 있을 뿐만 아니라 D.N.A.(Data, Network, AI) 생태계를 강화하여 양질의 일자리를 창출하기 위해 디지털 뉴딜이 포함된 한국판 뉴딜을 범정부적으로 추진하고 있다. 또한 세계적인 동형암호 원천기술도 보유하고 있다.

특히, 통계청은 통계작성을 위해 광범위한 데이터를 다루는 과정에서 국회, 조사 응답자, 자료 제공기관, 시민사회로부터 데이터 보호에 대한 요구가 급증하고 있는 실정이다. 이에 따라 통계청은 다양한 신기술과 방법론 등을 기반으로 민감한 통계데이터의 안전성을 확보할 수 있는 K-통계시스템을 구축하고, 이를 통해 데이터 간 연계·활용을 활성화하는 기반을 조성하고자 한다.

데이터는 규모의 경제와 범위의 경제가 적용되는 대표적인 재화로, 부처별로 산재한 각종 공공자료를 모을수록 데이터의 가치가 배가될 것이다. 이처럼 데이터의 경제적 특성을 살리면서 데이터 산업 및 경제 활성화를 창출하기 위해서는 기존의 데이터 활용과 정보보호기술 간 상충 문제를 극복하기 위한 새로운 암호기술을 도입함으로써 자료 축적 및 자료 결합과 자료 보안 간 강한 보완 관계를 창출해 나가야 한다. 개인정보 유출 및 데이터 보호에 대한 불안감을 불식시키고 다양한 이해관계자가 안심할 수 있는 수준으로 데이터가 보호되어야 한다.

통계청은 K-통계시스템을 기반으로 우리 경제 혁신의 큰 계기를 마련하기 위한 노력을 지속적으로 경주할 것이다. 세계 최초의 데이터 프라이버시 보호 기술 상용화를 통해 “K-방역” 모델에 이어 “K-프라이버시” 모델이 전 세계를 선도해 나갈 수 있도록 추진하고자 한다. 첨단 암호 기술을 통한 철저한 개인정보 보호와 다양하고 적극적인 정보 활용 방안이라는 두 날개를 장착한 K-통계시스템이 성공적으로 구축돼 우리 정부와 기업, 국민이 데이터 시대를 선도하는 디딤돌이 되기를 기대한다.





02

문화관광 분야에서의 빅데이터 활용

빅데이터의 이해

2016년 세계 경제 포럼(WEF: World Economic Forum)에서 '제4차 산업혁명'이란 용어를 사용하면서, 지금의 시대는 산업혁명시대로 지칭되고 있다. 4차 산업혁명은 인공지능, 사물인터넷, 빅데이터, 클라우드 등 첨단 정보통신기술(IT: Information Technology)을 기반으로 다양한 분야들이 융합되어 혁신적인 변화를 일으키는 것을 의미한다. 이 중에서 빅데이터(Big data)는 최근 몇 년 동안 사회에서 큰 주목을 받고 있는데, 그 이유는 사회변화와 IT발전으로 데이터가 기하급수적으로 증가했지만, 활용하지 못하고 쌓여만 있던 데이터를 새로운 가치 있는 정보로 제공할 수 있기 때문이다.¹⁾

빅데이터는 사회변화와 IT기술의 발달로 인해 기하급수적으로 증가했지만 활용되지 못했던 데이터를 정제 및 가공 분석을 통해 새로운 가치있는 정보로 만들 수 있는 데이터라 할 수 있다. 따라서 대용량의 데이터를 저장하고, 이를 분석할 수 있는 기반시설 및 기술이 필요하다. 또한 기하급수적으로 증가하는 데이터를 시의성 있게 활용하기 위해서는 빠르게 데이터를 정제하고 가공·분석하는 것이 중요하다. 그리고 기술의 발달로 정형화된 데이터가 아닌, 영상/그림(image), 음성(voice), 글자(text) 등 비정형데이터의 패턴 등을 분석하여, 기존의 전통적인 통계분석 방법과 차별화된 정보를 제공함으로써 더욱 가치있는 정보를 우리에게 제공하고 있다.

이러한 이유로 빅데이터를 구성하는 요소를 크기(Volume), 속도(Velocity), 다양성(Variety)으로 규정하고 있다. 또한 빅데이터의 목적은 가치(Value)있는 정보를 제공하기 위한 것이기 때문에, 최근에는 데이터의 절대적인 크기(Volume)가 아닌 정보 가치의 크기를 빅데이터의 주요 요소로 보기도 한다.

문화관광 분야에서의 빅데이터 필요성

문화관광 분야는 전통적인 방법으로 통계를 생산하여 왔다. 문화와 관광은 개인의 인식에 따라 인지하는 정도가 다르기 때문에, 외형적 행동 등을 가지고 직관적으로 구분하는 것이 어렵다. 예를 들어 국민의 여가시간을 파악한다고 할 때, 여가시간을 보내고 있는지는 본인만이 판단할 수 있다. 핸드폰으로 문자를 보내는 경우, 친구와 대화를 한다면 여가로 볼 수 있지만, 직장 또는 업무 담당자와 업무에 대한 문자를 주고받는다면 여가로 볼 수 없기 때문이다.

관광도 일상생활지역을 벗어나 다른 지역에서 관광활동이나 휴식활동을 하는 경우를 의미하며, 여행은 일상생활지역을 벗어나 다른 지역으로 관광, 출장, 친지방문 등을 의미한다. 이러한 경우, 일상생활 지역은 사람마다 다를 수 있는데, 집이 서울이고, 직장이 나주에 있어 광주에 전세를 살고 있다면 서울과 나주, 광주 모두 일상생활지역에 해당된다. 출장지역의 경우에도, 특정지역을 지속적으로 갈 경우 이 지역은 일상생활지역으로 볼 수 있다. 따라서 문화관광 분야는 빅데이터(행정자료 포함)로 통계를 산출하는 것이 매우 어렵다.

그러나 조사자료는(survey data)는 기준년도에 비해 공표되는 시기의 차이가 발생하기 때문에 시의성이 떨어져 중요하고 시급한 이슈가 발생했을 때 판단근거로 삼기에는 어려움이 있다. 과거의 메르스 시기에서나, 최근의 코로나19 등과 같이 문화생활을 영위하거나 관광활동을 저해하는 상황이 발생하

1) 박근화 외5인(2018), 문화·체육·관광 데이터 연계를 통한 빅데이터 생산 및 활용방안 연구



였을 경우 빠르게 현황을 파악하고, 상황을 진단한 후, 문화관광 정책을 시행할 수 있는 기초자료가 필요하지만 조사자료는 시의성이 시급한 경우에 활용할 자료가 없는 경우가 있어, 정책 활용에 한계가 있다.

이러한 이유로 문화관광분야에서도 최근 공공의 행정자료와 민간의 행정자료인 빅데이터²⁾를 활용하는 경우가 증가하고 있는데, 코로나19 이후에는 더욱 세밀하게 빅데이터를 분석하여 활용하고자 하는 경향이 증가하고 있다. 특히, 코로나19로 인해 개인이 관광 목적인 해외여행뿐만 아니라 국내여행도 감소됨에 따라 관광산업 역시 침체되어, 이에 대한 상황을 빠르게 파악할 수 있는 정보(또는 데이터)가 필요하다. 시급한 상황에서는 가능한 주기가 짧으며 지속적으로 생산되는 데이터가 필요한데, 빅데이터가 그러한 특성이 있어 활용하기에 적합하다.

2) 본 원고에서는 빅데이터를 민간과 공공에서 생산되는 행정자료로 정의



문화관광 분야에서 주로 활용되는 빅데이터

현재 문화관광에서 주로 사용하는 빅데이터는 신용카드데이터와 이동통신데이터이다. 지속적으로 생산되고 있으며, 기존에 생산되고 있지 않은 정보를 제공하고 있다는 점이 이들 데이터를 활용하는 주된 이유라 할 수 있다. 이들 데이터는 생산의 지속성 때문에, 데이터 가공 주기를 시간, 일, 주, 월 단위 등으로 정해 통계를 생산할 수 있다. 따라서 필요한 시기와 기준을 지닌 데이터를 실시간(real time)으로 활용하여 통계를 생산할 수 있으므로 시의성이 매우 좋은 통계자료 제공이 가능하다. 이 중에서 신용카드데이터는 지출액을 다양한 업종이나 품목으로 제공하기 때문에, 문화 또는 관광 소비 분석에서 실제 소비가 잘되고 있는 업종(품목)과 그렇지 않은 업종(품목)을 구분할 수 있다. 시간의 흐름에 따라 각 업종(품목)의 지출액 변화를 실시간으로 파악하여 현황을 진단할 수 있어 활용성은 매우 높다고 할 수 있다.

이동통신데이터는 관광분야 또는 특정 시설의 방문객 등을 파악하는데 주로 이용하고 있다. 이동통신 데이터는 출발지(Origin)와 도착지(Destination)를 파악할 수 있는 데이터이기 때문에, 이동에 대한 경로를 파악하기에 매우 적합한 데이터이다. 관광에서는 여행경로를 분석하고자 하는 요구가 지속적으로 있어왔는데, 이러한 여행경로는 기존 조사통계를 이용하여 파악하기에 한계가 있는 것이 사실이다. 출발지와 도착지의 관계를 통해 여행경로 분석을 실시하여 지자체별로 관광객 유치 등의 관광정책에 활용할 수 있도록 OD(출발-도착)통계를 생산할 수 있다.

또한 특정관광지, 축제장소, 문화시설 등에 얼마나 많은 인원이 오는지를 파악하는데도 이동통신데이터는 유용하게 사용된다. 이동통신데이터는 휴대폰을 이용하여 측정하기 때문에, 재방문, 이용시간, 어느 지역에서 왔는지 등을 파악할 수 있어, 방문객들의 특성을 분석하거나 시설 활용도 등의 계획을 세울 경우에도 유용하게 사용된다. 방문객들이 어느 지역에서 오고 얼마나 시설에 머물다 가는지, 재방문하는 비율은 어느 정도 되는지 등을 파악할 수 있다면 마케팅 등에 활용하기 매우 좋은 정보가 될 것이다.

이 외에 문화와 관광분야에서 많이 사용되는 빅데이터를 살펴보면, 법무부 출입국·외국인정책본부에서 작성하는 출입국자 현황자료, 영화진흥위원회에서 제공하는 영화관입장권통합전산망의 영화관련 데이터, 예술경영지원센터의 공연예술통합전산망의 공연 관련 데이터, 한국공항공사와 인천국제공항공사에서 제공하는 항공통계 등은 문화와 관광에서 자주 사용하게 되는 데이터이다. 관광에서는 날씨 데이터도 중요하게 사용되는 자료이다

문화관광 분야에서 빅데이터 활용

관광분야에서 신용카드데이터를 처음 분석한 것은 2013년도의 '외국인 신용카드 국내지출액 현황'인데, 현재까지 한국문화관광연구원에서 반기별로 생산하여 문화센터(<http://stat.mcst.go.kr/>)를 통해 제공하고 있다. 외국인이 국내에서 사용한 신용카드(VISA, MASTER, AMERICAN EXPRESS, JCB, 은련³⁾)사의 지출자료를 가공하여 지출액 현황을 작성하고 있다⁴⁾. 통계작성을 위해 대륙과 주요국가를 우리나라에 많이 오는 국가 위주로 조정하였으며, 업종은 쇼핑, 숙박, 음식, 의료, 교통, 체험, 스포츠 등으로 구분하였다.

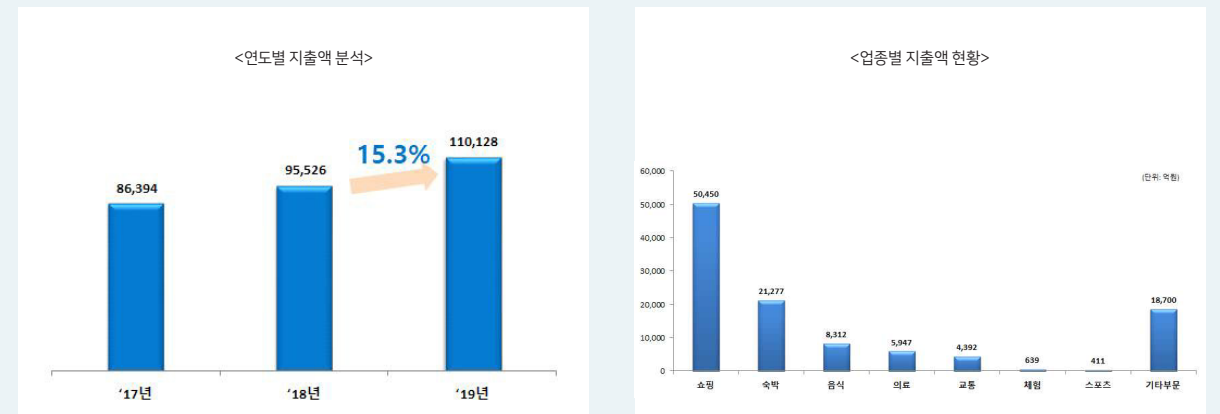
[그림 1]에는 2020년도에 작성한 2019년도의 외국인 지출액 현황을 제시하고 있는데⁵⁾, 이 자료를 통하여, 국가별 지출액 현황과 국내에서 많이 지출하는 장소와 품목 등을 파악할 수 있었으며, 관광목적으로 입국한 외국인과 사업(business)목적으로 입국한 외국인의 지출패턴 등을 분석할 수 있었다.

3) 은련(银联)카드는 세부 지출자료를 매입하지 않아 중국인의 다른 카드사의 업종별 지출비중으로 총 지출액 대비 세부 지출 자료를 추정하여 제공

4) 신한카드사에서 매입하여 제공하는 데이터 활용

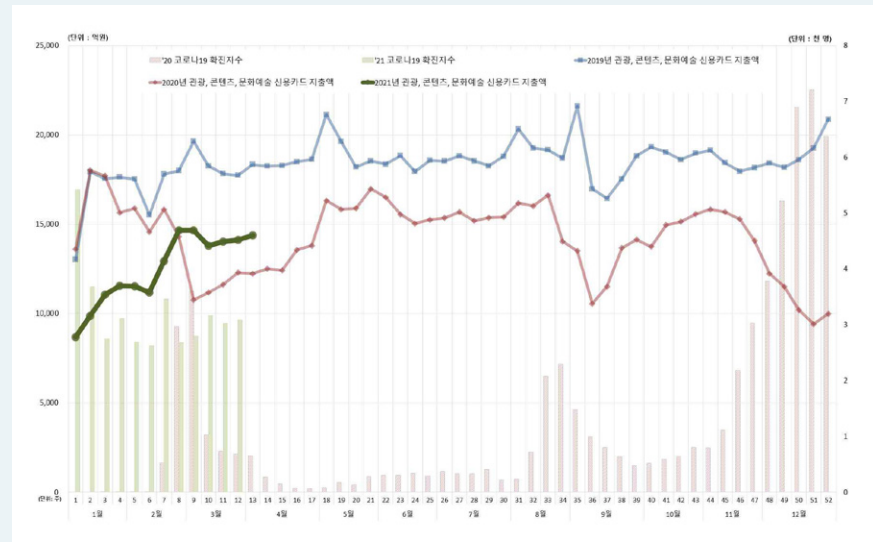
5) 한국문화관광연구원(2020), 연간 외국인 신용카드 지출액 현황 분석 보고서

[그림 1] 외국인의 국내 지출액과 업종별 지출액 현황



[그림 2]는 코로나19로 인하여 관광·콘텐츠·문화예술 분야에서 지출액의 변화를 파악하고자 만든 지표이며, 주간단위로 통계를 산출하고 있어, 주단위로 변화를 파악할 수 있다. 2019년도와 2020년도의 지출액을 같이 비교할 수 있도록 작성하고 있기 때문에, 2021년도의 지출액을 코로나 이전과 비교해서 어느 정도 회복되었는지 아니면 더 심각해졌는지 등을 파악할 수 있도록 분석한 것이다. 이를 살펴보면 2021년 3월 현재 관광·콘텐츠·문화예술 분야는 최근 지출액이 증가하고는 있지만, 아직 코로나 이전에 비해 여전히 소비가 감소한 상태로 이뤄지고 있음을 알 수 있다.

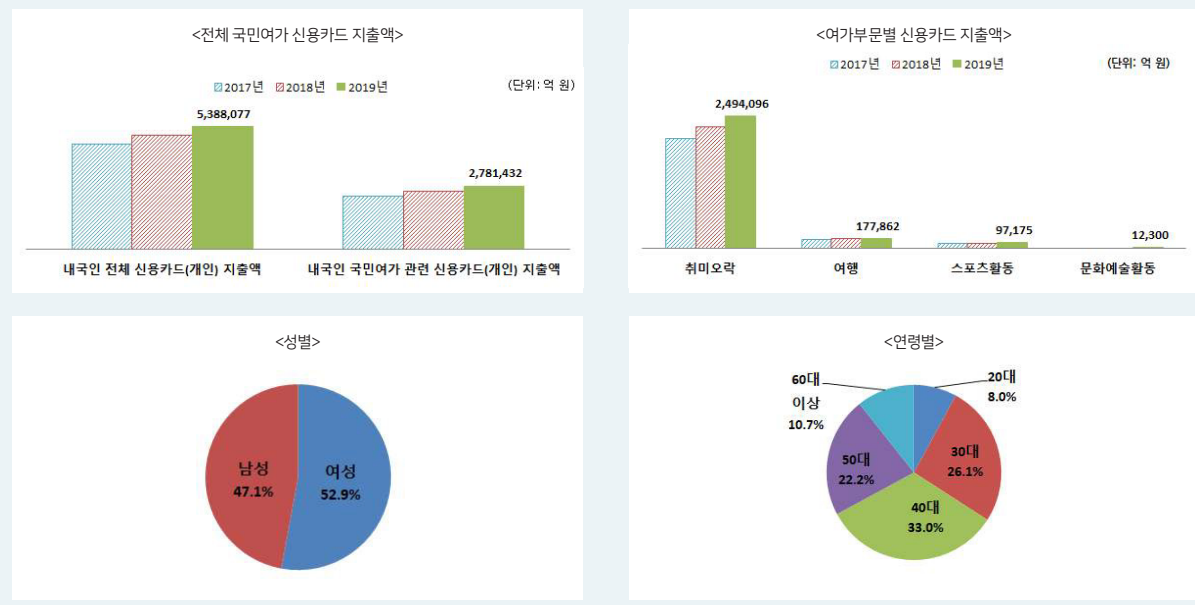
[그림 2] '21년 3월 관광·콘텐츠·문화예술 관련 신용카드 지출액 추이 분석⁶⁾



6) 한국문화관광연구원 (<https://www.kcti.re.kr>), 2021년 4월호 코로나19의 문화관광콘텐츠영향

[그림 3]은 우리나라 국민이 신용카드로 사용한 여가 관련 지출을 분석한 것인데, 신한카드사에서 제공한 데이터를 가공 분석하여 활용하고 있다. 카드사의 결과를 한국은행의 지급결제통계를 기준으로 반복비례가중법을 활용하여 세부 지출액을 추정하여 사용하고 있다. 여기서는 실제 여가활동을 하였는지는 알 수 없지만, 여가와 관련된 업종 또는 품목이라면 여가로 구분하여 과거와의 추이를 제시하여 변화를 분석하고 있다. 신용카드데이터는 카드이용자들의 정보를 가지고 있기 때문에, 인구통계학적인 측면에서 다양한 차원(dimension)으로 분석이 가능하다.

[그림 3] 2019년 국민여가 관련 신용카드 지출액 현황⁷⁾

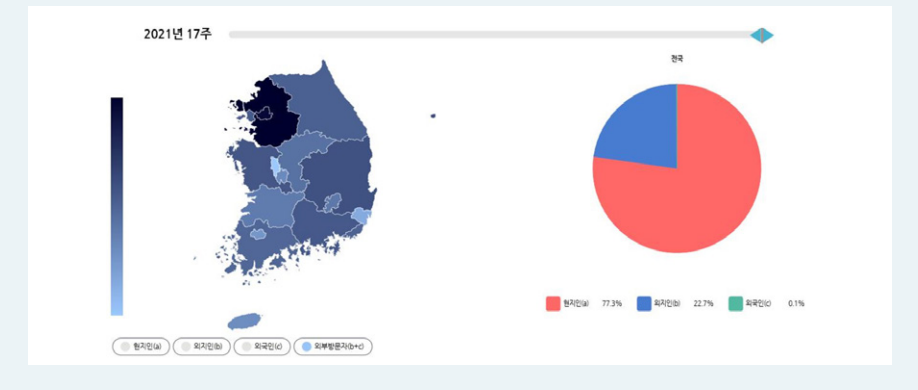


7) 한국문화관광연구원(2020), 2019 국민여가 관련 신용카드 지출액 현황

[그림 4]는 이동통신데이터를 이용하여 관광정보를 파악하고자 하는 것으로 광역지자체 방문자수를 주별 전국단위로 생산한 현황을 나타낸 것이다. 상단의 스크롤(scroll)을 움직이면 이용자가 지정한 기간동안의 주별, 월별로 누적된 방문자수의 현황을 제공한다. 지역별로 방문자수가 많으면 진하게 나타나고, 방문자수가 적으면 흐리게 나타나도록 하였다. 현지인과 외국인, 외국인을 구분하였으며, 해당기간의 현지인과 외국인의 비중이 나타나도록 하였다. 이동통신데이터 역시 가입자 등의 정보를 사전에 가지고 있으며, 방문한 지역의 기지국을 통해 파악이 가능하기 때문에, 주로 머무르는 장소와 일시적으로 머무르는 장소 등을 구분하여 분석할 수 있다. 따라서 이용자들의 특성 분석과 새로운 지역의 방문 등을 분석할 수 있다.

8) 한국관광공사에서 운영하는 한국관광 데이터랩 (<https://datalab.visitkorea.or.kr>)에서 인용

[그림 4] 광역지자체 방문자수⁸⁾

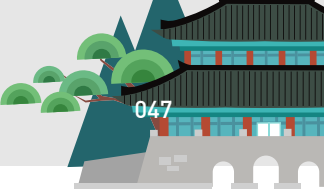


[그림 5]는 영화관입장권통합전산망(<http://www.kobis.or.kr>)에서 취합하여 한주간의 매출액과 관객수를 제시한 지표를 그래프로 나타낸 것이다. 전년도와 당해년도를 비교하고, 전주와 이번주를 비교하도록 구성하였다. 영화 현황에 대해 전체, 국내, 해외 등으로 이용자가 구분하여 살펴볼 수 있도록 하였다.

9) 한국문화관광연구원의 주별 주요지표에서 인용

[그림 5] 한 주간의 영화 매출액과 관객수 현황⁹⁾





빅데이터 활용의 한계와 유의점

지금까지 살펴봤듯이 빅데이터는 전통적인 통계에 비해 생성주기가 짧고, 실시간으로 정보를 파악할 수 있는 장점이 있다. 따라서 사회·경제적 변화를 쉽게 파악할 수 있으므로, 현재의 상황을 점검하는데 매우 유용하다. 이러한 이유로, 빅데이터를 통계로 만들어 활용하고자 하는 수요가 증가하고 있다.

그러나 많은 빅데이터는 실제 관광과 문화의 통계(또는 지표)를 생산하기 위해 마련된 데이터가 아니고, 신용카드 지출액 또는 모바일을 가지고 이동하면서 기지국과 교신하는 결과가 자동적으로 쌓여서 생기는 데이터이다. 따라서 값 그대로 이용하면 관광과 문화의 정보를 얻기 어렵기 때문에, 관광과 문화로 이용할 수 있도록 조작적 정의를 통해 가공이 이뤄져야 제대로 활용할 수 있다. 예를 들면, 이동한 결과는 여행이 아니다. 여행은 일상지역을 벗어나서 다른 지역에서 이뤄지는 행위이기 때문에, 일상지역을 정하는 규칙을 마련해야 한다. 밤에 머무르는 곳은 대부분 집이기 때문에, 밤 상주지역을 정하는 규칙을 마련해야 하며, 낮에 머무르는 곳은 회사 또는 학교 등이기 때문에 낮 상주지역도 정하는 규칙을 마련해야 한다. 이렇게 규칙을 정한 경우를 살펴보면 '한달에 밤과 낮 3시간 이상 8회 이상 머무른 곳'을 일상거주지역으로 판단하기로 조작적 정의를 세웠다면, 향후 이 지역은 일상 거주지역으로 판단하여 이 지역의 사람이 다른 지역으로 이동했을 경우에만 여행으로 판단하게 된다.

문화와 관광의 정보를 얻기 위해 조작적 정의를 하였다 하더라도, 통계로 활용하는 데는 한계가 있다. 실제 여가생활 또는 관광활동이 이동이나 지출행위 등으로 이뤄지는 것은 아니며, 본인의 인지 등이 중요한 요인이기 때문에, 조작적 정의를 잘 했다고 하더라도 정확한 통계를 산출했다고 말하기에는 어려움이 있다. 그러나 적절한 조작적 정의를 통해 지속적인 데이터의 변화를 살펴볼 수 있다면, 정확한 것과 상관없이 매우 중요한 정보로서 가치는 충분하다. 이러한 이유로 빅데이터는 통계로 생산하여 활용하는 것도 중요하지만, 살아있는(active) 정보 그 자체로 활용하는 것만으로도 충분한 가치를 지닌다.

이 외에도 빅데이터는 시간의 흐름에 따른 변화를 알 수 있는 정보를 제공하는 좋은 점이 있음에도 불구하고, 아쉬운점은 대부분 단일 변수만 생성된다는 점이다. 물론, 성별, 연령 등의 정보도 같이 제공되지만, 이들 변수들은 분석하는 단위로 사용되므로, 실제 단일 정보로 활용하는데 한계가 있다. 이와 달리 전통적 통계인 조사데이터(survey data)는 시의성의 한계와 데이터크기(data size)로 인해 공표범

위의 문제는 있지만, 한 번의 조사에서 다양한 정보를 같이 활용할 수 있도록 하는 장점이 있다. 따라서 어떤 행위나 사건(fact)에 대한 원인을 규명할 수 있는 것은 물론 향후의 변화까지 연결하여 분석할 수 있다. 빅데이터가 많은 정보를 줄 수 있다면 너무나 좋겠지만, 대부분이 행위나 결과(fact)만을 제공하고 있기 때문에 그 원인을 파악하는 것이 또 하나의 중요한 과제가 되며, 원인을 파악하였을 때, 새로운 business rule로 활용할 수 있게 된다.

빅데이터는 대부분 자동적으로 데이터가 생산된다. 신용카드데이터의 경우 과거에도 현재에도 미래에도 소비행위가 진행되면 카드결제가 이뤄지고, 카드결제가 이뤄지면 카드이용내역이 쌓이고, 이 결과들은 빅데이터로 나타날 것이다. 이러한 빅데이터는 살아있는(active) 데이터라고 할 수 있다. 따라서 데이터를 어떻게 보느냐에 따라서 값들이 달라진다. 조작적 정의에 의해서도 달라지고, 주기를 조정하여도 값이 변화한다. 이렇게 실시간으로 변화하는 빅데이터를 정지된(static) 데이터로 사용하는 경우가 많다. 빅데이터를 활용하는 대부분의 경우 특정 시점의 값이 얼마인지를 파악하는데 활용하고 있다. 이는 빅데이터의 가치를 제대로 활용하지 못하는 것이라고 할 수 있다. 빅데이터는 고정된 통계로 활용하는 것이 아닌, 능동적으로 변화하는 특성을 파악하여 분석하고, 예측 등에 활용될 때 제대로 된 가치를 보여준다.



문화와 관광분야에서 빅데이터의 활용방안

앞서 언급한 것처럼 빅데이터는 정확한 통계를 생산하기 보다 값들의 변화를 파악하고 분석하여 중요한 이슈로 제시하거나, 올바른 의사결정에 필요한 정보를 제공한다. 또한 어떤 행위의 사실(fact)을 하나의 값 또는 여러 개의 값으로 제공하기 때문에 그 원인을 파악하기 위한 또 다른 정보가 필요한 경우가 많다. 무엇보다 빅데이터의 가장 중요한 특성인 변화하는 데이터를 잘 이용할 수 있는 것이 중요하다.

따라서 제대로 된 빅데이터의 활용을 위해서는 서로 다른 데이터의 상호연계가 매우 중요하다. 데이터 3법(개인정보 보호법, 정보통신망법, 신용정보법) 개정을 통해 가명정보 등을 이용하여 서로 다른 데이터와 연계하는 것이 가능해진 만큼 이를 이용할 수 있는 방안을 마련하는 것이 중요하다. 그러나 빅데이터 간의 결합은 아직 어려움이 많으며, 특히, 조사데이터는 응답자들에게 개인정보 이용에 대한 동의를 받아야 되는 문제가 남아있다.

빅데이터간의 결합은 너무나 큰 데이터가 생성되기 때문에, 결합 후에도 분석 등과 같은 활용성에 큰 어려움이 있다. 개인정보 문제 등으로 인해 데이터베이스를 이용한 결합이 아닌 하나의 완전한 데이터로 결합해야 하기 때문에, 결합한 데이터의 용량은 기하급수적으로 증가하고, 결합 후에는 분석할 수 있도록 새롭게 데이터를 분해하는 등의 과정이 필요하다. 이러한 대용량 데이터를 저장할 시스템이나 기술(분석도구 등)을 보유한 기관은 많지 않아 활용에 어려움이 크다.

본 원고에서는 빅데이터와 조사데이터간 결합의 중요성을 강조하고자 한다. 조사데이터와 빅데이터

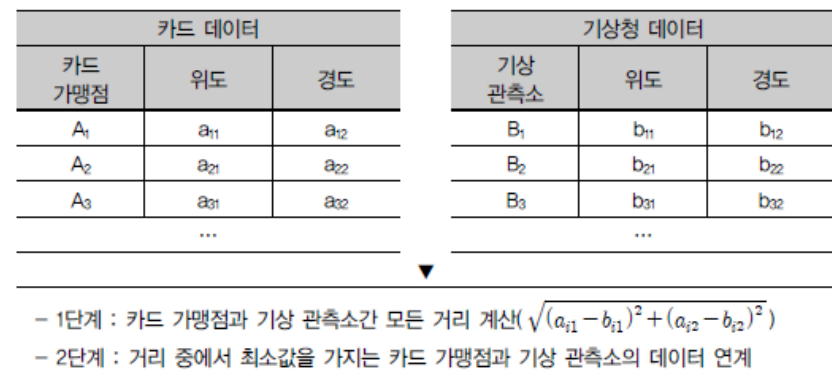


의 결합 또는 연계분석은 매우 활용성이 클 것으로 생각된다. 조사데이터는 예산과 시간 등의 표본데이터가 지니고 있는 한계로 소지역(small area) 통계 등의 세부적인 분석은 어렵지만, 구조화된 조사표의 조사항목 중 해당 통계의 값을 통해¹⁰⁾, 현황 및 실태에 대한 인과관계를 파악할 수 있다. 이에 반해 빅데이터는 실시간으로 변화하는 값을 파악하는데 용이 하지만, 결과의 인과관계 등을 파악하는데 어려움이 있다. 따라서 빅데이터와 조사데이터를 상호 연결할 수 있다면, 정확한 모집단 값의 파악이 가능하며, 결과에 대한 인과관계 분석도 가능할 것이다.

변화하는 데이터란 빅데이터 특성은 미래를 예측하거나, 세분화(segmentation)된 분석을 할 때 유용할 것으로 생각된다. 능동적인 데이터는 변화를 예측하고 개개인(또는 작은 집단)의 단위로 값들을 제공하고 해석할 수 있을 때 의미가 있다. 예를 들면 기업은 고객세분화를 통해 고객 개개인별로 맞춤형 마케팅 정보를 실시간으로 제공함으로써 큰 효과를 기대할 수 있다. 최근에는 인공지능(AI)기법을 적용하여 이러한 고객분석방법을 확대하고 있다.

문화관광분야에서는 직접적인 데이터 연계를 통해 다양한 분석을 하고자 하는 노력하고 있다. 박근화 외 5인(2018)의 연구진은 신용카드데이터와 기상청데이터를 연계하여 날씨정보에 따른 여행지 또는 관광지지의 소비지출 차이를 분석하는 연구를 수행하였다. [그림 6]은 연구방법을 제시한 것이며, 날씨에 따라 여행지 또는 관광지에서의 신용카드지출액 차이가 발생한다는 결과를 도출하였다. 즉, 날씨가 여행 또는 관광 관련 지출 활동에 영향을 미치고 있음을 제시하고 있다.

[그림 6] 신용카드데이터와 기상청 데이터의 연계방법¹¹⁾



11) 박근화 외5인(2018), 문화·체육·관광 데이터 연계를 통한 빅데이터 생산 및 활용방안 연구

다음은 국민여행조사 데이터와 빅데이터를 연계하여 분석하는 것을 시도한 사례인데, 직접적인 연계는 조사데이터의 응답자들에게 개인정보 이용 동의를 받지 못했기 때문에, 지역단위로 데이터를 결합 또는 연계하여 분석하였다. 국민여행조사는 월 4,200표본을 조사하고 있지만, 기초자치체인 시군단위의 통계는 생산하지 못하고 있다. 이러한 이유로 조사데이터를 정답으로 두고, 이동통신데이터와 인구통계 데이터 등을 연계하여 소지역 단위의 통계를 생산하였다. [그림 7]은 소지역 단위의 통계를 생산한 결과물로 시·도단위의 OD(출발-도착)통계를 생산한 것이다.



[그림 7] 시도단위의 여행회수의 출발-도착 통계¹²⁾

	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
서울	-	2,995,143	992,894	4,636,505	842,987	1,560,010	567,955	370,135	16,208,079	11,271,965	2,398,015	4,614,587	3,251,810	3,381,030	2,665,792	2,056,924	2,658,992
부산	995,227	-	724,554	242,569	145,509	323,524	1,891,440	55,356	584,659	791,304	392,299	425,582	676,894	1,444,576	3,232,999	8,224,057	868,531
대구	584,332	1,214,756	-	149,269	59,526	262,750	412,158	42,844	375,239	643,790	420,108	327,108	409,731	512,904	6,161,150	1,750,325	492,023
인천	2,559,115	644,511	191,626	-	195,154	382,876	121,674	103,603	4,178,305	2,818,662	785,005	1,990,701	953,173	1,043,562	729,109	530,560	779,302
광주	448,374	306,843	64,244	135,340	-	177,792	37,810	35,399	287,249	223,762	104,450	348,521	1,461,562	5,806,883	187,435	392,461	384,352
대전	742,691	468,631	195,667	232,737	117,427	-	90,413	585,486	587,591	647,254	1,210,025	3,184,577	1,116,125	578,909	609,493	484,162	344,997
울산	276,271	1,620,456	361,607	86,103	41,052	114,229	-	19,691	189,432	357,955	168,419	166,337	194,163	334,600	2,154,129	1,218,123	243,990
세종	176,553	87,516	37,442	52,819	32,701	440,007	17,990	-	139,058	129,082	226,077	489,228	160,066	109,224	103,634	81,106	68,638
경기	13,452,168	3,149,269	1,112,834	6,955,455	919,509	1,944,156	632,536	491,168	24,607,879	14,658,301	4,002,425	8,066,929	4,382,064	4,407,480	3,608,028	2,648,044	3,107,008
강원	830,148	228,884	99,280	289,915	41,674	145,727	56,577	30,543	1,003,342	4,401,341	399,459	254,459	179,522	220,301	380,768	179,477	223,369
충북	743,406	359,661	161,659	310,941	69,469	863,542	89,989	284,460	878,131	1,262,674	2,574,553	1,182,680	473,236	425,107	722,233	358,694	332,502
충남	1,054,375	456,026	174,277	545,021	130,835	1,406,056	120,233	349,767	1,301,565	905,145	774,366	4,867,493	1,057,435	734,373	551,457	457,118	374,918
전북	625,672	315,009	79,478	230,511	407,712	454,712	59,755	74,106	499,245	368,624	218,149	1,154,815	6,027,680	1,379,751	252,311	433,745	346,921
전남	520,685	489,909	85,086	204,375	1,743,157	191,296	86,297	37,913	396,782	265,039	127,534	365,459	1,055,474	5,890,801	261,812	616,988	380,240
경북	656,461	1,305,405	2,830,644	216,682	67,014	339,234	615,213	52,049	536,306	1,014,235	559,961	409,106	360,176	6,442,003	1,108,611	373,513	
경남	706,741	5,026,813	779,668	203,167	132,129	315,565	814,438	50,395	474,373	596,546	303,045	392,738	650,187	1,280,266	1,726,659	8,462,404	598,166
제주	238,126	172,684	40,914	69,081	46,563	37,801	21,608	7,647	117,410	80,159	38,223	49,975	75,099	153,982	64,080	73,168	971,328

12) 박근화 외(2020), 데이터 연계를 통한 소지역 통계 생산방안 연구

문화관광분야의 통계는 대부분 1년 단위로 조사하여 통계를 생산하고 있는데, 시의성 문제와 함께 표본수의 문제로 소지역(small area)단위 또는 소영역(small domain)단위의 통계는 생산하지 못하고 있다. 그리고 문화관광분야의 소비지출 관련 통계는 현재 매우 부족한 실정이다. 이러한 이유로 신용카드데이터와 이동통신데이터를 활용하고 있지만, 빅데이터의 중요한 특성인 변화를 제대로 활용하지 못하고 통계값으로만 이용하는데 그치고 있다. 향후에는 다양한 데이터의 상호연계와 가공 등을 통해 여러 유형의 정보를 활용할 수 있는 방안을 마련해야 할 것이다.

또한 문화관광 분야에서도 분석방법을 인공지능(AI)기법 등으로 확대 및 적용하여 예측과 세분화된 정보를 보다 정교하게 제공할 수 있도록 지속적인 연구가 필요할 것이다. 이러한 방법을 적용하기 위해서는 데이터 수집 및 분석 등과 같은 전문가들을 양성하고, 대용량 데이터를 분석할 수 있는 환경을 제공하는 것이 필요하다. 특히 문화관광분야는 IT기술이 아직 직접적으로 적용되지 않은 영역이라 할 수 있다. 따라서 신기술과 이를 활용할 수 있는 환경이 문화관광분야에도 조성되어야 할 것이다.



구자룡 벨류바인 대표

03

빅데이터 시대의
데이터 리터러시:
데이터를 다루는 습관

빅데이터 시대가 도래하면서 모두가 데이터 과학자가 되어야 하는가에 대한 의문이 높아지고 있다. 당연한 말이지만 모두가 데이터 과학자가 될 필요는 없다. 자동차가 우리의 일상에서 이동을 위한 필수품이 된 이후 모두가 운전을 하지만 자동차 전문가가 아닌 것과 같다. 데이터 과학자가 되지 않아도 자동차를 이용하듯이 데이터를 이용할 수 있으면 된다. 운전은 면허가 필요하지만 데이터를 다루는 일은 관심만 있으면 가능하다. 왜냐하면 우리는 현업에서 언제든지 데이터를 활용할 수 있는 상황에 노출되어 있기 때문이다. 자동차를 이용하듯이 그냥 이용하면 된다. 다만 자동차를 자기 마음대로 운전하면 사고가 나듯이 데이터도 자기 마음대로 사용하면 문제가 된다. 따라서 데이터를 제대로 다룰 수 있어야 하고, 분석할 수 있어야 하고, 시각화 할 수 있어야 한다. 그래야 데이터와 숫자에 속지 않고 제대로 활용할 수 있다. 데이터가 업무수행을 위한 필수 역량이 되고 있다.



🔍 데이터를 다루는 습관이 들어야 한다.

일반적으로 한 번 들인 습관은 쉽게 바뀌지 않는다. 그러나 제대로 습관이 들면 바꿀 필요가 없다. 데이터를 다루는 습관이 바로 여기에 해당한다. 현업에 있는 대다수의 사람들은 그 동안 데이터를 다룰 기회가 많지 않았다. 데이터의 수집 및 분석과 의사결정에 다양한 제약과 장벽이 있었기 때문에 전문가의 영역으로 치부되었다. 이 부분은 아직도 데이터과학자의 영역으로 일부 존재한다. 현업의 실무자가 그 장벽을 넘어야 할 이유는 없다. 현업에서 필요한 수준의 데이터는 이미 풍부한 상태이며, 스스로가 데이터에 접근할 수 있는 정도에서 가치를 만들기만 하면 된다. 다만 스스로가 깊은 고민을 하지 않고 막연하게 두려움을 가지고 있었기 때문에 풍부한 데이터를 활용하기 위한 고민을 하지 못했을 뿐이다. 어쩌면 지금이 데이터로 업무의 성과를 높이는 현명한 의사결정자로 탈바꿈할 수 있는 기회다.

데이터를 다루는 습관을 들일 때 현명한 의사결정권자가 되고, 우리 사회에서 요구하는 데이터 리터러시 역량을 갖춘 사람이 되는 것이다. 데이터 리터러시는 데이터를 읽고 쓰는 능력을 의미하는데, 이는 어떤 문제가 있다는 것을 인식하면 이 문제를 정의하고, 문제를 해결할 수 있는 데이터를 수집하고, 수집된 데이터를 분석하고 시각화하여, 문제해결 목적에 맞게 활용함으로써 새로운 가치를 만들어 내는 능력이다.

어느 순간 빅데이터가 넘쳐나면서 모두 빅데이터에서 보배를 찾으려고 한다. 하지만 빅데이터는 보배가 아니라 수많은 구슬에 불과하다. 사막에서 바늘을 찾는 격이다. 그런데 무슨 바늘인지 모르면서 사막으로 달려가는 것은 시간 낭비일 뿐이다. 데이터를 말하기 전에 먼저 우리가 가지고 있는 문제를 구체적으로 정의해야 할 이유이다. 데이터 기반의 의사결정을 위한 첫걸음은 바로 문제정의다. 그 문제를 해결하는데 필요한 데이터가 무엇인지, 우리가 그런 데이터를 가지고 있는지, 없다면 어디서 수집할 수 있는지를 고민해야 한다.

데이터를 확보했다면 데이터에 맞는, 그리고 문제해결에 적합한 데이터분석 방법을 찾아야 한다. 이런 과정이 결코 가벼운 과정은 아니다. 그러나 현업에 종사하는 많은 사람들은 현업의 특성을 잘 알고 있기 때문에 그렇게 어렵지 않게 접근이 가능하다. 데이터 분석 역시 전문적인 데이터 과학자가 아닌 이상 엑셀 정도의 도구로도 충분히 분석할 수 있다. 중요한 것은 문제해결을 위한 데이터 분석이기 때문에 빅데이터가 아니어도 괜찮다. 세계에서 가장 큰 규모의 빅데이터를 가지고 있는 구글이나 아마존조차 전체 빅데이터가 아니라 그 속에 있는 데이터의 일부만 추출하여 데이터 분석, 모델링, 예측, 통찰의 제공 등을 실행하기 위한 의사결정을 한다.

현업 차원에서 문제해결을 위해 분석 가능한 데이터를 내가 다룰 수 있는 분석도구로 활용할 수 있으면 된다. 현업 중심의 셀프서비스 데이터분석가가 되어야 하는 이유다. 중요한 것은 내가 어느 정도 데이터 분석을 할 수 있어야 한다는 점이다. 데이터 분석은 분석 도구와 통계에 대한 기본적인 이해를 바탕으로 가능하다. 이 부분을 스스로가 수행해 보면서 전체적인 프로세스와 의사결정 구조를 이해할 수 있다. 결과적으로 분석적 마인드가 형성되어야 데이터를 원활하게 다루고 활용할 수 있게 된다. 셀프서비스로 데이터를 분석하는 연습을 하고 또 연습하는 수고로움을 감내해야 리터러시 역량을 높일 수 있다.



데이터를 다루는 습관을 기르기 위해서는 어떤 문제에 직면했을 경우 가장 먼저 데이터 자체에서부터 시작할 것인가? 아니면 비즈니스 문제에서 시작할 것인가를 고민해야 한다. 데이터를 활용하는 목적은 비즈니스상의 문제나 이슈의 예방과 해결에 있기 때문에 비즈니스 문제에서 출발할 때 가장 큰 가치를 만들 수 있다.

데이터에 접근하는 관점과 프로세스(비즈니스 문제에서 출발)



🔍 나도 모르게 데이터에 속고 있다

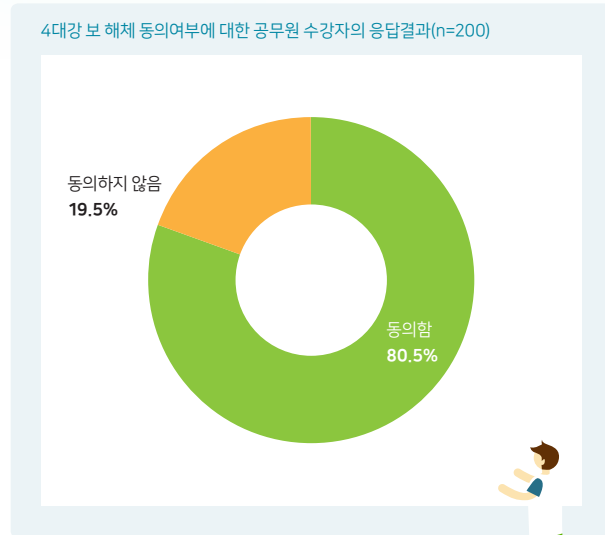
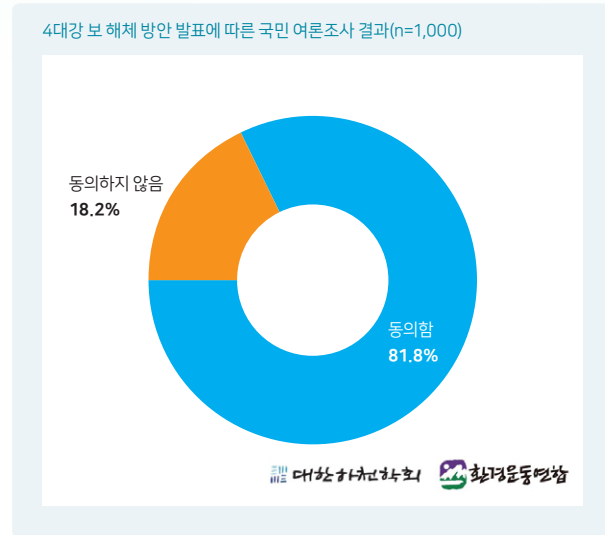
최근 들어 데이터 리터러시 강의요청이 많다. 빅데이터 시대에 데이터 기반으로 현명한 의사결정을 하고 싶은 기업이나 공공부문에서 데이터를 통한 성과를 만들기 위한 기본적인 욕구라고 생각한다. 수강자들은 대체로 고등교육을 받았고, 경력도 꽤 된다. 그렇지만 사회 초년생이든 경력이 높은 관계 없이 많은 사람들이 너무나 쉽게 데이터에 속고 있다. 경우에 따라서는 알고도 속는다. 물론 수강자만을 대상으로 한 경험하기에 선택편향이 존재한다는 점을 감안하더라도 너무나 잘 속고 있다는 점에서 깜짝 놀라기도 하고 걱정이 되기도 한다.

우선 독자 여러분도 아래의 질문에 한 번 응답해 보기를 권한다.

지난 2월 환경부에서는 보 가운데 3개를 해체하고, 2개의 보를 상시 개방하는 방안을 제안했습니다. 그 근거로 △보의 효용성이 부족하고, △보가 없어도 물이용에 어려움이 크지 않으며, △수질·생태계가 개선되고, △유지·관리 비용 절감 효과가 크다는 내용을 고려했다고 밝혔습니다. 귀하께서는 이 처리방안에 대해 어떻게 생각하십니까?
(1) 동의한다
(2) 동의하지 않는다

위의 질문은 정답이 있지는 않다. 그냥 여러분의 생각을 보기(1)과(2) 중에서 하나를 선택하면 된다. 온라인 설문으로 필수질문이기 때문에 보기 중에서 꼭 응답해야 다음 질문으로 넘어갈 수 있다. 이제 응답을 했다면, 여러분은 어떤 선택을 했을까?

이 조사는 2019년 5월에 발표된 “4대강 보 해체 방안 발표에 따른 국민 여론조사”다. 위의 질문은 이 조사에서 가장 중요한 질문의 하나였다. 전국의 19세 이상 성인 남녀를 대상으로 16개 광역시도별 성, 연령별 인구비례할당에 의해 1,000명이 응답했는데, ‘동의한다 81.8%’, ‘동의하지 않는다 18.2%’로 보고되었다.



동일한 질문을 온라인 실시간 설문조사 플랫폼인 슬라이도(sli.do)를 이용하여 2021년 상반기에 진행한 데이터 리터러시 강의(4곳의 인재개발원에서 총 7회)에서 200명의 수강자를 대상으로 조사해봤다. 응답자는 올해 신규로 임용된 공무원(5급, 7급, 9급)과 승진자 및 중앙부처 핵심인력이 포함되어 있다. 전체 200명의 응답자 중에서 '동의한다 80.5%', '동의하지 않는다 19.5%'로 나타났다. 5급 신규자의 경우 23명이 응답했는데(30세 전후의 5급 공채) 동의한다에 83%가 나와서 일반 국민보다 높게 나타났다.

더욱 고민스러운 것은 위 질문에 어떤 문제가 있는지 찾아보기 위해 온라인 협업프로그램인 패들렛(padlet)을 이용하여 하나의 작업공간에 의견을 자유롭게 제시할 수 있도록 했다. 이 질문의 문제점으로 응답자들은 '부정적인 측면에 대한 언급이 없어서 객관적인 판단이 어렵다', '긍정적인 대답을 유도한다', '응답자의 생각과 답변을 왜곡한다' 등의 의견을 제시했다. 질문에 문제가 있다는 점을 인식했는데도 불구하고 실제 응답에서는 유도한 방향으로 응답한 것이다. 아니 문제가 있다는 전제로 의견을 구하기 전까지는 문제가 있다는 점을 인식하지 못했을 수도 있다. 30세 전후의 5급 공채 신규자들로 정부부처의 핵심인재들이고 사회참여와 공정성에 대한 관심이 높은 MZ세대의 특성을 고려할 때 우려되는 상황이다. 다른 교육에서도 비슷한 결과가 되풀이 되었다. 즉, 유사한 형태로 속고 있다는 점을 확인할 수 있었다.

예시한 내용은 여론조사에서 하나의 질문만을 대상으로 살펴본 한계가 있다. 그렇지만 위 조사의 목적이 사회적 갈등이 심화된 특정의 사안에 대해 국민들의 여론을 수렴하고 그 결과에 따라 정책 결정을 하고자 했다는 점을 생각해야 한다. 그 중대성에 비해 누군가가 의도적으로 데이터를 왜곡하고자 마음을 먹는다면 너무나 쉽게 데이터로 국민을 속일 수 있다는 것을 확인할 수 있다. 공무원조차도 이렇게 쉽게 속는데 일반 국민은 어떻게 생각하면 등골이 오싹하다.

4대강 보 해체 동의여부 질문의 문제에 대한 5급 공채 공무원들의 의견

padlet

이 조사의 문제는 무엇일까?

4대강 보 해체 관련 조사의 문제점을 찾아보자.

구구름 20/11/13 10:23

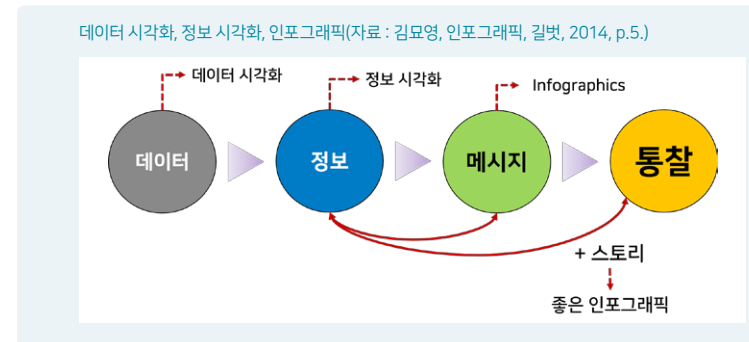
익명 21/02/08 02:46
보 해체의 부정적 측면에 대한 언급이 없어 해당 자료만 보고는 객관적인 판단이 어렵습니다.

익명 21/02/08 02:46
긍정적인 내용만 담아 답변을 유도하고 있는 것으로 보입니다.

익명 21/02/08 02:46
질문이 긍정적인 대답을 유도하네요.

익명 21/02/08 02:46
강의 가치, 생태계적 부분을 어필함으로써 응답자가 보의 방향성에 대한 생각이 왜곡될 수 있도록 구성되어 있으며, 이후 질문이 보 해체 방안과 다른 목적을 가진 부분에 대한 질문이 있다.(일관성X)

익명 21/02/08 02:46
답안을 유도하는 질문이므로 응답자의 생각과 답변을 왜곡시킬 수 있음



데이터 시각화로 통찰을 제공해야 한다

데이터를 수집 및 처리하여 시각적으로 도식화 하는 기법을 데이터 시각화라고 한다. 데이터 시각화는 왜 필요할까? 데이터 시각화는 아이디어를 탐구하고 정보를 빠르고 효과적으로 전달할 수 있기 때문이다. 따라서 보기 좋게 시각화하는 것이 아니라, 복잡한 데이터와 도표에 존재하는 이면을 현실적으로 볼 수 있게 만들어야 한다. 좋은 시각화는 보다 쉽게 차트를 보고 통찰할 수 있게 만든다. 가공되지 않은 데이터로부터 어떻게 통찰을 이끌어낼지 생각하고 또 생각해야 한다. 좌측 그림과 같이 데이터 시각화를 통해 정보를 도출하고, 정보 시각화를 통해 메시지를 발굴하고, 여기에 스토리를 더하여 인포그래픽으로 표현할 때 보다 쉽게 통찰할 수 있다. 즉, 데이터 시각화는 메시지를 제시해야 하고 통찰을 제공해야 한다.

최근에 특히 공공부문에서 인포그래픽을 통한 정책 설명과 홍보를 많이 하고 있다. 각 정부부처에서 수행하고 있는 정책이나 계몽 활동, 그리고 정책의 성과를 시각화로 표현하는 트렌드가 강하다. 대표적으로 2가지 사례를 비교해서 살펴보면, 어떤 인포그래픽이

좋은 인포그래픽인지 확인할 수 있다.

2020년은 그 누구도 예상하지 못했던 코로나19 팬데믹 상황이었다. 우리나라는 K-방역으로 세계적인 관심을 받았다. 그 중심에 마스크 착용이 있다. 과연 마스크를 착용하면 코로나19 예방효과가 있을까? 모두가 의심하던 시기에 인포그래픽 한 장으로 명쾌하게 설명하고 계도하여 모두가 마스크를 착용하는 것이 당연하다는 인식을 심어주었다. 2020년 8월경 '마스크 착용에 따른 코로나 19 전파'라는 인포그래픽이 언론과 지자체, 그리고 소셜미디어를 통해 공유되었다. 출처는 한국 질병관리본부(현 질병관리청)로 되어 있었다. 하지만, 정작 인포그래픽 자료의 출처는 질병관리본부(현 질병관리청)가 아니라고 공식 발표한 바 있다.



▲ 마스크 착용에 따른 코로나19 전파에 대한 각종 인포그래픽(각 기관 및 원본)

이 자료의 출처를 추적해보면, 미국치과협회의 온라인 세미나에서 한 발표자가 중환자 전담의사인 친구로부터 받은 자료라면서 마스크 착용으로 감염 확률을 1.5%까지 줄일 수 있다고 한 것이다. 국내-외에서 다양하게 인용되었는데, 특히 국내에서는 질병관리본부에서 발표한 것으로 둔갑하여 여러 기관(특히 지자체)에서 재인용되었다. 여전히 지금도 사실인양 공유되고 있다. 당시 정은경 질병관리본부장(현 질병관리청장)은 "마스크를 쓰지 않으면, 마스크를 쓴 사람보다 감염 가능성이 5배가 높다. 턱에만 걸치는 것은 소용이 없다"며 마스크 착용의 중요성을 강조한 바 있는데 이 말이 공식 입장이었다.

그렇다면 사실이 아님에도 사실인양 널리 공유된 이유는 무엇일까? 메시지가 분명한 인포그래픽이기 때문이다. 출처와 근거가 불명확했지만 마스크를 착용해야 한다는 인식을 심어주는 데는 더 없이 좋은 자료가 되었기 때문이다. 이 사례에서 알 수 있는 것은 명확한 메시지에 적절한 스토리가 입혀지면 그것을 보는 사람들이 쉽게 통찰을 얻을 수 있다는 점이다. 그러나 원래의 데이터가 잘못되었거나 근거가 부족한 내용은 더 큰 오류를 만들 수도 있다는 점을 간과하면 안 된다. 이 사례에서 출처와 근거가 명확했다면 최고의 인포그래픽으로 남았을 텐데 하는 아쉬움이 있다.

그럼에도 불구하고 마스크 착용의 예방효과는 분명한 것으로 검증되고 있다. 국제학술지 랜싯(The Lancet)에 게재된 '코로나19의 사람 간 전파 예방을 위한 물리적 거리 두기, 마스크 및 눈 보호구' 논문에 따르면, 비감염자가 N95 등 마스크를 쓴 채 확진자에게 노출될 경우 마스크를 쓰지 않았을 때보다 감염 위험이 85%나 감소한 것으로 나타났다. 정부에서도 마스크 착용을 적극 권장하고 있으며 의무화까지 했다.



다른 사례를 보자. 아래 제시한 '2021년 달라지는 세법'과 '흡연자에게 더 위험한 코로나19' 인포그래픽은 과도한 정보 제공으로 메시지가 무엇인지 확인하기 어렵고, 무엇을 해야 하는지에 대한 통찰을 주지 못하고 있다. 핵심을 놓치고 있지 않나 의심되는 인포그래픽이다. 인포그래픽의 표적 고객이 스스로 찾고 싶은 메시지를 일목요연하게 제시해야 통찰이 일어난다. 한편, '국민연금 A TO Z' 인포그래픽을 보면, 한 눈에 국민연금의 현재 상황을 파악할 수 있고 내가 앞으로 수령할 수 있는 금액을 추정할 수 있다. 대부분의 직장인은 60세 전후 정년퇴직을 하는데 이후 연금 수령액만으로 편안한 노후생활이 어렵다는 것을 통찰할 수 있다. 은퇴 후의 소득활동을 어떻게 할지 고민해야 한다는 것을 말해주고 있다. 좋은 시각화는 구구절절이 설명하는 것이 아니라 숫자와 그래프로 통찰을 제공해 줄 수 있어야 한다.

다만, '2021년 달라지는 세법'과 '흡연자에게 더 위험한 코로나19' 인포그래픽은 과도한 정보 제공으로 메시지가 무엇인지 확인하기 어렵고, 무엇을 해야 하는지에 대한 통찰을 주지 못하고 있다. 핵심을 놓치고 있지 않나 의심되는 인포그래픽이다. 인포그래픽의 표적 고객이 스스로 찾고 싶은 메시지를 일목요연하게 제시해야 통찰이 일어난다. 한편, '국민연금 A TO Z' 인포그래픽을 보면, 한 눈에 국민연금의 현재 상황을 파악할 수 있고 내가 앞으로 수령할 수 있는 금액을 추정할 수 있다. 대부분의 직장인은 60세 전후 정년퇴직을 하는데 이후 연금 수령액만으로 편안한 노후생활이 어렵다는 것을 통찰할 수 있다. 은퇴 후의 소득활동을 어떻게 할지 고민해야 한다는 것을 말해주고 있다. 좋은 시각화는 구구절절이 설명하는 것이 아니라 숫자와 그래프로 통찰을 제공해 줄 수 있어야 한다.

다만, '2021년 달라지는 세법'과 '흡연자에게 더 위험한 코로나19' 인포그래픽은 과도한 정보 제공으로 메시지가 무엇인지 확인하기 어렵고, 무엇을 해야 하는지에 대한 통찰을 주지 못하고 있다. 핵심을 놓치고 있지 않나 의심되는 인포그래픽이다. 인포그래픽의 표적 고객이 스스로 찾고 싶은 메시지를 일목요연하게 제시해야 통찰이 일어난다. 한편, '국민연금 A TO Z' 인포그래픽을 보면, 한 눈에 국민연금의 현재 상황을 파악할 수 있고 내가 앞으로 수령할 수 있는 금액을 추정할 수 있다. 대부분의 직장인은 60세 전후 정년퇴직을 하는데 이후 연금 수령액만으로 편안한 노후생활이 어렵다는 것을 통찰할 수 있다. 은퇴 후의 소득활동을 어떻게 할지 고민해야 한다는 것을 말해주고 있다. 좋은 시각화는 구구절절이 설명하는 것이 아니라 숫자와 그래프로 통찰을 제공해 줄 수 있어야 한다.



▲ 2021년 달라지는 세법과 흡연자에게 더 위험한 코로나19, 그리고 국민연금 A TO Z

어떻게 해야 데이터와 시각화에 속지 않을까?

어떤 경우에 데이터나 시각화가 거짓말을 할까? 다시 말해 차트를 만드는 사람은 어떤 경우에 거짓말을 하고 싶어질까? 데이터 시각화의 세계적 권위자인 알베르토 카이로 교수는 <숫자는 거짓말을 한다>라는 책에서 여섯 가지 이유를 제시했다. “디자인이 잘못되었을 때, 잘못된 데이터를 사용할 때, 표시된 데이터의 양이 너무 많거나 적을 때, 불확실성을 숨기거나 헛갈리게 할 때, 잘못된 패턴을 제시할 때, 그리고 사람들의 기대나 편견에 영합할 때”라고 했다. 데이터는 올바르게만 그래프가 틀린 경우, 그래프는 멋지지만 데이터가 틀린 경우를 흔히 보게 된다. 앞에서 살펴본 사례를 포함하여 잘못된 시각화의 많은 부분들이 바로 이런 경우에 해당한다. 아울러 사람들의 기대나 편견에 영합하는 경우를 특히 조심해야 한다. 차트 작성자가 주관적인 판단으로 좋은(?) 차트를 만들어 의사결정권자에게

잘 보이고 싶은 유혹에 빠질 때, 그 의사결정권자 뿐만 아니라 그 의사결정으로 영향을 받는 많은 국민과 소비자들은 고통을 받을 수도 있다. 차트를 볼 때 한 번 더 눈을 부릅뜨고 작성자가 어떤 의도를 가지고 작성했는지 먼저 살펴야 한다.

카이로 교수는 거짓말하는 차트에 속지 않으려면, 차트를 볼 때 다섯 가지를 유의하라고 한다. 첫째, 제목, 설명, 출처를 먼저 읽어라. 둘째, 측정 대상, 단위, 척도, 범례 등을 확인하라. 셋째, 시각적 부호화(원의 크기 등)의 의미를 파악하라. 넷째, 주석(핵심이나 요점을 강조하기 위한 짧은 설명 등)을 읽어라. 마지막으로 폭넓은 시야로 패턴과 동향, 그리고 변수들의 관계를 파악하라. 거짓말하는 차트에 속지 않기 위한 유의사항을 명심하면서 혹시 의심이 든다면 꼭 원본 자료를 찾아보아야 한다. 언론에서 다루었다고, 정부기관에서 제시했다고 무조건 신뢰하면 오류에 빠질 수 있다. 어쩌면 언론과 정부가 데이터와 차트로 독자와 국민을 가장 속이고 싶어 할지도 모른다. 영향력과 이해관계가 있기 때문이다. 스스로 속지 않기 위한 데이터 리터러시가 필요한 순간이다.

데이터를 분석하고 시각화를 하는 이유는 개인적 경험에 의한 의사결정보다 증거에 기반을 둔 의사결정이 보다 현명하다고 느끼기 때문이다. 데이터를 목적에 맞게 활용하는 데는 데이터 수집도, 데이터 분석도, 그리고 데이터 시각화도 중요하다. 데이터를 다루는 능력이 필수적으로 요구되는 사회에서 리터러시를 높이기 위한 노력은 단순한 개인의 능력이 아니라 사회적 정보격차의 해소차원에서 조직적으로 역량 강화를 위한 노력이 선행되어야 한다. 특히 디지털 전환이 이루어지는 사회 환경에서 데이터는 새로운 원유이고 미래의 자원이 되기 때문에 데이터를 기반으로 의사결정을 하는 개인, 조직, 사회, 국가의 문화가 형성되어야 한다. 나만 데이터 리터러시 역량이 뛰어나다고 되는 것이 아니다. 사회 전체적으로 데이터 기반의 의사결정 문화가 만들어질 때 디지털 사회의 새로운 가치가 창출된다.





동형 암호화의 개념 및 현황

■ 송영수 ■

■ 서울대학교 수리과학부 교수 ■

최근 데이터 기반 연구 및 산업이 발전함에 따라 개인정보 보호 이슈도 함께 떠오르고 있습니다. 예를 들어 IT 기업들은 다수의 고객 데이터를 이용하여 서비스를 구축하거나 개인 사용자에게 맞춤형 서비스를 제공하는 과정에서 개인정보의 수집과 분석이 일어나게 됩니다. 문제는 이 과정에서 민감한 개인정보가 노출될 여지가 크다는 데 있습니다.

개인정보를 보호하기 위해 데이터에 대한 접근을 막을 수 있지만 이는 데이터의 사용을 원천적으로 차단하게 되어 올바른 해답이 아닙니다. '프라이버시 보호'의 목적은 데이터의 유용성과 기능을 보존하면서 데이터 제공자들이 의도하지 않은 방향의 정보 오남용을 방지하는 것입니다.

데이터 프라이버시를 달성하려는 노력은 정책수립, 기술적 연구 등 다양한 방향으로 지속되어 왔습니다. 대표적으로 미국의 HIPAA, EU GDPR과 같은 개인정보 보호에 대한 규정들이 있으며 국내에서는 최근 데이터 3법을 개정하는 등 개인정보와 데이터 기반 과학 및 산업발전 두 마리 토끼를 잡으려는 노력을 하고 있습니다. 대표적인 프라이버시 기술로 고전적으로는 익명화, 매스킹, 랜덤화, 범주화 등 데이터의 정보 일부를 삭제하거나 변형하는 방법들이 주로 쓰였는데 프라이버시 보호를 위해 손실되는 정보량으로 인해 활용성이 저하된다는 고질적인 단점이 있었습니다.

다행히도 최근 학계에서는 프라이버시 보존에 사용가능한 시큐어컴퓨팅 기술의 급격한 발전이 이루어지고 있습니다. 기존의 암호 기술이 단순히 암호화, 디지털 서명 등을 이용하여 데이터 보호와 신원확인에 초점을 맞췄다면 최근에는 이를 넘어서서 데이터 결합과 계산과정을 보호하는 등 추가적인 기능을 제공하는 다양한 기반기술들이 등장하고 있습니다. 현재 가장 주목받는 시큐어컴퓨팅 기술로는 차등정보보호(Differential Privacy), 영지식 증명(Zero-Knowledge Proof), 다자간 연산(Multi-Party Computation), 그리고 오늘의 주제인 동형암호(Homomorphic Encryption) 등이 있습니다.

동형암호란?

동형사상(Homomorphism)이란 대수학 분야에서 사용하는 용어로 두 집합 사이의 대수적 구조를 보존하는 함수들을 지칭하는 말입니다. 다시 말해서 동형사상이 주어진다면 <연산을 수행한 뒤에 함수 값을 계산>하는 경우와 <함수를 먼저 계산하고 연산을 수행>하는 두 가지 방법이 동일한 결과를 도출하게 됩니다(그림 1).

그림 1. 동형사상의 성질

$$\begin{array}{ccc}
 x_1, x_2 \in X & \xrightarrow{F} & F(x_1), F(x_2) \in Y \\
 + \downarrow & & + \downarrow \\
 x_1 + x_2 \in X & \xrightarrow{F} & F(x_1) + F(x_2) \\
 & & = F(x_1 + x_2) \in Y
 \end{array}$$

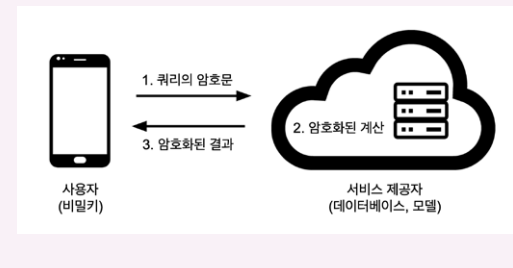
그렇다면 동형암호란 무엇일까요? 동형암호란 간단히 말해서 암호화 과정이 동형사상인 암호시스템입니다. 동형암호에서는 우리가 평문(plaintext)에서 계산을 진행하듯이 암호문(ciphertext) 위에서 특정한 연산들을 수행할 수 있으며 그 결과로 얻어진 암호문을 비밀키 소유자가 복호화하면 마치 평문 위에서 같은 계산을 진행한 것과 동일한 결과를 얻을 수 있습니다(그림 2).

그림 2. 동형암호 개요

$$\begin{array}{ccc}
 m_1, m_2 \in M & \xrightleftharpoons[Dec]{Enc} & E(m_1), E(m_2) \in C \\
 + \downarrow & & + \downarrow \\
 m_1 + m_2 \in M & \xrightleftharpoons[Dec]{Enc} & E(m_1) + E(m_2) \\
 & & = E(m_1 + m_2) \in C
 \end{array}$$

동형암호는 처음에 언급했던 데이터 프라이버시 문제를 해결할 수 있는 가장 전도유망한 기술 중 하나로 평가되고 있습니다. 대표적인 용례로 IT 기업들이 제공하는 맞춤형 서비스에서의 개인정보 보호를 꼽을 수 있는데 사용자가 서비스 제공을 위해 보내야 하는 쿼리(query)에 민감한 정보가 포함되어 있다면 이를 암호화하여 전송하고 서비스 제공자는 암호화된 쿼리 상에서 계산을 진행하는 방법을 통해 개인정보의 유출을 막을 수 있습니다(그림 3). 또한 정부, 금융기관 등 다수의 기관에서 산발적으로 보유하고 있지만 공유가 금지되어 있는 데이터를 안전하게 취합하고 분석하여 그 활용성을 높이는 데에도 사용할 수 있습니다. 특히 동형암호를 통계데이터 분석에 적용한다면 데이터베이스 연계를 통해 더 다양하고 정확한 통계지표를 제공하는 국가통계 분석시스템 구축에 큰 역할을 할 것입니다.

그림 3. 동형암호를 이용한 프라이버시 보존 개인맞춤형 서비스



동형암호의 종류

동형암호는 보통 지원하는 연산을 기준으로 몇 종류로 나뉩니다. 일반적으로 동형암호는 덧셈과 곱셈을 모두 지원하는 암호시스템을 지칭하지만 이 중 하나의 연산만 지원하는 스킴은 부분동형암호(Partial Homomorphic Encryption)로 불리우며 RSA, ElGamal 등이 대표적인 예입니다. 반면 두 가지 연산 모두 계산이 가능하지만 연산의 횟수에 제한이 있는 경우는 준동형암호(Somewhat Homomorphic Encryption), 마지막으로 연산의 종류와 횟수에 제한 없이 임의의 계산을 수행 가능한 암호시스템을 완전동형암호(Fully Homomorphic Encryption)로 분류합니다. 개념적으로는 완전동형암호가 우월하지만 실제 응용분야에서는 효율성과 편의성 등의 이유로 준동형암호 역시 좋은 선택이 될 수 있습니다.

동형암호의 안전성

잠시 동형암호에 대해 흔히 생기는 오해 중 하나를 소개하려 합니다. 만약 0을 암호화하여 암호문 $c=Enc(0)$ 를 얻고 이 암호문을 두 번 더한다면 동형암호 성질에 의해 그 결과물 $c+c$ 역시 $0+0=0$ 의 암호문일 것입니다. 그렇다면 $c+c=c$ 가 성립해야 하고 결과적으로 $c=0$ 이 되어 안전성이 보장되지 않는다고 생각할 수 있습니다. 위와 같은 문제가 발생하지 않은 이유는 하나의 평문에 대응되는 암호문이 유일하지 않기 때문입니다. 동형암호의 암호화 과정은 일종의 확률적 알고리즘으로 주어진 평문에 대응하는 수많은 암호문 중 하나를 랜덤하게 도출하게 됩니다. 이렇게 생성된 암호문 상에서 연산을 진행하면 그 결과에 대응하는 수많은 암호문 중 하나를 얻게 되므로 이로부터 평문에 대한 정보를(일부라도) 얻어내는 것은 불가능합니다.¹⁾

현재 가장 널리 쓰이고 있는 효율적인 동형암호 시스템들은 대부분 격자(lattice) 이론에 기반하고 있습니다. 격자이론은 수학과 컴퓨터과학 내에서 오랫동안 연구된 주제로 암호시스템의 설계와 안전성 증명에 널리 사용되고 있습니다. 기존 RSA, ECDSA 등에 사용되었던 소인수분해(integer



1) 동형암호의 암호화 과정은 함수가 아니라 확률적 알고리즘이므로, 엄밀하게 말하자면 위에서 동형암호를 정의할 때 사용했던 '암호화 과정이 동형사상'은 잘못된 표현이며 정확히는 '복호화 과정이 동형사상인 암호시스템'으로 정의해야 합니다.

factoring)나 이산로그(discrete logarithm)와 다르게 격자기반 문제들은 이를 효율적으로 풀 수 있는 양자 알고리즘이(아직까지는) 존재하지 않기 때문에 양자컴퓨터의 등장 이후에도 사용 가능한 차세대 암호의 가장 강력한 후보로 여겨지고 있습니다.

동형암호의 역사

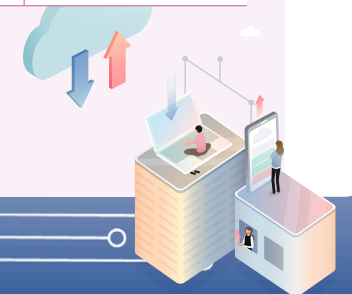
동형암호의 개념은 1978년에 이미 Rivest, Adleman 그리고 Dertouzos에 의해 구상되었습니다. 하지만 함께 제시되었던 스킴들의 취약점이 발견되면서 안전한 동형암호의 설계문제는 암호학계의 성배로 불리며 30년이 넘는 시간동안 미해결 상태로 남아있었습니다. 2009년 Gentry가 소위 재부팅(Bootstrapping)으로 불리는 방법을 제시하며 증명가능한 안전성을 가지는 최초의 완전동형암호를 설계하면서 본격적인 동형암호 연구의 장을 열었습니다.

동형암호 연구 초창기에는 학계 내에서도 동형암호의 실용화는 수십 년 이상의 시간이 걸릴 것이라는 부정적인 의견이 많았습니다. 그도 그럴 것이 2011년에 발표된 최초의 동형암호 개념증명(proof-of-concept) 구현결과는 재부팅 한 번에 30분이라는 비현실적인 수준의 실행시간을 보여줬기 때문입니다. 하지만 그 뒤로 약 10년간 다양한 측면에서 급격한 발전이 이루어졌습니다.

먼저 2012년 유한체 위에서의 모듈러 연산(modular arithmetic)을 지원하는 BGV, FV 스킴이 등장하였고 이와 함께 수백에서 수천 개의 숫자를 하나의 암호문에 저장할 수 있는 패킹(packing) 기술이 개발되었으며, 같은 해 이를 구현하여 3만개의 Boolean gate로 이루어진 AES-128 복호화를 동형암호 상에서 36시간 내에 계산하는 결과가 보고되었습니다. 다음 해 IBM에서는 다양한 최적화 과정을 거쳐 최초의 공개 동형암호 라이브러리 HElib을 출시하였고 이를 바탕으로 위와 같은 AES-128 계산시간을 4분으로 줄이는 성과를 거두었습니다. 한편 지원하는 연산의 종류를 다양화하는 방향의 노력도 있었는데 2016년 Chillotti 등이 개발한 TFHE 스킴은 단일 비트 연산에 적합하며, 마지막으로 2017년에 발표된 CKKS(a.k.a. HEAAN) 스킴은 기본적으로 근사계산을 지원하고 데이터 패킹이 가능하여 기계학습 등 실수연산을 포함하는 빅데이터 분석에 가장 적합하다는 평가를 받고 있습니다. 2020년에는 동형암호 기술의 산업적 활용을 목적으로 ISO/IEC 표준화가 제안되어 현재 표준화 과정이 진행 중입니다.

표 1. 동형암호 스킴과 특징

스킴	지원하는 연산	패킹	재부팅	속도
BGV, FV	유한체 위에서의 모듈러 연산	Yes	Optional	곱셈: 10~100ms (데이터 갯수 211-215)
TFHE	단일 비트연산(XOR)	No	Default	XOR+재부팅: 13ms
CKKS	실수(복소수) 위에서의 근사연산	Yes	Optional	곱셈: 10~100ms (데이터 갯수 211-215)



*** 동형암호의 성능과 응용

그 외에도 다양한 동형암호 스킴들이 연구되었지만 지금으로서는 위에 언급된 세 종류의 동형암호가 가장 경쟁력 있는 후보들이며 이들을 기반으로 학계 및 기업 다수의 기관에서 동형암호 라이브러리를 개발하고 관리하고 있습니다(표2). 동형암호의 성능은 무어의 법칙(Moore's Law)을 상회하는 속도로 발전하고 있는데 현재 공개된 라이브러리를 기준으로 수천 개의 실수를 저장하고 있는 BGV나 CKKS 암호문의 곱셈 연산은 인수(parameter) 선택에 따라 약 10ms 에서 0.1s의 성능을 보여주고 있으며 TFHE 암호문의 재부팅은 약 13ms가 소요됩니다(표 1).

동형암호 계산에 필요한 알고리즘 대부분은 병렬처리가 용이한 형태를 가지고 있습니다. 위에 소개된 모든 실험수치는 CPU 상에서 얻어진 결과이며 GPU나 FPGA등 다중코어를 이용한 구현을 통해 손쉽게 개선할 수 있습니다. 작년부터 미국 국방부 방위고등연구계획국(DARPA)에서는 동형암호 구현을 위한 하드웨어 가속기 개발에 3300만 달러 예산 규모의 사업에 착수하여 Microsoft, Intel과 업무협약을 체결한 상태입니다.

표 2. 대표적 동형암호 라이브러리 목록

라이브러리	개발기관	BGV or FV	TFHE	CKKS
HElib	IBM	✓		✓
Microsoft SEAL	Microsoft Research	✓		✓
HEAAN	Seoul National University			✓
FHEW	UC San Diego		✓	
TFHE	Inpher		✓	
Lattigo	EPFL	✓		✓
PALISADE	Duality Tech	✓		✓

동형암호 기초 라이브러리의 구축 이후에는 자연스럽게 이를 바탕으로 좀 더 복잡하고 실용적인 문제들에 적용하려는 시도들이 있었습니다. 대표적으로 미국 국립보건원(NIH)에서 후원하는 iDASH Security & Privacy Competition이 있는데 매년 생명과학 분야에서의 주요 프라이버시 문제를 선정하고 이를 동형암호를 포함한 최신 암호기술을 이용하여 해결하는 대회가 진행되고 있습니다. 최근 몇 년간의 결과를 살펴보면 현재 동형암호 기술이 로지스틱 회귀분석이나 소수의 층을 가지는 인공신경망 등 비교적 간단한 기계학습 모델의 학습(training)이나 추론(inference) 계산을 무리 없이 해내는 수준에 도달했음을 알 수 있습니다.

본격적인 동형암호의 사용 예로는 최근 Microsoft Edge에서 시작한 'Password Monitor'를 들 수 있는데 이 서비스는 사용자의 비밀번호가 유출되었는지 여부를 확인하고 경고하는 것을 목표로 합니다. 기업에서 보유하고 있는 유출된 비밀번호 목록과 사용자 계정 비밀번호 목록에 공통된 항목이 있는지를 암호화된 상태에서 계산하는 방식을 통해 사용자 비밀번호 목록에 대한 접근 없이 원하는 결과만을 도출할 수 있습니다.

*** 진행중인 연구와 결론

동형암호의 빠른 발전에도 불구하고 기술의 상용화를 위해 해결해야 할 과제들이 남아 있습니다. 암호화된 상태에서는 덧셈, 곱셈과 같이 제한된 연산만 가능하며 평균 상태와 비교해서 전혀 상이한 계산 복잡도를 가집니다. 그러므로 동형암호 상에서의 프로그래밍은 기술에 대한 깊은 이해를 요구하며 같은 계산을 수행하는 프로그래머더라도 최적화 수준에 따라 막대한 성능 차이를 보일 수 있습니다. 최근 비전문가의 진입장벽을 낮추기 위해 원하는 계산을 수행하는 스킴의 인수 및 알고리즘을 자동적으로 선택하여 동형암호 배경지식이 없이도 합리적인 수준의 프로그래밍을 가능하도록 하는 컴파일러 연구들이 진행되고 있습니다. 지금까지는 행렬-벡터 곱셈이나 인공신경망 추론을 도와주는 기초적인 수준이지만 근 미래에 통계적 분석과 기계학습 등 보다 복잡하고 일반적인 프로그래밍을 지원하는 컴파일러가 등장할 것으로 생각됩니다.

한편 데이터 3법 개정 및 데이터기반 산업의 팽창에 따라 동형암호를 포함한 최신 암호기술의 개발 및 전문가에 대한 수요가 빠르게 늘어날 것으로 예상하고 있습니다. 학계, 산업계 및 정부기관에 소속된 다양한 분야의 전문가들의 암호분야 진입과 전공자 배출을 통한 인력 육성을 통해 데이터 분석과 공유경제 시대에 대비하는 것이 절실한 시점입니다.



뇌과학과 인공지능의 융합을 추구하는 한양대학교 심리뇌과학과

김성신

한양대학교 인텔리전스컴퓨팅학부 심리뇌과학과

필자가 고등학생이었던 1996년도에 딥블루라는 IBM이 만든 컴퓨터가 세계 체스 챔피언 카스파로프를 이겼다는 뉴스를 접한 기억이 난다. 많은 사람들이 그 뉴스를 보면서 체스나 장기 정도는 컴퓨터가 이길 수 있겠지만 바둑은 어렵도 없고 앞으로도 인간을 이길 수는 없을 것이라고 생각했다. 필자도 바둑을 잘 두진 못하지만 대략적인 경기방식을 알고 있었고 고등학교 때 배운 확률이론으로 따져보더라도 경주의 수가 너무나 많아서 컴퓨터가 결코 인간을 이길 수 없다는 생각에 동의하였다.

20년이 지난 2016년 구글 딥마인드사가 개발한 알파고의 경기 소식을 들었을 때도 컴퓨터가 감히 인간을 이길 수 없을 것이라는 생각에는 변함이 없었다. 하지만 인간 대표인 이세돌은 한 판을 제외하고 알파고에 모두 패하였고 그 뒤로 알파고는 더 이상 인간을 이길 수 없는 상대가 되면서 급기야 은퇴를 선언하기에 이르렀다(그림 1). 그 후로도 인간이 둔 바둑의 기보를 학습하지 않고 바둑의 게임 규칙을 가지고 100% 강화학습 방식만을 이용한 알파고-제로는 기존의 알파고를 훨씬 뛰어넘게 되었다.



그림 1 구글 딥마인드의 알파고와 최종 경기에서 첫 수를 두고 있는 이세돌(제공: 한국기원)

인공지능의 한계와 목표 구글 딥마인드사의 최고경영자인 데미스 하사비스는 카이스트에서 했던 강연에서 딥마인드의 최종 목표는 범용인공지능(Artificial General Intelligence: AGI)을 개발하는 것이라고 하였으며 범용인공지능의 유일한 예는 우리의 뇌라고 하였다. 인공지능은 바둑의 예에서 볼 수 있듯이 예전에는 불가능하게만 여겨졌던 많은 일들을 가능하게 하였다. 자연어처리에 인공지능을 적용하여 인간의 언어를 이해하고 음악과 미술 분야에서 예술작품을 창조하는 것도 가능해 졌다. 자율주행 기술의 발전으로 운전자가 필요 없는 세상이 올 것이며 인간이 해결하지 못한 난치병들도 정확히 진단하고 치료하는데 인공지능이 폭넓게 활용될 것이다.

하지만 현재까지의 인공지능이 가진 한계점은 크게 두 가지로 생각해 볼 수 있는데 첫째는 인간의 뇌처럼 다양한 과제를 학습하고 적용할 수 있는 일반화된 지능을 갖추지 못한 것이며, 둘째는 학습에 필요한 데이터와 에너지가 인간의 뇌에 비해 비교할 수 없이 많아서 학습효율이 떨어진다는 것이다. 구체적으로 이세돌을 바둑에서 이겼던 알파고는 바둑 이외의 과제를 수행하는데 활용될 수 없고(물론 이후로 바둑, 장기, 체스를 동시에 학습할 수 있는 시스템이 개발되었으나 이 또한 범용지능과는 너무나 먼 이야기이다) 학습과정에서 엄청난 양의 전력을 소모하면서 한 개인이 평생 둘 수 없는 양의 바둑을 두었다. 이에 반해 바둑경기를 두는 동안 이세돌의 뇌가 사용한 에너지는 오렌지 주스 한 잔에 해당하는 에너지만큼도 되지 않으며 또 평생 두었던 바둑 경기의 수를 따져보아도 수 만 경기에 지나지 않을 것이다. 이미지를 인식하고 자연어를 처리하는 딥러닝시스템을 학습시키기 위해서 필요한 데이터의 양을 따져보아도 인간이 평생 동안 학습하는 것과

는 비교가 되지 않을 정도로 많다. 인간은 만 3세만 되어도 최소한의 학습자료만으로도 사물을 인식하는 법을 배우고 의사소통에 필요한 언어의 기본적인 문법까지도 구사할 수 있게 된다. 정리하자면 인간의 뇌처럼 최소한의 에너지를 소모하면서 일반적인 지능을 갖추는 진정한 의미의 인공지능을 개발하는 길은 아직까지 요원해 보인다.

비교적 짧은 뇌과학의 역사 그렇다면 새로운 차원의 인공지능을 개발하기 위해서 참고할 수 있는 유일한 예인 뇌에 대한 연구는 어느 정도 이루어져 왔는가? 수학, 물리학, 화학, 생물학과 같은 다른 기초과학에 비해서 뇌과학의 역사는 매우 짧다. 달리 말하면 우리 인류가 뇌에 대해서 알고 있는 지식은 대략 150년 전과 수천년 전을 비교했을 때 큰 차이가 나지 않는다. 가령, 19세기에 유행하였던 골상학은 머리의 모양과 크기로 인간의 심리적인 특징을 알 수 있다고 주장하였다. 물론 현재 골상학의 주장들은 혈액형으로 성격을 알 수 있다는 주장만큼이나 터무니없는 것으로 여겨지고 있다. 20세기 들어서도 인간의 뇌기능에 대한 이해가 매우 제한적이어서 두개골에 구멍을 뚫고 뇌의 전두엽 일부를 자르거나 다른 부분과의 연결 부위를 끊는 뇌 절제술이 정신질환을 치료하기 위한 목적으로 이루어졌다. 지금으로서는 상상할 수 없는 이러한 수술은 70년대 초까지 이어졌는데(얼음을 깨는 송곳을 사용한 간단한 시술로까지 발전함) 이 방법을 개발한 안토니우 모니스는 1949년에 노벨 생리의학상을 수상하기까지 하였다. 또 한 가지 중요한 예는 뇌과학 역사상 가장 중요한 실험 참가자라고 할 수 있는 환자 H.M.(헨리 몰라이슨: 1926~2008)이다(그림 2). 그는 뇌전증을 치료하기 위한 목적으로 당시에는 그 기능을 이해하지 못했던 해마체를 제거하는 수술을 받았으며 이 후

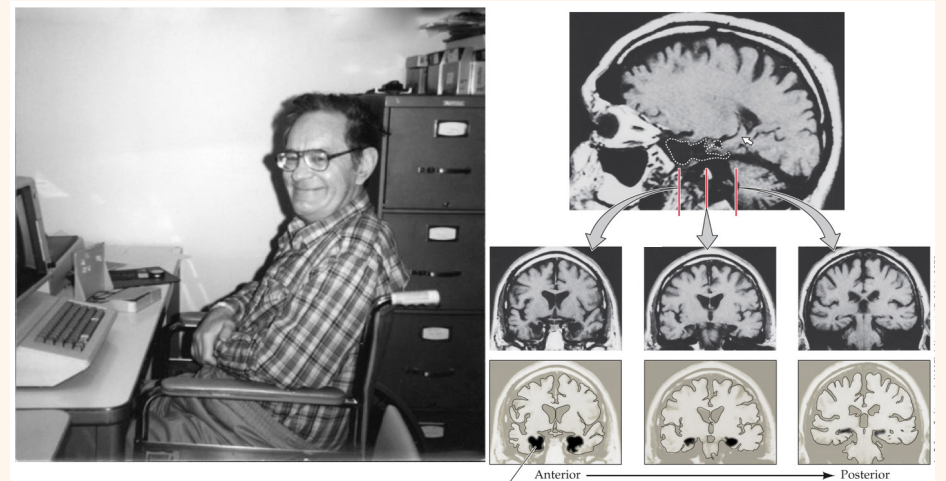


그림 2 해마체 제거 수술을 받은 환자 H.M.으로 알려진 헨리 몰레이슨(좌: 알마 제공), MRI로 촬영한 헨리 몰레이슨의 뇌, 양쪽 해마체가 제거된 것을 볼 수 있다(우: Corkin et al., Journal of Neuroscience).

로 새로운 기억을 형성하지 못하는 심각한 부작용으로 평생 동안 고통을 받았다.

이처럼 인간의 뇌에 대해 현재 우리가 알고 있는 거의 대부분의 지식들은 최근 50년 이내에 얻어진 것들이다. 최근 50년 동안 뇌과학의 발전을 이끄는 혁신이 이루어졌는데 그 중 대표적인 것이 핵자기공명영상(MRI)을 포함한 다양한 뇌영상 기술의 개발이다. 이러한 혁신에 힘입어 2013년도에 미국은 인간의 유전자지도도를 완성한 휴먼 게놈 프로젝트와 같이 인간의 뇌지도도를 완성하려는 목적으로 수십억 달러 예산규모인 브레인 이니셔티브를 출범하였고 유럽연합, 중국 및 일본, 그리고 우리나라에서도 뇌연구에 대한 지원이 확대되고 있다. 특히 최근 5년동안 인공지능에 대한 관심과 연구는 전 세계적으로 가히 폭발적으로 증가하고 있으며 이에 따라 뇌연구에 대한 관심도 높아지고 있다.

Neuroscience-inspired AI(뇌과학에 영감을 받은 인공지능 연구)

다시 데미스 하사비스의 말을 인용해 본다면 범용지능의 유일한 예인 뇌에 대한 연구는 지금과는 다른 차원의 인공지능을 개발하는데 결정적인 아이디어를 제공할 수 있다. 예를 들어 오늘날 인공지능의 핵심기술이 된 딥러닝에서 사용하는 합성곱 신경망(Convolutional Neural Network: CNN)의 기본적인 아이디어는 시각피질에서 일어나는 위계적인 정보처리과정과 유사하다. 시각 자극의 가장 기본적인 요소인 자극의 방향이나 경계와 같은 정보는 시각피질의 가장 뒤쪽에 자리잡고 있는 제 1차시각피질에서 처리되고 이 정보는 그 다음 몇 단계의 시각피질을 거쳐가면서 점점 복잡하고 일반적 또는 추상적인 정보를 처리하게 되며 최종적으로 하측두엽(inferior temporal lobe)에서 시각자극과 관련된 객체에

대한 인식이 일어난다. 다른 예를 들면, 구글 딥마인드에서 2015년 네이처지에 발표한 논문에서 사용한 Deep Q-Network(DQN)는 강화학습 알고리즘을 사용하여 컴퓨터가 오락실용 고전게임인 아타리 게임을 인간 수준 이상으로 잘 할 수 있도록 학습시킬 수 있다는 것을 보여주었다. DQN의 핵심 아이디어는 메모리 버퍼에 학습 데이터를 저장하였다가 임의적 샘플링을 통해서 다시 학습하게 만드는 것인데 이는 뇌에서 명시적으로 표현할 수 있는 기억을 생성하는 해마체의 반복재생(replay) 현상을 모사한 것이다. 실제로 데미스 하사비스의 박사학위 논문 주제는 해마체의 기능에 대한 것이었다. 필자가 노스웨스턴대학에서 박사후연구원으로 일할 때 연구실의 동료였던 Jane Wang은 현재 구글 딥마인드에서 일하고 있는데 경두개자극자극기로 해마체의 기억능력을 향상시키는 가능성을 보여준 연구를 수행하여 사이언스지에 출간한 적이 있는 뇌과학자이다.

이 밖에도 자연어처리에 많이 활용되는 재귀적 피드백을 가지고 있는 Recurrent Neural Network(RNN)도 작업기억(working memory)을 구현하는 전두엽의 기능과 매우 유사하다. 애초에 CNN, DQN, RNN과 같은 딥러닝 기술들은 인공신경망에 기반을 두고 있으며 결국 최초의 아이디어는 뉴런과 시냅스로 이루어진 신경계의 구조와 정보처리과정을 모방한 것에서 시작한 것이라고 할 수 있다.

그렇다면 인공지능은 반드시 뇌를 모방해야만 하는 것인가라는 질문을 할 수 있다. 물론 뇌가 인공지능 개발에 많은 아이디어를 준 것은 맞지만 대부분의 인공지능 분야 연구자들은 실제로 뇌과학에 관심을 갖거나 공부하지 않는다. 애초에 실리콘 반도체를 기

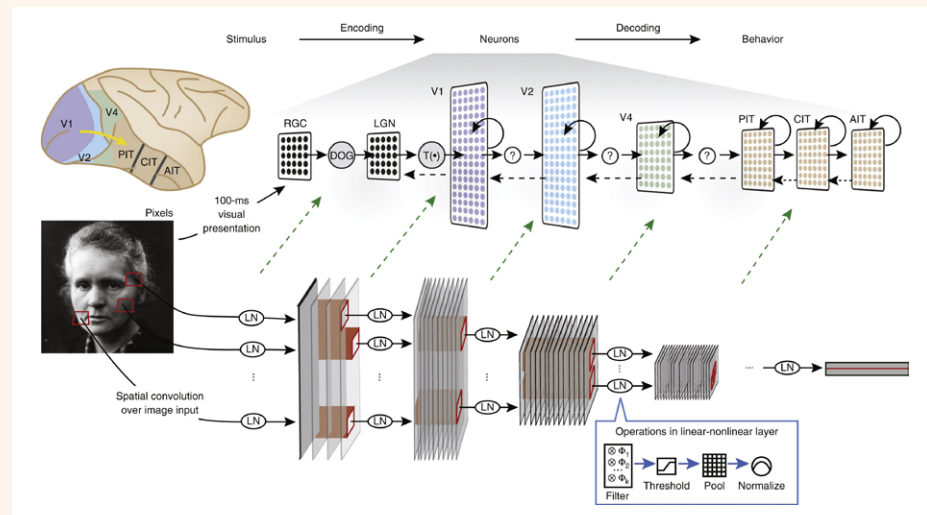


그림 3 | 시각중추의 위계적인 정보처리 과정을 모방한 Deep Convolutional Neural Network (출처: Yang and Wang, Neuron, 2020)

반으로 구현되는 인공지능과 단백질과 같은 유기체로 구성된 뇌가 같은 작동원리를 갖는 것은 불가능하다. 그리고 인공지능은 시스템이 수행하고자 하는 기능을 최적화하는데 목적이 있기 때문에 뇌가 가지고 있는 감정이나 의식과 같이 기능적인 측면에서 다소 불필요한 요소를 구현할 이유도 없을 것이다. 이러한 논의를 할 때 자주 나오는 이야기는 인류 최초로 동력비행기를 만든 라이트 형제에 대한 것이다. 라이트 형제는 새의 나는 방식에서 비행기를 만드는데 필요한 기본적인 아이디어를 얻은 것은 사실이고 애초에 하늘을 나는 새라는 존재가 없었다면 인간은 하늘을 날려고 하는 시도조차 하지 않았을지도 모른다. 하지만 라이트 형제가 새의 비행방식에만 집착하였다면 발명에 실패하였을 것이다. 실제로 라이트 형제 이전에 레오나르도 다빈치를 비롯한 많은 과학자들이 새가 나는 방식을 모방하려고 하였으나 실패하였고 실제로 새가 나는 정확한 원리는 동력비행을 성공하고 난 이후에서야 알 수 있게 되었다(그림 4).

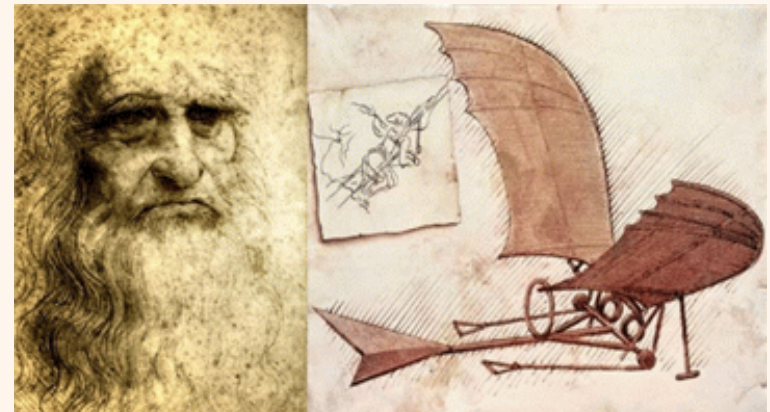


그림 4 | 레오나르도다빈치가 디자인한 사람의 팔 다리를 이용하여 날개를 퍼덕이며 날아오르게 디자인한 비행기(위). 라이트형제가 발명한 세계최초의 동력비행기(아래, Getty Images 제공)

인류는 이제 하늘을 나는 것을 넘어서서 화성에 탐사선을 보내는 수준에 도달하였고 더 나은 우주선을 개발하기 위해서 새가 나는 법을 연구할 필요가 없어졌다. 하지만 인공지능이 바둑과 같이 극히 제한된 과제를 학습하는 능력이 인간보다 뛰어나다고 하더라도 앞서 언급한 것처럼 진정한 지능을 갖추었다고 말하기엔 부족하며 따라서 뇌에 대한 연구를 통해 얻을 수 있는 아이디어는 아직도 무궁무진하다. 사실 인공지능을 이야기하기 전에 우리는 아직 뇌에 대해서 아는 것이 별로 없다.

AI-inspired Neuroscience(뇌과학에 영감을 받은 인공지능)

지금까지 논의한 예들처럼 대다수의 사람들이 뇌에 대한 연구가 인공지능 분야의 발전에 미친 영향들에 대해서 이야기하지만 반대의 경우를 생각할 필요가 있다. 사실 대다수의 뇌과학자들은 이 부분에 더 많은 관심을 갖고 인공지능과 관련된 연구에 주목하고 있다. 물리학에서도 이론 물리학자들이 이론적으로 예측한 현상이나 물질의 존재를 실험적으로 증명하는 방식으로 연구가 이루어진 경우가 많다. 대표적인 예로 아인슈타인의 일반상대성이론이 예측하는 중력에 의한 시공간의 휘어짐은 1919년 유명한 개기일식에서 보인 별의 위치를 통해서 확인이 되었으며, 또한 일반상대성이론이 예측하는 블랙홀의 존재도 증명이 되어 2020년에 노벨물리학상이 수여되기도 하였다.

뇌과학의 역사에서 필자가 가장 흥미롭게 생각하는 사례는 80년대 컴퓨터 과학자들이 개발한 시간차 학습이라는 강화학습 알고리즘에 대한 것이다. 이 알고리즘은 보상을 통해서 행동을 교정하여 보상을 최대화시키는 방법을 학습하는 것인데 이 알고리즘이 추정하는 보상예측오류의 크기가 뇌에서 분비되는 도파민의 양과 관련이 있다는 것을 발견한 것이다. 즉, 보상에 대한 기대값을 예측하는 강화학습 알고리즘을 이용하여 도파민이 보상의 크기 자체보다는 실제 보상과 기대한 값의 차이에 반응한다는 것을 발견한 것이다. 이러한 결과는 '기대가 크면 실망도 크다'라는 속담이 갖는 의미와 왜 상대방이 모르게 깜짝 선물을 준비하는지 이유를 잘 설명해 준다. 또한 최근 2020년에는 구글 딥마인드가 발표한 네이처 논문에서 쥐의 뇌에서 예측하는 보상의 값이 다양한 도파민 뉴런들이 존재한다는 것을 발견하였는데 이는 강화학습 모델 중 학습 성능이 뛰어난 '분포강화학습'이 실제로 뇌에서도 이루어진다는 것을 입증한 것으로서, 이 역시 인공지능의 주요한 알고리즘이 실제로 뇌에서 구현이 된다는 것을 보여준 예이다.

최근 유행하는 딥러닝과 관련한 흥미로운 연구는 이미지를 인식하거나 아타리 게임을 수행하면서 학습된 딥네트워크(Deep

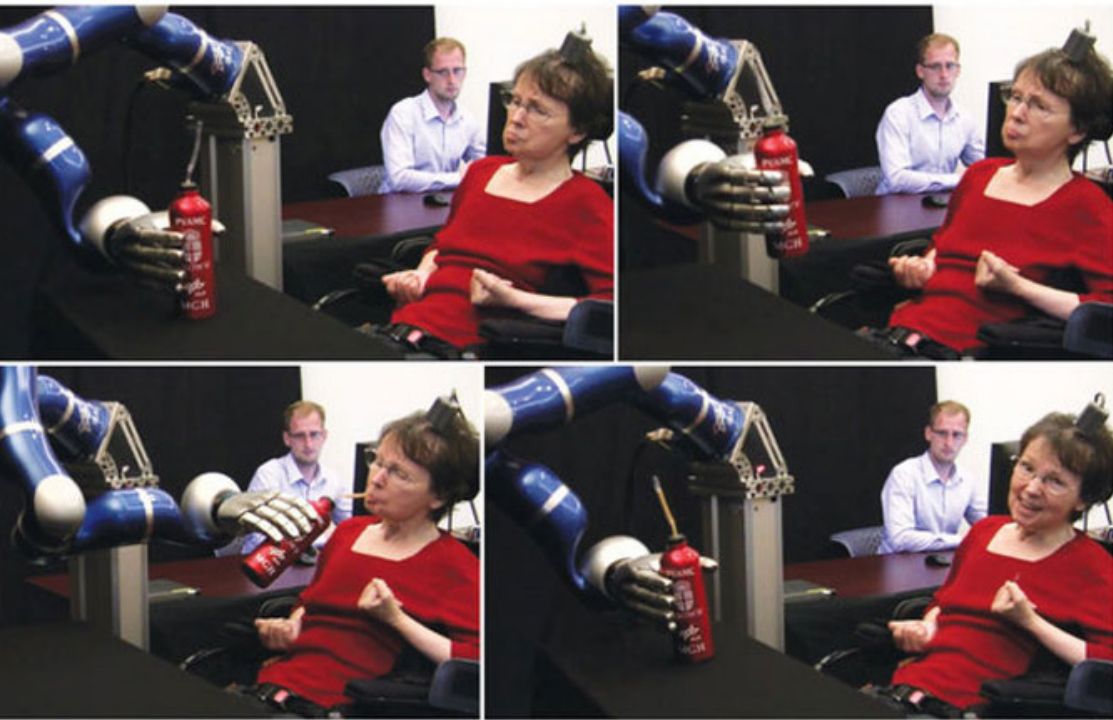


그림 5
브라운 대학의 연구진이 사지마비 환자가 생각만으로 로봇팔을 움직여서 음료수 병을 잡고 먹게 하는 뇌-기계 접속 시스템을 개발함.
(출처: Hochberg et al., Nature, 2012)

network) 를 모델로 같은 과제를 수행할 때 실제 fMRI 에서 활성화 된 패턴을 잘 예측하는 것을 보여주었다. 더 나아가 뇌에서의 정보가 처리되는 위계관계가 딥네트워크의 층들간 위계관계와 일치한다는 것을 보여줌으로써 뇌의 정보처리 기능을 모방한 딥러닝 방식이 개발되었는데 이제 역으로 딥러닝이 뇌가 실제로 정보를 처리하는 방식을 이해하는데 사용되고 있는 것이다.

뇌과학의 도구로서의 인공지능 인공지능 알고리즘과 뇌의 기능 간 연관관계를 밝힌 위에 언급된 예들 이외에도 인공지능 연구를 통해 개발된 다양한 알고리즘들은 뇌연구를 하는데 폭넓게 활용되고 있다. 2000년대초에 최초로 소개되었고 최근에는 일론 머스크가 투자한 뉴럴링크에서도 선보인 뇌-기계 접속 기술은 뇌에서

발생하는 뉴런들의 발화 패턴을 학습하고 이를 예측하는 기계학습 알고리즘들을 이용한다(그림 5). 2019년, 2020년 네이처와 그 자매지에 소개된 연구에서는 RNN(Recurrent Neural Network) 이라는 인공지능 알고리즘을 이용하여 대뇌피질에서 얻어진 전기적인 신호를 피험자가 의도한 언어와 문장으로 변환함으로써 뇌-기계 접속을 통해 의사소통이 가능할 수 있음을 보여주었다.

기능성자기공명영상(fMRI)을 이용한 인간의 인지적 기능에 대한 연구도(그림 6) 2000년대 이후 기계학습 알고리즘들을 도입하였는데, 특히, 시각중추의 활성화 패턴을 분석하여 주어진 자극의 이미지나 영상까지도 비록 뚜렷하지는 않지만 유의미하게 재구성하는 연구가 성공적으로 이루어졌다. 심지어 꿈에서 보았던 이미지

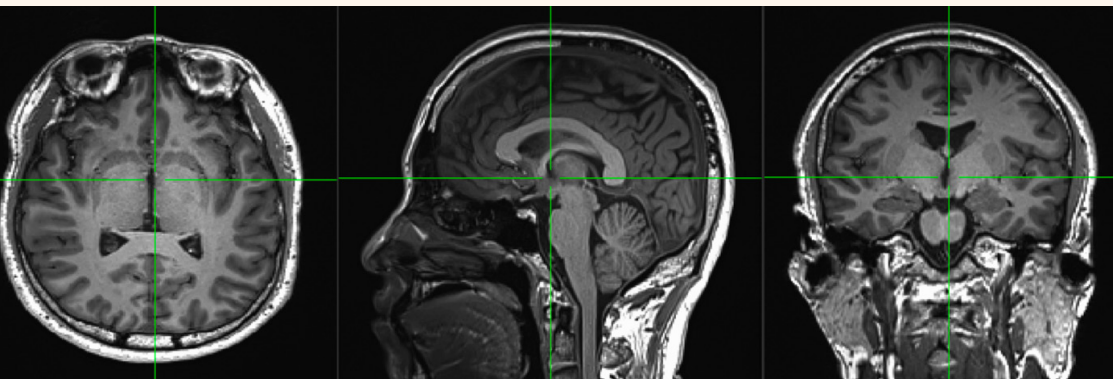


그림 6
성균관대학교 소재 뇌과학이미징연구단의 MRI 로 찍은 필자의 뇌. 위에서 본 모습(좌), 옆에서 본 모습(중), 뒤에서 본 모습(우)

를 예측하는 연구도 2013년 사이언스지에 소개된 바가 있다.

최근에는 인공지능 알고리즘 중 혁신적인 생성모델로 주목을 받고 있는 VAE(Variational Autoencoder) 와 GAN(Generative Adversarial Network) 모델을 이용하여 뇌의 시각중추에서 fMRI 로 얻어진 활

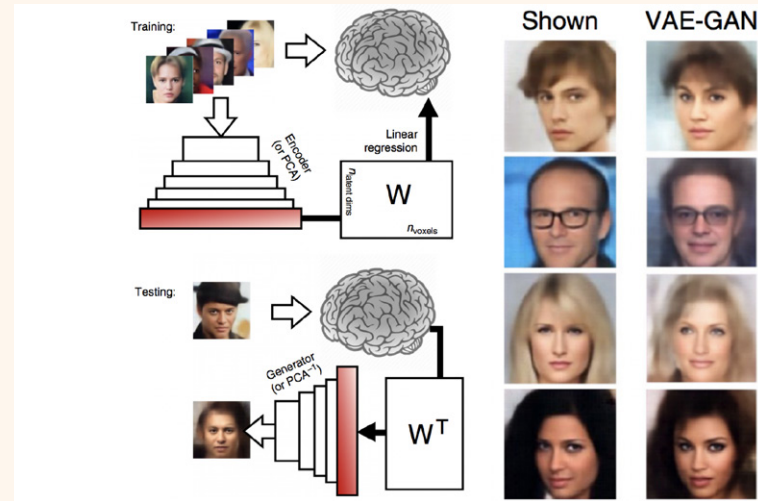


그림 7
시각중추를 촬영한 fMRI 데이터에 인공지능 알고리즘인 VAE 와 GAN을 적용하여 피험자가 보았던 얼굴의 이미지를 재구성해 낸 연구(출처: VanRullen a& Reddy, Communications Biology, 2019)

성화 패턴으로 피험자가 보았던 얼굴의 이미지를 재구성하는 연구가 소개되었다(그림 7). 또한 인공지능 연구에서 자주 사용되는 확률적 추론방법 중 하나인 베이저언 추론 방식이 뇌가 외부의 자극을 처리하고 학습하는 과정을 설명하는 주요한 이론으로 사용되고 있다. 뇌를 하나의 거대한 네트워크로 이해하는 브레인 커넥톰이라는 개념도 주목을 받고 있는데 이 또한 인공지능의 기술에도 적용되는 네트워크 이론들이 뇌의 복잡한 인지과정을 설명하는데 매우 중요한 도구로 사용되고 있는 예이다. 결국 뇌의 활동을 측정하는 기술들이 다양해지고 보다 정교해지면서 실험에서 얻어지는 데이터의 품질이 높아지고 그 양도 많아지면서 이를 다루기 위한 인공지능의 기술들이 중요해지고 있는 것이다. 필자의 경우도 초고해상도 7 테슬라 fMRI를 사용할 경우 한 개의 프로젝트를 수행하는 과정에서 생성되는 데이터가 수십테라바이트 정도는 쉽게 넘어서기도 한다. 지금까지 살펴본 다양한 사례들에서 인공지능과 뇌에 대한 두 연구분야는 현재까지 상호 간에 긴밀한 교류가 이루어져 왔으며 앞으로 이러한 경향은 더 가속화될 것이다.



한양대학교 심리뇌과학과 소개

이러한 배경 가운데 한양대학교에서 데이터사이언스학과와 심리뇌과학과로 구성된 인텔리전스 컴퓨팅학부를 신설하였다(그림 8). 필자는 작년 가을에 심리뇌과학과에 신입 조교수로 임용이 되어 올해 첫 신입생을 받게 되었다. 위에서 다소 길게 배경에 대해 설명한 것은 한양대 심리뇌과학과에서 중점을 두고 수행할 교육과 연구 분야가 최근의 이와 같은 흐름과 밀접하게 관련이 있기 때문이다.

한양대 심리뇌과학과는 대학입시에서 자연계와 인문계 학생 모두 지원할 수 있지만 교육과정은 공과대학에 소속된 학과들과 유사해서 국내 대학에 있는 기존의 심리학과와는 확연하게 차이를 보인다. 굳이 유사한 심리학 분야를 생각해 본다면 인간의 행동을 모델링하고 정량적으로 분석하는 계량심리학에 가깝다. 뇌과학 분야도 연구 방식과 주제에 따라서 매우 다양한 접근이 가능하지만 한양대의 심리뇌과학과는 인간의 뇌 기능을 연구하는 인지뇌과학, 그 중에서도 뇌에서 일어나는 정보처리 방식, 의사결정, 기억 및 학습과 같은 주제에 집중하고 있다. 따라서 교과과정 전체가 프로그래밍을 기본으로 데이터를 다루고 해석하는 분석적인 능력을 기를 수 있도록 구성되어 있다.

구체적으로 1학년 때는 프로그래밍의 기초를 배우고 데이터사이언스에 대한 기본적인 개념들과 이

를 이해하기 위한 기초적인 수학과목들을 수강한다. 특히 최근 고교 교육과정에서 빠졌던 행렬과 벡터를 다루는 선형대수 과목이 필수과목으로 지정이 되어 있다. 미적분 과목도 공과대학의 교육과정과 유사하게 1학년 1, 2학기에 수강할 수 있도록 구성되어 있다. 2학년 때는 컴퓨터공학의 핵심적인 과목들, 예를 들어 자료구조론, 알고리즘, 데이터베이스 시스템과 같은 과목들을 수강하고 선형대수와 함께 데이터를 다루는데 가장 중요한 개념인 확률과 통계 과목을 수강한다. 동시에 기본적인 기계학습 알고리즘을 이해하고 1학년 때 배운 파이썬과 같은 프로그래밍을 통해서 실제적인 문제에 적용하는 법을 배우는 과목을 수강한다. 2학년 과정을 마치면 뛰어난 학생인 경우 현업에서도 인턴십 정도는 할 수 있는 수준이 되는 것을 목표로 교과 과정을 구성하였다. 2학년 2학기에 인지뇌과학의 개론을 처음 접하고 3학년 때부터는 계산인지과학, 인지행동모델링, 뇌영상기법과 같은 수업을 통해서 뇌의 정보처리과정, 기억 및 학습, 의사결정 과정 등 뇌의 고등인지 기능을 배우게 된다. 동시에 인공지능의 주요한 과목인 딥러닝, 베이지언이론, 강화학습이론과 같은 과목을 수강하는데 이러한 교과과정 구성이 한양대 심리뇌과학과의 가장 큰 강점이라고 생각한다.

정리하자면, 2학년 때까지 데이터를 다루고 이해하는 기본적인 과목인 프로그래밍과 수학과목들을 수강한 후 3학년 때부터 뇌과학과 인공지능의 주요 과목들을 수강하도록 하는 교육과정은 국내 어느 학교에서도 시도하지 못한 것이다. 인공지능이나 데이터사이언스를 가르치는 학과에서 뇌과학을 접하기 쉽지 않고 뇌과학도 분야가 워낙 넓기 때문에 인공지능과 가장 관련성이 높은 계산인지신경과학에 초점을 맞춰서 교과과정을 구성하는 경우가 없기 때문이다. 심리뇌과학과와 쌍둥이 같은 데이터사이언스학과가 1년 먼저 같은 인텔리전스컴퓨팅학부 내에 신설된 것도 이와 같이 두 분야에 접점을 만들고자 하는 한양대의 시도라고 할 수 있다.

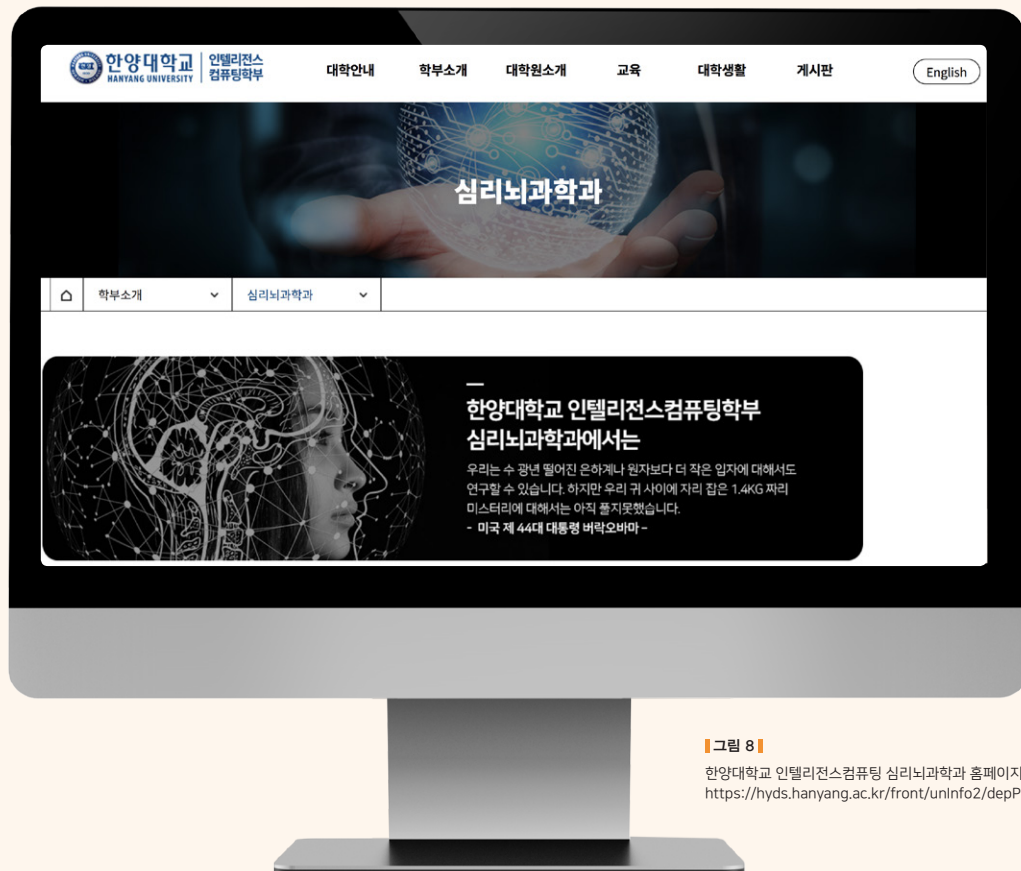


그림 8 | 한양대학교 인텔리전스컴퓨팅 심리뇌과학과 홈페이지: <https://hyds.hanyang.ac.kr/front/uninfo2/depPsy>

한양대학교 심리뇌과학과 전망

현재 미국과 유럽의 소위 과학 선진국에서는 뇌과학, 심리학, 인공지능의 융합이 더 이상 낯설지 않다. 대부분의 연구 중심 대학은 뇌과학과가 학부과정에 있으며 심리학과 내에서도 뇌를 연구하는 학자들이 많고 인공지능 연구와도 밀접하게 관련이 있는 경우가 많다. 하지만 이처럼 뇌와 인공지능의 융합 연구와 관련된 산업은 구글의 딥마인드, 일론 머스크가 투자한 뉴럴링크와 같은 회사를 제외하고 아직까지 많지 않다. 그나마 언급된 두 회사도 이윤을 만드는 것보다는 미래기술을 연구하는 것에 초점을 두고 있기 때문에 현재로서는 연구소에 가깝다고 할 수 있다.

딥마인드의 데미스 하사비스가 한 말처럼 범용인공지능을 만들고자 하는 그들의 목표는 1960년대 달에 가려는 아폴로 프로젝트와 같이 인류의 삶에 직접적인 도움이 되지 않을 수 있으나 인류의 꿈을 이루는 일처럼 여겨진다. 하지만 이러한 과정 가운데 인류는 우리에게 닥친 수없이 많은 문제를 해결할 수 있는 열쇠를 찾을 수 있을 것이다. 우울증과 치매, 자폐를 비롯해 인류가 쌓아올린 과학기술을 한없이 무기력하게 만드는 각종 뇌질환을 극복하게 될 날이 올지도 모르고 지금보다 비교할 수 없이 진보된 인공지능을 활용하여 인류가 겪고 있는 수많은 과학의 난제들을 풀고 기후변화와 같은 인류가 맞닥뜨린 위기를 극복하게 될지도 모른다. 궁극적으로 뇌와 인공지능에 대한 연구를 통해 지능이란 무엇인가에 대해 답을 함으로써 최종적으로 우리 인류는 의식의 본질, 즉 인간의 본질을 이해할 수 있을지 모른다. 학과의 전망에 대해서 너무 거창하게 말하였지만 한양대학교 심리뇌과학과는 이처럼 위대한 가치를 추구하면서 동시에 현실적인 문제해결 능력을 갖춘 학생들을 길러내고자 한다.

필자는 미국 유학 중 라이트형제가 첫 비행에 성공한 장소인 노스캐롤라이나의 Kill Devils Hills 라는 곳을 가본 적이 있다(그림 9). 하늘을 날고자 했던 형제들의 위대한 도전 정신과 목숨을 건 불굴의 의지와 신념을 생각할 때마다 경외감을 느낀다. 그곳에 세워진 라이트형제의 기념탑에 새겨진 다음과 같은 말로 이 글을 마무리 하고자 한다.

“
IN COMMEMORATION OF THE
CONQUEST OF THE AIR BY
THE BROTHERS WILBUR AND
ORVILLE WRIGHT. CONCEIVED
BY GENIUS. ACHIEVED BY
DAUNTLESS RESOLUTION AND
UNCONQUERABLE FAITH.
”



그림 9 | 필자가 10 여년 전 방문한 미국 노스캐롤라이나 Kill Devils Hills 에 세워진 라이트형제 기념탑

R에 도전하자... 따라가다 보면, 나도 R유저 ⑨

'R에 도전하자' 지난 호의 ①부터 ⑧에서는 R의 설치부터 시작하여 몇 가지 기능을 따라 해보고 R에서의 색깔처리, R의 기초적인 통계함수 및 R의 벡터 및 데이터 프레임을 사용한 자료처리, 함수작성, 그래픽 장치, ggplot 패키지 등에 대해 알아보았다. 이번 호에서는 지난 호의 ggplot2 패키지를 사용한 R의 그래픽 기능에 대해 좀 더 알아보기로 하자.

1 | ggplot2 패키지

이번 호에 사용할 프로그램은 모두 ggplot2 패키지에 대한 내용이므로 이 패키지를 먼저 설치하여야 한다. 설치하는

```
> install.packages("ggplot2")
```

라고 명령하면 설치할 패키지를 내려받을 사이트의 목록이 나오고 이 목록에서 아무 곳이나 선택하면 인터넷을 통해 패키지를 내려받아 로컬 컴퓨터에 설치가 된다. 패키지 탑재는 library 함수를

```
> library(ggplot2)
```

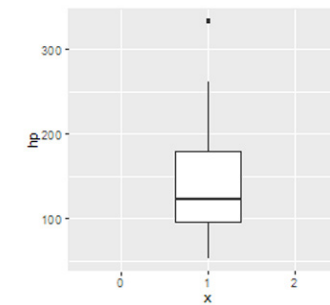
로 사용하여 ggplot2를 탑재하며 설치된 R의 버전이 ggplot2의 버전보다 오래된 것이면 경고메시지가 출력되기도 하는데 이런 경우 대부분 그냥 사용하여도 되나 R을 최신 버전으로 업데이트하면 해결된다.

2 | 상자 그림(Box Plot)

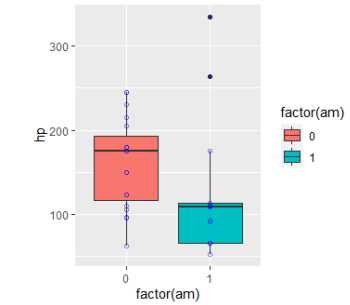
상자 그림은 최대, 최소 및 25%, 50%, 75% 백분위 수를 상자로 표시하는 그림으로 R의 boxplot 함수나 ggplot2의 qplot 함수로도 그릴 수 있으나 여기서는 geom_boxplot 함수로 그리는 방법을 살펴보자. geom_boxplot 함수에서는 상자의 크기 및 이상치(outlier)의 색깔, 크기, 모양, 수염의 최대길이 등을 추가로 설정할 수 있다.

보기 1: 상자 그림은 geom_boxplot으로 그린다. 한 변수의 자료에 대해 상자 그림을 그릴 때는 x에는 상자 그림의 x 좌표, y에는 상자 그림을 그릴 변수를 설정한다. 아래는 mtcars 데이터 프레임의 마력(hp)에 대한 상자 그림을 그린 것이다. mtcars 데이터 프레임은 R에 내장된 데이터 프레임으로 자동차의 모델, 마력(hp), 변속기 종류(auto/manual) 등이 저장되어 있다. 다음 그림의 왼쪽은 hp(마력)에 대한 상자 그림이고 오른쪽은 am(변속기의 종류로 0은 자동변속기, 1인 수동변속기)의 값에 따른 hp의 상자 그림이다. 왼쪽 그림에서 사용한 scale_x_discrete는 그림에서의 가로축 x의 범위를 이산형으로 0, 1, 2로 설정한 것이다. 이 부분을 생략하면 상자 그림의 폭이 크게 그려진다.

```
> ggplot(mtcars) +  
  geom_boxplot(aes(x='1', y=hp)) +  
  scale_x_discrete(limits=factor(0:2))
```



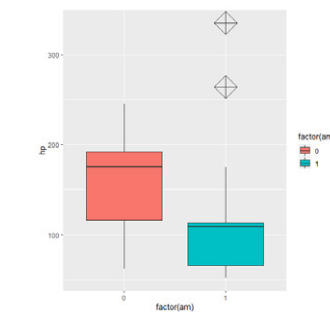
```
> ggplot(mtcars, aes(factor(am), hp)) +  
  geom_boxplot(aes(fill=factor(am))) +  
  geom_point(alpha=.5, shape=21, colour="blue")
```



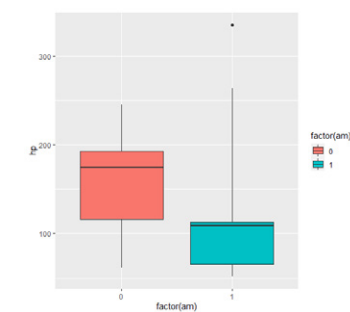
이 그림의 왼쪽에서 aes 함수 내에 x의 값을 '1'로 설정한 것은 aes 함수가 가로축과 세로축이 모두 설정되어야 한다. 한 변수에 대한 상자 그림에서 x축의 값을 정할 수가 없으므로 이를 설정한 것이다. 오른쪽 그림은 fill에서 변속기의 종류별로 그려달라고 했으므로 자동으로 x축이 변속기가 되며, 원자료의 값을 산점도로 추가하기 위해 geom_point 함수를 사용하였다.

다음의 예들은 점의 크기, 모양 및 수염길이를 추가로 설정한 예이다.

```
> ggplot(mtcars, aes(factor(am), hp)) +  
  geom_boxplot(aes(fill=factor(am)),  
  outlier.shape=9, outlier.size=10)
```



```
> ggplot(mtcars, aes(factor(am), hp)) +  
  geom_boxplot(aes(fill=factor(am)),  
  coef=4)
```



이 그림의 왼쪽은 상자 그림의 최대 수염길이를 넘어가는 자료에 대해서 9번 점모양(plot 함수의 pch 참고), 점의 크기는 10으로 설정한 결과이다. 상자 그림의 수염길이는 IQR(상자 그림의 상자의 높이)의 1.5배까지만 늘이고 그 바깥의 자료는 별도의 점으로 표시하는데 이 수염의 길이를 IQR의 4배까지 확장하도록 설정한 결과이다.

3 | 원그래프(Pie chart)

ggplot 패키지에 원그래프를 그리기 위한 geom은 따로 제공되지 않는다. 따라서 원그래프를 그리기 위해서는 약간의 편법으로 막대그래프를 그린 후 이를 각도로 변환하는 방법을 사용한다. 각도로 변환하기 위해서는 coord_polar 함수 등을 사용한다. 이 함수는

```
coord_polar(theta = "x", start = 0, direction = 1)
```

로 사용할 수 있다. 이 함수에서 필요한 theta는 각도를 설정하는 것이 기본인데 geom_bar의 계산에서 얻어올 수 있으므로 빈도에 비례하는 각도를 추가로 계산하지는 않는다. start는 원그래프의 시작 각도이며 direction은 부채꼴이 생성되는 순서로 1은 반시계방향, -1은 시계 방향이다.

보기 2: 앞의 mtcars 데이터 프레임에서 cyl은 자동차 엔진의 실린더 개수다. 이 자료에 대해서 각 실린더 개수에 해당하는 자동차 수의 빈도에 대한 원그래프를 그려보자. 이를 위해 먼저 cyl 변수의 빈도수를 다음과 같이 얻는다.

```
> freq_df <- data.frame(table(mtcars$cyl))
```

Var1	Freq
1	4
2	6
3	8

이는 도수분포표의 결과를 데이터 프레임으로 변환하여 저장한 것이고, 원그래프의 범례에서 사용하기 위해 각 범주의 이름을 설정해보자. 이름은

```
> freq_df$name <- paste(freq_df$Var1, "cyl", sep="-")
```

와 같이 주어 범주의 이름은 cyl 다음에 -4, -6 또는 -8로 Var1 변수의 값이 붙게 하고 이를 기준 데이터 프레임 freq_df의 name이라는 변수에 저장하였다. 최종 데이터 프레임의 값은 다음과 같다.

```
> freq_df
```

	Var1	Freq	name
1	4	11	4-cyl
2	6	7	6-cyl
3	8	14	8-cyl

이제 이 데이터 프레임의 Freq 변수값을 사용하여 원그래프를 그려보자. 기본 원그래프의 경우 x축이 없으므로 막대그래프 정보를 얻어올 때 x를 없앤다. 따라서

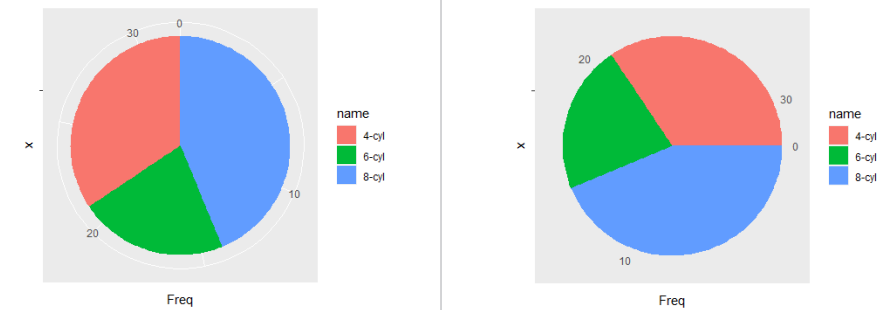
```
> p0 <- ggplot(freq_df, aes(x="", y=Freq, fill=name)) +
  geom_bar(stat = "identity")
```

로 막대그래프 그림을 얻고 이 막대그래프를 사용하여

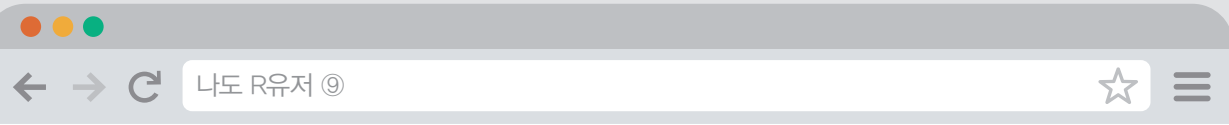
```
> piep <- p0 + coord_polar("y", start=0)
> piep
```

로 coord_flip 함수를 추가하면 아래의 왼쪽 그림과 같은 원그래프를 얻는다.

원그래프는 막대그래프를 그리기 위해 계산된 결과에 각도 0(12시 방향)에서 시작하여 반 시계 방향으로 호를 그린다. 이러한 경우 start에 0을 설정하며 12시 방향에서 시계 반대 방향으로 호를 그리기 시작하였다(왼쪽 그림).



원그래프에서 바깥쪽의 원을 없애려면(이 원은 극좌표의 좌표 표시이기도 함) 다음과 같이 panel.grid를 element_blank()를 추가하면 된다.



```
> piep <- ggplot(freq_df, aes(x="", y=Freq, fill=name)) +
  geom_bar(stat = "identity") + coord_polar("y", start=pi/2)
> piep + theme(panel.grid = element_blank())
```

이 그림에서는 시작하는 각도를 90도(=pi/2)로 하여 실제로는 수평선에서부터 각 슬라이드를 그리도록 하였다.

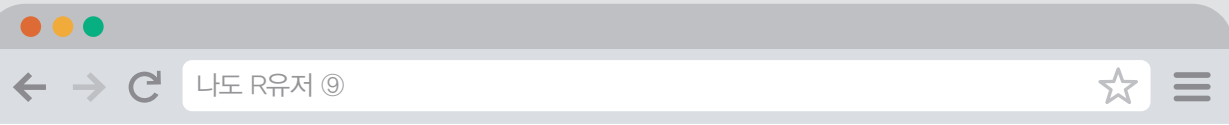
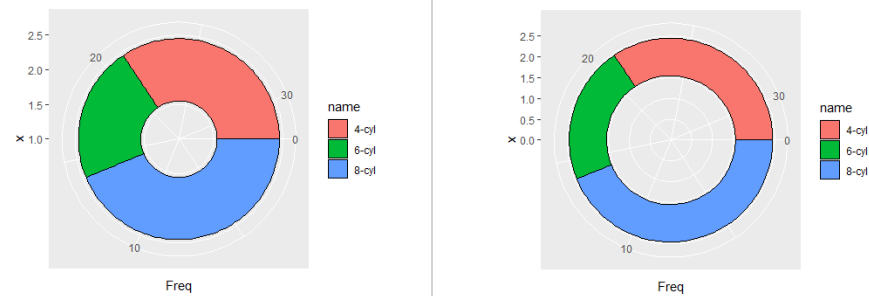
4 | 도넛 그래프와 이중 원그래프

도넛 그래프는 원그래프에서 가운데 작은 원을 만들어 이 부분을 비워둔 형태의 그림을 말한다. ggplot에서 도넛 그래프는 ggplot 함수에서 x값을 설정하여 원그래프의 x 좌표를 적절하게 설정한 후 그림을 그릴 x의 범위를 설정하면 된다. 다음은 ggplot에서 x에는 2를 설정하고 마지막 그림의 x 범위를 xlim 함수로 1부터 2.5까지 설정하였다.

```
> piep <- ggplot(freq_df, aes(x=2, y=Freq, fill=name)) +
  geom_bar(stat = "identity", color="black" ) +
  coord_polar(theta="y", start=pi/2) +
  xlim(1, 2.5)
```

위 명령의 결과는 아래의 왼쪽 그림과 같다. 이 그림 왼쪽 위의 눈금 표시를 보면 중간이 1이고 가장 높은 곳이 2.5임을 알 수 있다. xlim의 값을 적절히 수정하면 도넛의 중간부분의 크기를 조정할 수 있다. 아래의 오른쪽 그림은 위의 코드와 같으나 xlim의 값만 0.0부터 2.5로 바꾼 것이다.

```
> piep <- ggplot(freq_df, aes(x=2, y=Freq, fill=name)) +
  geom_bar(stat = "identity", color="black" ) +
  coord_polar(theta="y", start=pi/2) +
  xlim(0.0, 2.5)
```



5 | 누적 막대그래프, 가로막대 및 side-by-side 막대그래프

기본 막대그래프는 지난 호에서 다루었으므로 이번엔 이의 응용에 대해서 알아보자. 막대그래프의 경우 그룹별 범주의 빈도를 그리는 경우가 있는데 이런 경우 막대를 쌓아서 표시하거나 그룹별 막대를 옆에 나란히 붙이는(side-by-side) 경우가 있다.

보기 3: 다음과 같이 성별(gender)에 따라 좋아하는 TV 프로그램 영역을 얻었다고 하자.

```
> f.tab2 <- data.frame(drama=c(20, 60), sports=c(40, 10), news=c(40, 20),
  gender=c("male", "female"))
```

```
> f.tab2
```

	drama	sports	news	gender
1	20	40	40	male
2	60	10	20	female

이와 같은 자료는 요약된 것이고 막대그래프 등을 그리려면 빈도가 저장된 변수를 지정해야 한다. 이 자료는 빈도가 세 개의 변수에 나누어 들어가 있으며, 이를 통해 성별, 프로그램 영역별 빈도를 얻어야 하므로 자료의 형식을 바꾸어야 한다. 이를 위해서는 reshape2 패키지의 melt 함수를 많이 사용한다.

```
> library(reshape2) # 패키지 설치가 안 되었으면 install.packages 먼저!!
```

```
> f.tab2.m <- melt(f.tab2, id.vars='gender')
```

```
> f.tab2.m
```

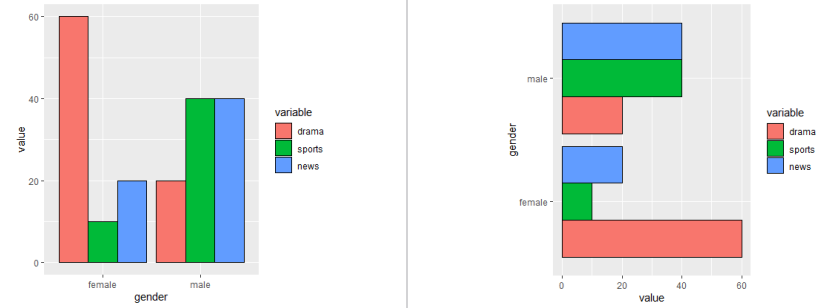
	gender	variable	value
1	male	drama	20
2	female	drama	60
3	male	sports	40
4	female	sports	10
5	male	news	40
6	female	news	20

이 자료를 사용하여 성별에 따른 막대그래프를 그리려면 ggplot에서 x는 성별을 y에는 빈도를(막대 그래프의 기본 설정) 설정한다. 프로그램 종류별로 색을 달리하여 그리기 위해서는 fill에 프로그램 종류를 설정하였다.

```
> piep2 <- ggplot(f.tab2.m, aes(x=gender, y=value, fill=variable)) +
  geom_bar(position="dodge", stat = "identity", colour="black")
```

```
> piep2
```

side-by-side로 막대를 그리기 위해서는 geom_bar 함수에서 position에 "dodge"를 설정하면 된다. 위의 명령으로 아래의 왼쪽 그림과 같은 side-by-side 막대그래프를 얻을 수 있다.



왼쪽 그림은 세로 막대이지만, 이를 오른쪽 그림과 같이 가로막대 그래프로 얻으려면 coord_flip() 속성을 추가해주면 된다.

```
> piep2 + coord_flip()
```

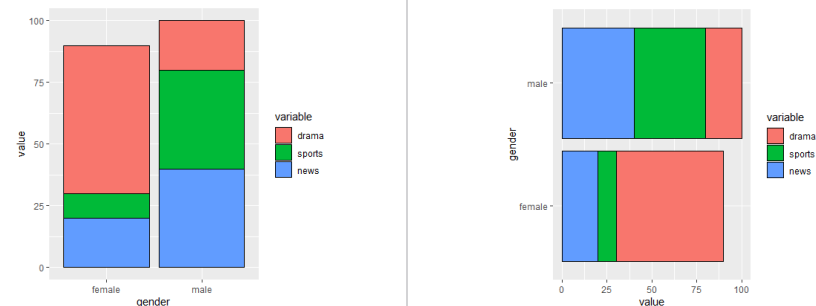
누적 막대그래프는 position에 "stack"이라고 설정하거나 position을 설정하지 않으면 된다. position의 기본값이 "stack"이기 때문이다.

```
> piep3 <- ggplot(f.tab2.m, aes(x=gender, y=value, fill=variable)) +  
  geom_bar(stat = "identity", colour="black")
```

명령의 결과는 아래의 왼쪽 그림과 같다. 아래의 오른쪽 그림은 가로 누적 막대그래프로 앞에서 본 것처럼 세로로 그린 다음 좌표만 "flip" 해주면 된다. 즉,

```
> piep3 + coord_flip()
```

으로 아래의 오른쪽 그림을 얻는다.



6 | 그룹별 다중 원그래프

앞에서 한 변수(mtcars에서의 실린더 개수별 빈도수)에 대한 원그래프와 도넛 그래프를 그려보았는데 만일 그룹별 원그래프를 각 그룹에 대해서 따로 그리지 않고 한 번에 그리고 싶으면 어떻게 할지 알아보자. 원그래프는 막대그래프를 그린 후 이를 coord_polar 함수를 사용하여 원그래프로 바꾸어 주었으므로 이 경우도 정확하게 같은 방법으로 그릴 수 있다.

그룹별 원그래프를 그릴 경우 원래 자료의 빈도수 합이 그룹별로 동일하지 않으면 원그래프를 함께 그릴 때 어떤 그룹은 360도를 다 채우지만, 또 다른 그룹은 그렇지 않을 수도 있다. 앞 절의 성별 선호 프로그램 자료도 남자는 100명이 조사되었지만, 여자는 90명만 조사되었을 경우 같은 중심을 가진 동심원을 그릴 때 남자 기준으로 그리면 여자는 10명이 모자라게 된다. 따라서 이런 경우에는 원그래프를 절대 빈도가 아닌 상대빈도(비율)로 계산하여 그리는 편이 좋다.

성별에 따라 좋아하는 프로그램의 비율을 계산하기 위해 다음과 같이 해보자. 먼저 남자인 자료의 인덱스값과 여자인 자료의 인덱스 값을 다음과 같이 얻고(남자 1, 3, 5, 여자 2, 4, 6이다)

```
> male.idx <- which(f.tab2.m$gender == 'male')  
> female.idx <- which(f.tab2.m$gender == 'female')
```

성별에 따른 합을 각 범주(남자, 여자)의 빈도로 나눈 '성별 상대비율'을 구하면 다음과 같이 얻을 수 있다.

```
> f.tab2.m$prop[male.idx] <- f.tab2.m$value[male.idx] / um(f.tab2.m$value[male.idx])  
> f.tab2.m$prop[female.idx] <- f.tab2.m$value[female.idx] / |sum(f.tab2.m$value[female.idx])
```

성별 선호 프로그램의 비율은 계산하여 f.tab2.m의 prop라는 변수에 저장하였으며, 필요하면 위에 계산에 100을 곱하여 백분율을 사용하여도 된다.

이제 원자료인 절대 빈도를 동심원 원그래프로 성별에 따라 그려보면 먼저 막대그래프를 앞 절의 그룹별 막대그래프 설정과 동일하게 한 후 마지막에 coord_polar 함수를 호출하면 된다. 즉, 원자료인 절대 빈도를 그룹별 원그래프로 얻으려면 y에 value를 설정하고 다음과 같이 명령한다.

```
> ggplot(f.tab2.m, aes(x=gender, y=value, fill=variable)) +  
  geom_bar(stat = "identity", colour="black") +  
  coord_polar("y")
```

위 명령의 결과로 아래의 왼쪽 그림이 얻어지며 첫 두 줄만 명령하면 그룹별 누적 막대가 얻어지고 이

를 극좌표로 변경하여 원그래프를 얻는다. 이 그림에서 안쪽 원그래프는 여자인 경우의 원그래프, 바깥쪽은 남자인 경우의 원그래프(내지는 도넛 그래프)인데 여자의 전체 빈도수가 90, 남자는 10이기 때문에 차이인 10만큼 여자 그래프는 비어 있다.



이제 비율을 사용하려면 매개변수 y의 값만 절대 빈도인 value 대신 상대빈도인 prop로 설정하면 된다. 즉,

```
> ggplot(f.tab2.m, aes(x=gender, y=prop, fill=variable)) +
  geom_bar(stat = "identity", colour="black") +
  coord_polar("y")
```

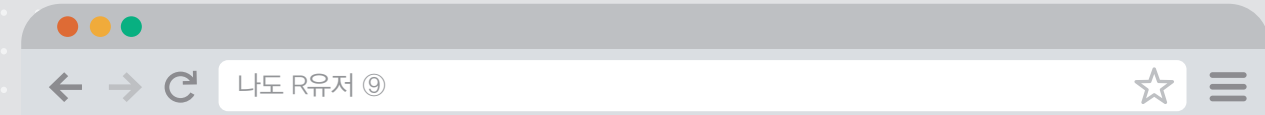
로 위의 오른쪽 그림을 얻게 되고 이 경우 두 그룹 모두 합이 1이므로 두 그룹의 원그래프가 모두 차 있음 알 수 있다. 그래프의 바깥쪽 숫자를 확인하면 더 확실하게 알 수 있다.

7 | 막대그래프의 응용 - 인구 피라미드

인구 피라미드는 성별, 연령대별 막대그래프를 말한다. 예를 들어 우리나라의 2017년의 성별, 연령별 인구 추정값은 다음과 같은데 이를 인구 피라미드로 그려보자. 데이터 프레임 pop10.2017에 이 자료가 저장되었다면 다음과 같이 값을 확인할 수 있다.

```
> pop10.2017
```

	age	male	female
1	0-9	2354265	2227092.5
2	10-19	2909993	2676221.0
3	20-29	3526732	3169220.0
4	30-39	3850486	3688221.0
5	40-49	4418589	4297638.0
6	50-59	4155589	4092335.0
7	60-69	2502779	2648484.5
8	70-79	1369560	1804078.5



9	80-89	404486	854029.5
10	90+	35337	127837.0

인구 피라미드에서 중요하게 보아야 할 점은 대개 가로막대를 사용하고 성별 인구를 중앙의 중심으로 할 경우 빈도수가 많으면 서로 반대 방향으로 막대의 길이가 길어진다는 것이다. 이에 착안해보면 두 성별 중 하나의 빈도수는 음수로 변환하여 막대를 그리면 될 것이라는 점이다. 이를 위해 앞에서 본 reshape2 패키지의 melt 함수를

```
> df1 <- melt(pop10.2017, measure = c('male', 'female'), variable.name = 'gender',
  value.name='pop')
```

로 사용하여 인구(pop)를 하나의 열로 갖는 데이터 프레임 df1을 만들면 다음과 같다(공간을 줄이기 위해 자료 중 성별 두 줄만 보임).

```
> df1
```

	age	gender	pop
1	0-9	male	2354264.5
2	10-19	male	2909992.5
	:		
11	0-9	female	2227092.5
12	10-19	female	2676221.0

이제 이 데이터 프레임에서 인구 중 하나의 성별에 대해서는 음수로 변환하기 위 다음의 단계를 거치면

```
> df2 <- df1 # 나중을 위해서 원래의 df1은 살려 두기 위해 df2로 복사
> idx_F <- which(df2$gender == 'female') # 여자인 인덱스
> df2$pop[idx_F] = (-1) * df2$pop[idx_F] # 여자 인구는 음수로
```

얻어진 데이터 프레임 df2는 다음과 같다.

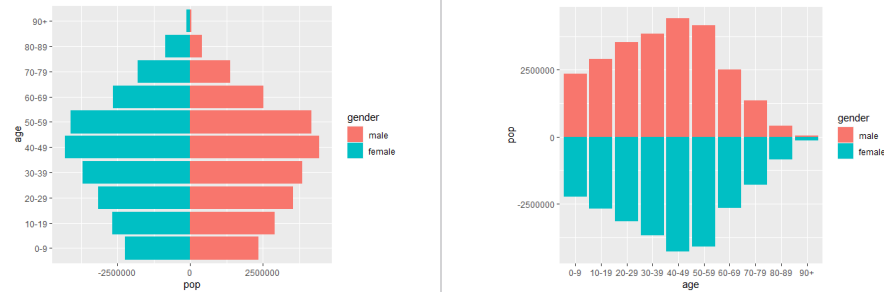
```
> df2
```

	age	gender	pop
1	0-9	male	2354264.5
2	10-19	male	2909992.5
	:		
11	0-9	female	-2227092.5
12	10-19	female	-2676221.0
	:		

이제 이 자료를 사용하여 빈도수 pop에 대해서 막대그래프를 그리되 x축의 좌표는 age를 사용하도록 설정한다. pop의 값은 막대의 높이로 설정한 후 최종적으로 coord_flip을 사용하여 가로로 막대를 그리도록 다음과 같이 명령하면

```
> ggplot(data=df2, aes(x=age, y=pop, fill=gender)) +
  geom_bar(stat="identity") +
  coord_flip()
```

아래의 왼쪽 그림과 같은 인구 피라미드가 된다.



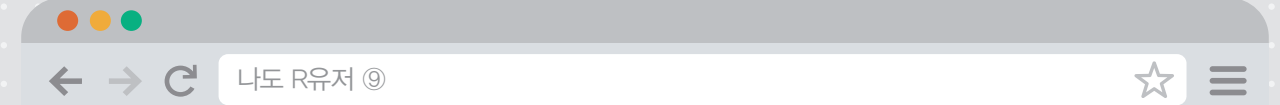
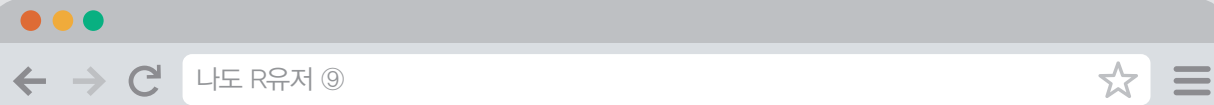
인구 피라미드를 기준으로 설명하였지만 한 범주에서 두 개 이상의 값을 갖는 누적 막대그래프에서 y의 값이 음수가 가능하므로 만일 인구자료가 아닌 다른 자료에서 위의 오른쪽 그림과 같은 형태의 그래프를 얻으려면 위의 명령과 같이 막대그래프를 생성한 후 coord_flip을 빼면 된다.

앞에서 df2를 만드는 과정을 설명하기 위해 세 줄에 걸쳐 df2를 만들었으나 df2를 만들지 않고 다음과 같이 한 줄로 양숫값을 음수로 바꾸어도 된다.

```
> df1$pop[df1$gender=='female'] = -df1$pop[df1$gender=='female']
```

이 명령 후에 위 ggplot 함수의 데이터 프레임 이름 설정에서 df2 대신 df1을 설정하여 ggplot 함수를 호출하면 동일한 결과를 얻는다.

또한 melt 함수를 사용하여 롱 포맷으로 자료를 변환하는 과정도 따로 실행하지 않고 ggplot 함수 내에서 ifelse 구문을 사용하여 여자 인구 값을 음수로 바꾸면 같은 결과를 얻을 수 있다. ifelse 함수는 한 줄로 if-else 구문을 작성하며(이를 보통 inline if-else라고 한다) 사용법은



```
ifelse(test, yes, no)
```

이며 test에는 조건문을, yes에는 test가 참일 때 실행할 명령을, no에는 test가 거짓일 때 실행할 명령을 설정한다. 즉,

```
> pp1 <- ggplot(data = df1, mapping = aes(x = age, fill = gender,
  y = ifelse(test = gender == "female", yes = -pop, no = pop))) +
  geom_bar(stat = "identity") +
  scale_y_continuous(labels = abs, limits = max(df1$pop) * c(-1, 1)) +
  coord_flip()
```

으로 원자료를 직접 사용하여도 같은 결과를 얻는다. 위의 명령에서 사용한

```
y = ifelse(test = gender == "female", yes = -pop, no = pop)
```

는 y 값 gender가 female이면 pop 값에 -1을 곱해준 값으로 바꾸고, 그렇지 않으면 pop 값을 그대로 사용한다.



이수동
 (주)인포그래픽스 대표 겸(사)한국인포그래픽협회 대표, sam9332@naver.com

데이터 인포그래픽 강좌 series 12

해외 통계 자료로 시장분석 보고서 만들기

INFOGRAPHICS



보고서를 작성할 때 다른 나라 통계 자료를 분석하여 활용하는 경우가 많다. 해외 자료를 활용할 때는 무엇보다 자료의 신뢰성이 중요하므로 정부가 공식적으로 발표한 자료가 있다면 우선순위로 활용하도록 하자. 또한 가장 최근 자료를 사용하는 것이 중요하므로 2018년 자료와 2019년 자료가 있다면 당연히 2019년까지 조사한 자료를 우선 활용한다. 이번 호에서는 원본에 있는 표의 해석방법을 통해 어떤 관점에서 표를 재편집하고 시각화할 수 있는지 등에 대해 알아보도록 하자.



해외 통계 자료 표 해석법

해외 자료를 분석할 때 가장 중요한 것은 자료의 원본을 읽어 보는 습관이다. 자료에 궁금한 점이 있다면 원본을 읽어 보고 해답을 찾아보는 것이 조사의 기본이다. 해외자료는 본인의 의도대로 해석하는 경우가 많으므로 작성자의 의도를 배제하고 분석한 후 그 데이터를 가지고 재편집하는 것이 좋다.

가. 외국인 노동자 연도별 추이

[그림1] 외국인 노동자 연도별 추이

人数	平成20年	平成21年	平成22年	平成23年	平成24年	平成25年	平成26年	平成27年
全国籍計	486,398	562,818	649,982	686,246	682,450	717,504	787,627	907,896
中国	210,578	249,325	287,105	297,199	296,388	303,886	311,831	322,545
韓国	20,661	25,468	28,921	30,619	31,780	34,100	37,262	41,461
フィリピン	40,544	48,859	61,710	70,301	72,867	80,170	91,519	106,533
ベトナム	※その他に含む	※その他に含む	※その他に含む	※その他に含む	26,828	37,537	61,168	110,013
ネパール	※その他に含む	※その他に含む	※その他に含む	※その他に含む	※その他に含む	※その他に含む	24,282	39,056
ブラジル	99,179	104,323	116,363	116,839	101,891	95,505	94,171	96,672
ペルー	15,317	18,548	23,360	25,036	23,267	23,189	23,331	24,422
G8+オース	39,968	43,714	46,221	50,321	51,156	53,584	57,212	61,211
その他	60,151	72,581	86,302	95,931	78,273	89,533	86,851	105,983

割合	平成20年	平成21年	平成22年	平成23年	平成24年	平成25年	平成26年	平成27年
中国	43.3%	44.3%	44.2%	43.3%	43.4%	42.4%	39.6%	35.5%
韓国	4.2%	4.5%	4.4%	4.5%	4.7%	4.8%	4.7%	4.6%
フィリピン	8.3%	8.7%	9.5%	10.2%	10.7%	11.2%	11.6%	11.7%
ベトナム					3.9%	5.2%	7.8%	12.1%
ネパール							3.1%	4.3%
ブラジル	20.4%	18.5%	17.9%	17.0%	14.9%	13.3%	12.0%	10.6%
ペルー	3.1%	3.3%	3.6%	3.6%	3.4%	3.2%	3.0%	2.7%
G8+オース	8.2%	7.8%	7.1%	7.3%	7.5%	7.5%	7.3%	6.7%
その他	12.4%	12.9%	13.3%	14.0%	11.5%	12.5%	11.0%	11.7%



일본 정부 기관인 후생노동성(www.mhlw.go.jp)에서 외국인 노동자 연도별 추이를 발표한 자료다. 공식 발표한 자료는 모두 2개의 표로 이루어져 있다. 첫 번째 표는 나라별 노동자 수를 연도별로 정리한 내용이고, 두 번째 표는 전체 노동자 수에서 나라별 노동자 수가 차지하는 구성비를 비율로 나타낸 자료다. 언뜻 보면 같은 자료라고 생각할 수 있지만 사실 어떤 기준(독립변수: Independent Variables)으로 자료를 편집하느냐에 따라 표가 전달하는 메시지가 달라질 수 있는 통계 자료이다.



1 나라별 외국인 노동자 수를 나타낸 표 분석

1 전체 분석

- 표를 분석할 때 가장 중요한 것은 표의 열과 행이 어떤 항목으로 이루어져 있는가를 확인하는 것이다. 특히 표는 1열이 가장 중요하다. 1열은 전국합계, 중국, 한국 등이 표시(국가별)되어 있다.
- 당연히 우리나라가 해당 자료를 활용하는 주체이므로 한국보다 상위에 있는 항목은 사용 1순위이며 가장 중요한 분석 대상이다. 따라서 일본 전체 합계, 중국 노동자수 추이는 중요한 비교 대상이 된다.
- 1행은 우리가 알아보기 힘든 '평성(平成)'이라는 단어가 들어 있다. '평성(平成)'은 일본이 쓰는 연호다. 일본은 1989년 1월 8일부터 '평성(平成)'을 사용 중이다. 예를 들어 '평성(平成) 20년'은 2008년을 말한다. 따라서 '평성(平成) 27년'은 2015년이다.
- 해외 자료에서 표의 1열과 행의 분석이 끝났다면 데이터 절반은 해석한 것과 다름없다. 해당자료는 2008년부터 2015년까지의 일본 내 '해외 노동자 취업자 수'를 조사한 데이터라는 것을 알 수 있다.

2 표가 지닌 다양한 의미 해석

우선 표를 분석할 때는 전체 흐름과 세부 항목이 지닌 의미를 순서대로 해석하는 것이 중요하다. 2015년 전국의 모든 외국 취업자 수는 90만7천 명이다. 2014년의 경우는 78만7천 명이니 약12만 명이 1년 동안 증가한 셈이다. 외국인 노동자 수가 가장 많은 국가는 역시 중국이다. 한국은 2번째로 많은 수를 기록하고 있다. 한국인의 취업자 수는 2015년 32만2천 명을 나타내고 있다. 중국은 한국보다 약 3배 정도 많다.

- 의미1 : 일본 내 외국인 노동자 수는 꾸준히 증가하고 있다. 1위는 중국, 2위는 한국이다.
- 의미2 : 2015년 한국인 노동자 수는 전년 대비 4천 명이 증가했지만 같은 기간 중국은 1만1천명 증가했다. 2008년부터 분석하면 한국은 꾸준히 증가하고 있지만 중국은 2011년~2012년 사이 오히려 감소했다.

3 표의 재구성

[그림2] 연도별 일본 내 외국인 노동자 수 추이

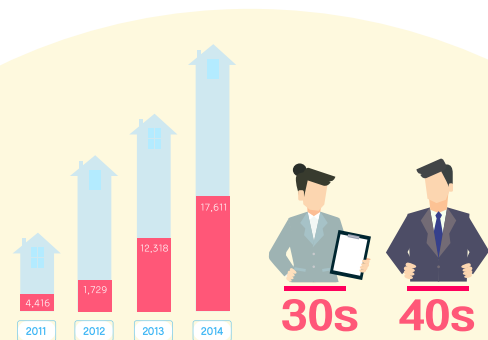
외국인 노동자수	2008	2009	2010	2011	2012	2013	2014	2015
전국 합계	486,398	562,818	649,982	686,246	682,450	717,504	787,627	907,896
한국	20,661	25,468	28,921	30,619	31,780	34,100	37,262	41,461
중국	210,578	249,325	287,105	297,199	296,388	303,886	311,831	322,545

* 다른 나라의 데이터는 제외

- 1열에서는 전국합계, 한국, 중국 항목만 사용하고 나머지는 제거하며 표에 이를 명시한다.
- '평성'의 연호는 '연도'로 변경하며, 한국의 데이터를 '전국합계' 바로 아래에 위치하도록 순서를 중국과 변경한다. 메시지의 핵심은 한국인 일본 내 취업자 수를 파악하는 것이므로 표에서 한국을 중국보다 위에 표현하는 것이 좋다. 표를 재 편집하는 경우 서열이 매우 중요하다.
- 표에서는 폰트 굵기, 폰트 크기로 강조할 부분을 부각시킬 수 있다.

2 그래프 또는 시각적 표현

해당 표를 그래프로 전환한다는 것은 연도별 흐름을 한눈에 살펴보기 위함이 크다. 다만 모든 연도를 넣을 것인지 아니면 최근 5년간의 흐름만을 보여줄 것인지에 따라 사용하는 그래프가 다를 수 있다. 2008~2015년까지의 8개년 흐름을 보려면 선그래프나 연속하는 막대그래프로도 표현 가능하다. 이때 전국, 한국, 중국 등 그래프 3개가 한 쌍을 이뤄 8년간을 나타내야 하므로 총 24개 막대를 그려야 한다. 또 다른 방법은 일본 전국합계와 중국은 제거하고, '한국의 연도별 노동자 수 추이' 데이터만 추출해서 보여줄 수도 있다. 이는 '한국인 노동자 수가 꾸준히 증가'한다는 메시지를 부각하는 데 유리한 정보기획 방법이다.



KEY NOTE

- 1 전국합계, 한국, 중국까지 들어 있는 표를 시각화해 볼 수 있다.
- 2 8년간의 흐름을 선 그래프로 나타내고 증가 흐름을 비교해 볼 수 있다.
- 3 전국합계와 한국 또는 한국의 데이터만 추출해서 그래프로 나타내 볼 수 있다.
- 4 전체 8년간의 흐름을 그래프와 표가 아닌 한국인 노동자를 상징하는 아이콘 + 숫자(2012년 vs 2015년 2개 연도만 비교)로 만들어 볼 수 있다.

3 외국인 노동자 수를 비율로 나타낸 분석

1 전체 분석

표를 종합적으로 분석해 보자. 데이터는 전체 외국인 취업자에서 차지하는 비율이며, 1행은 2008년부터 2015년까지의 연도를 말한다. 중국은 2008년부터 꾸준히 40% 이상의 비율이었으나 2014년 처음으로 30%대로 하락했다. 특히 2015년 큰 폭으로 하락했다. 이 밖에 2012년부터 3년간 꾸준히 노동자 수 비율도 하락하고 있음을 알 수 있다. 한국은 2008년부터 2015년까지 꾸준히 4%대 중반을 유지하고 있으며, 중국 + 한국 등 2개 국가의 비율을 합하면 절반 가까운 비율임을 추가로 파악할 수 있다.

2 표가 지닌 다양한 의미 해석

- 의미1 : 일본 내 한국인 노동자 수 비율은 지난 8년간 '4%'선을 유지하고 있다.
- 의미2 : 미미한 비율이지만 한국은 2013년까지 상승하다 이후 0.1%p씩 꾸준히 하락하고 있다.
- 의미3 : 중국은 2014년 처음으로 40%대가 무너졌으며, 한국은 큰 변화가 없다.

3 표 편집

2번째 표에서도 1열은 한국, 중국 항목만 표시했다. '평성'의 연호는 '연도'로 변경하였으며, 한국을 중국 위에 표시하였다.

[그림3] 일본 전체 외국인 노동자 수 대비 한국, 중국 노동자 비율

노동자 비율	2008	2009	2010	2011	2012	2013	2014	2015
한국	4.2	4.5	4.4	4.5	4.7	4.8	4.7	4.6
중국	43.4	44.3	44.2	43.3	43.4	42.4	39.6	35.5

4 그래프 또는 시각적 표현

한국과 중국의 상관관계 흐름을 보여 주기 위해 그래프를 선택할 수 있다. 우선 두 개의 2중 막대그래프로 연속하는 높낮이를 나타낼 수 있으나 8년간의 흐름을 보여주는 데는 선그래프가 유용하다. 이 밖에 한국과 중국의 흐름이 바뀌는, 즉 '변곡점'이 되는 2013년 이후부터의 독립변수만 가지고 그래프를 나타낼 수 있다. 모든 연도를 다 사용하지 말고 전달하려는 '메시지 목적'에 맞게 'X축 독립변수'를 선택적으로(2013~2015년) 활용할 수 있다.

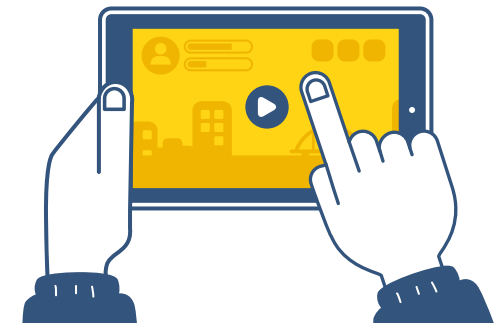
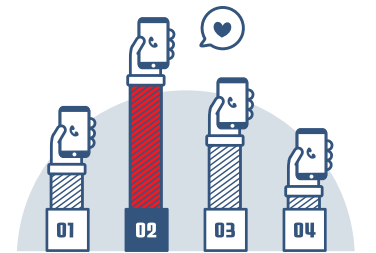
▶ 전문가 TIP!

통계 데이터의 고수가 되려면 X, Y축을 유기적으로 선택할 수 있어야 한다. 이는 통계를 어떻게 바라 봐야 하는가의 문제이기도 하지만 사람들을 설득할 수 있는 커뮤니케이션 방법을 스스로 결정할 수 있다는 뜻이기도 하다. '빅데이터'에서도 데이터의 가치를 찾을 수 있어야 하고 이를 다시 재해석할 수 있는 통찰력을 지니는 것이 중요하다.

KEY NOTE

시각화(디자인) 방법

- 1 '표'로 나타낼 경우 국기, 폰트 조정, 배경색 등을 통해 표 디자인을 할 수 있다.
- 2 8년간 추이를 선 그래프로 나타내고, 중요한 변곡점을 표시할 수 있다.
- 3 3개년 정도만 추출해서 그래프로 표현이 가능하다.
- 4 아이콘(노동자 또는 의자 등) + 지표(숫자)를 가지고 2015년 vs 2012년 등 2개 연도만 비교하여 시각화해 볼 수 있다.





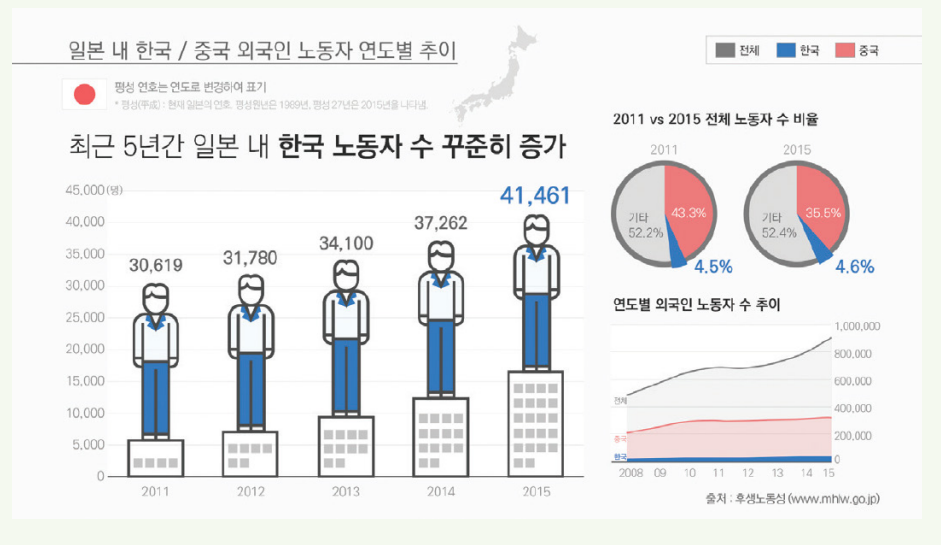
나. 인포그래픽 시각화 및 디자인

시각화를 통해 표와 그래프를 볼때 가장 눈에 띄는 부분이 어디일까? 아무래도 가장 크고 진한 부분일 것이다. 많은 양의 통계자료를 기획할 때 전달하려는 의도에 따라 재편집하여 여러 해석을 만들어낼 수 있다면, 시각화 및 디자인 단계에서도 기획할 때 재편집한 내용을 반영한 표현이 가능하다. 여러 가지 복잡한 그래프를 한 장(화면)에 구성할 때 내용에서 통일시킬 수 있는 부분과 강조해야 할 부분이 무엇인지를 먼저 생각해보면 좀 더 수월하게 디자인을 할 수 있다.

▶ 디자인 포인트

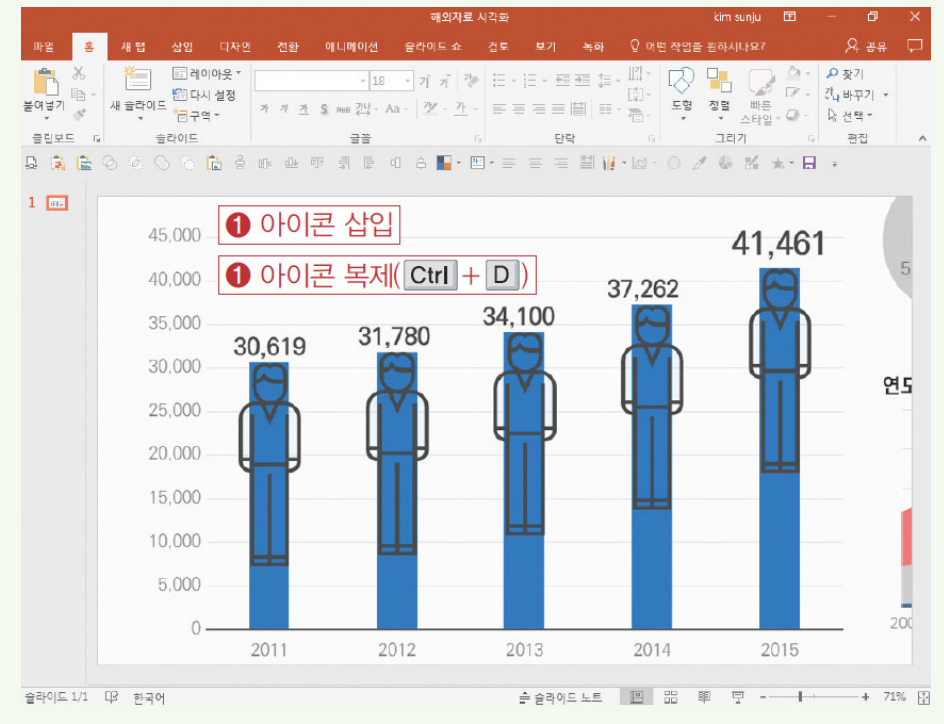
- 1 일본 내 여러 나라의 노동자 수를 다룬 자료이지만, 일본에서 한국의 노동자 수가 지속적으로 증가한다는 부분에 중점을 두고 작성한 예다. 다른 나라 데이터는 제외하고 전체 합계와 상위 국가인 중국과 한국의 데이터를 비교한 후 추이 그래프를 통해 전체적인 흐름을 보여주는 형태로 구성하였다.
- 2 기획단계에서 뽑아낸 내용의 우선순위를 정하고 강조해야 할 데이터를 기준으로 시선의 흐름에 따라 배치한다. 비교 대상 없이 한국의 데이터를 별도로 보여줌으로써 증가율을 실감할 수 있도록 표현한다. 그래프의 크기를 다르게 하여 강약을 조절하고 불필요한 부분을 최소화해 간결하게 나타내는 것이 좋다.
- 3 각 그래프마다 범례가 같다면, 전체적으로 통합해주는 것도 좋다. 공간을 활용하여 상위에 범례를 표시해준다.
- 4 해외자료를 인용할 때 사전에 알아야 할 정보나 사실을 간략히 넣어주는 것도 좋은 방법이다. 이와 관련하여 폰트를 선택할 때 다국어 지원이 가능한 폰트를 사용한다

[그림4] 일본 내 한국, 중국 외국인 노동자 연도별 추이



위의 최종 결과물에서 중요한 과정은 사람을 막대그래프 대신 노동자 이미지로 시각화 해 그래프 속에 넣는 과정이다. 연습을 통해 일반적인 세로형 막대그래프 대신 PPT와 엑셀을 활용해 제작이 가능하다.

[그림5] 시각 그래프 제작 과정



제작과정에서 사용한 폰트의 크기, 종류, 그림내 컬러 값을 정리하면 그림 6과 같다.

[그림 6] 텍스트 제작에 필요한 글꼴 및 색채 예시

텍스트	글꼴/글꼴 크기	글꼴 색(RGB)
평성 연호는~	KoPub돋움체 Medium/12.5	130/130/130
평성 설명	KoPub돋움체 Light/10.5	185/185/185
출처	KoPub돋움체 Medium/12.5	90/90/90

특히 폰트의 저작권은 중요한 부분이라 상업용 무료 폰트를 사용해야 한다. 인터넷에서 검색을 하거나 카페 등에서 무료라고 다운로드 하라는 메시지가 있어도 일단 의심하고 먼저 저작권 가이드를 확인하는 습관을 가져야 한다.

표와 그래프는 처음 만드는 사람의 주관적 견해가 매우 강한 표현방법이다. 색을 넣고, 글자를 강조하면 그 쪽으로 시선이 움직이기 때문이다. 비대면 시대 모니터로 정보를 보고 또 자료를 찾고 하는 흐름이 대세가 됐다. 중요한 것은 원래의 데이터를 인용할 때 어떤 의미가 숨어 있고 이를 어떻게 해석해야 하는지 스스로 중심을 잡고 비판적인 사고력을 갖는 것이다

출처
데이터인포그래픽 디자인제작실무 (이수동, 김선주 예문사)



이수동: sam9332@naver.com
(주)인포그래픽웍스 대표 겸 (사)한국인포그래픽협회 대표(현) 국내 인포그래픽 및 데이터 시각화 개척자이자 전문가다. 인포그래픽기획실전략(베스트셀러), 인포그래픽인사이드57(중국 인민체신출판사 독점계약), 데이터인포그래픽 디자인제작실무(베스트셀러), 4차 산업혁명 멘토링(세종도서 선정) 등을 포함해 다수의 책을 집필했다.

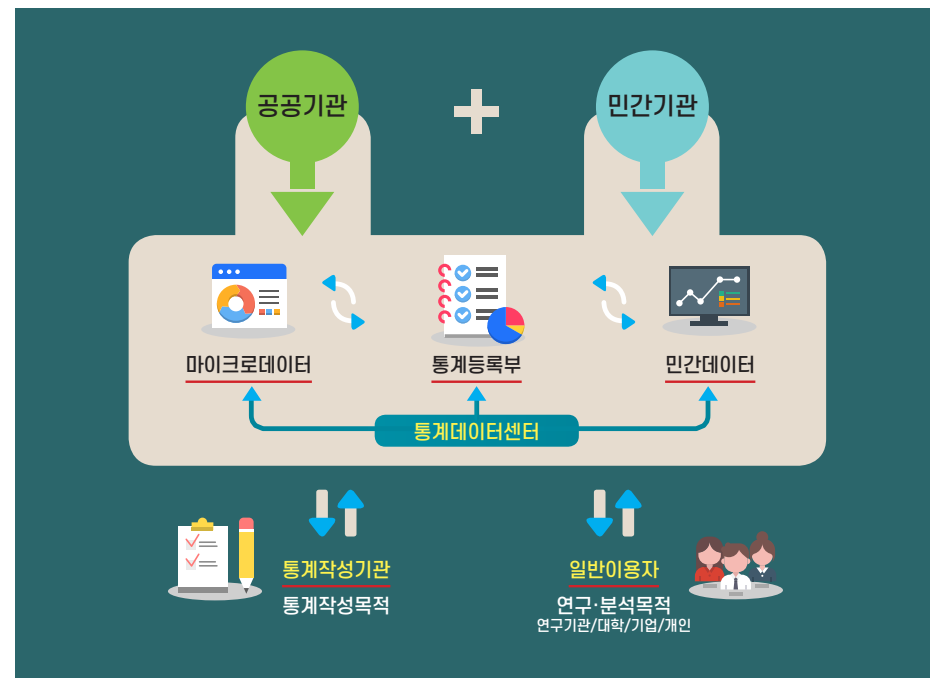
통계데이터대구센터 개소! 대구·경북지역 경제발전의 디지털 초석을 마련

김옥경

통계청 통계데이터기획과 사무관

4차산업혁명 이후 AI(인공지능)와 빅데이터 시대로 발전됨에 따라 이에 필요한 데이터 수요가 점차 증가하는 추세다. 공공분야에서는 데이터 전문기관인 통계청이 주체가 되어 행정통계자료와 통계기초자료 등 공공데이터와 통신데이터 등 민간데이터를 구축하여, 국가정책 수립 및 기업경영 의사결정, 개인 자영업자의 창업결정 등의 이용자 니즈를 충족할 수 있도록 노력하고 있다.

데이터 활용 활성화 및 이용자 만족도 제고를 위해 통계청이 핵심적으로 노력하고 있는 부분은 국민 누구나 통계데이터센터에서 쉽고 편리하게 데이터를 연계·활용할 수 있는 환경을 구축하여 모든 이용자가 만족하는 최상의 서비스를 제공하는 것이다.



이의 일환으로 통계청은 지역별 이용자의 접근성과 편의성을 향상시키고자 2019년 3개 센터(대전, 서울, 부산)에서 시작하여 데이터서비스의 사각지대였던 거점지역을 중심으로 지역센터를 점차 확충하여 서비스를 확대해 왔다. 2020년에는 호남지역 이용자들을 위한 광주센터, 2021년에는 대구·경북 지역 이용자 지원을 위한 대구센터를 개소하여 지역민의 데이터 활용을 지원하고 지자체 및 지역 연구센터, 대학교와의 연계·활용을 위한 기반을 마련하였다.

- 1 통계데이터대구센터 개소식
- 2 통계데이터대구센터가 위치한 SW융합테크비즈센터 외부 전경
- 3 통계데이터센터 실내 전경



통계데이터 대구센터는 대구·경북 지역 소재의 정부·지자체·연구기관·기업 등 다양한 데이터 수요층이 쉽게 접근할 수 있도록 대구광역시 수성구 SW융합테크비즈센터 내에 위치하고 있으며, 15석 규모의 분석 플랫폼을 갖추고 있다.

이처럼 통계데이터 대구센터가 신설됨으로써 기존 대구·경북 지역 내 공공기관, 대학교, 연구기관 등의 이용자들이 데이터 분석을 위해 타 지역으로 가야만 했던 불편함을 해소하고 해당 지역 내에서 직접 다양하고 민감한 공공 및 민간데이터의 연계·활용을 통하여 지역경제 활성화를 위한 다양한 연구를 시도할 수 있게 되었다.

통계데이터 대구센터가 위치해 있는 SW융합테크비즈센터는 대구공항에서 15분 거리에 있으며, 경부고속도로 동대구JC 및 수성IC와 인접하여 이용자들이 쉽게 접근하여 이용할 수 있는 지리적 장점을 가지고 있다. 또한 대구광역시에서 운영하고 있는 대구빅데이터활용센터가 통계청의 통계데이터대구센터와 같은 장소에 위치하고 있어 두 센터간 유기적인 상호 협력 관계를 통해 데이터 활용의 시너지효과가 더욱 커질 것으로 기대하고 있다.



지하철	<ul style="list-style-type: none"> 지하철 2호선 대공원역 4번출구(도보 20분)
버스	<ul style="list-style-type: none"> 연호지 정거장(609번, 903번, 수성3번 15분정도 소요) 경기장네거리(403번, 609번, 939번 10분정도 소요)
차량	<ul style="list-style-type: none"> 서울방향 : 경부고속도로 > 천안JC > 남이JC > 대전JC > 김천JC > 구미 > 금호JC > 동대구JC > 수성IC 강원방향 : 중앙고속도로 > 만촌JC > 제천 > 서안동 > 군위 > 금호JC > 동대구JC > 수성IC ▶경부고속도로 > 양산 > 언양JC > 경주 > 영천 > 동대구JC > 수성IC ▶대구부안고속도로 > 밀양 > 청도 > 수성IC

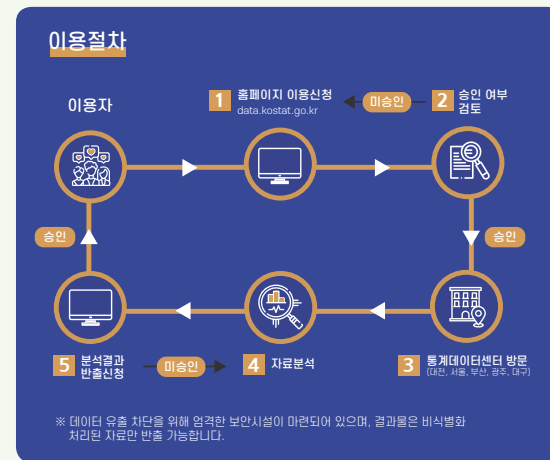


통계데이터 대구센터는 가장 최근에 개소된 만큼 단기간에 이용 활성화를 끌어올릴 수 있도록 통계데이터센터가 보유하고 있는 자료와 분석플랫폼을 개소일부터 올 연말까지(4월 15일 ~ 12월 31일) 무료로 제공하고 있다. 이는 이용수수료 납부에 대한 부담을 잠시 제거하여 이용자들이 데이터 이용에 대한 강점을 직접 체감함으로써 이용활성화를 도모하고자 함이다. 이 외에도 데이터 이용 및 분석방법 등에 대한 교육프로그램을 운영하고, 통계데이터센터 교육홈페이지를 통하여 교육컨텐츠를 무료 제공함으로써 보다 폭넓은 이용자층으로 확대하고자 노력하고 있다.

현재 통계청은 데이터 이용자들에게 이원화된 서비스를 제공하고 있어 다소 이용 편의성 및 활성화를 저해하고 있는 것이 사실이다. 즉, 행정자료 중심의 데이터는 통계데이터센터에서 제공하고, 이외 일반 통계자료인 마이크로데이터는 마이크로데이터이용센터(RDC, Research Data Center)에서 제공하고 있다. 이에 2021년 적극행정 정책추진의 일환으로 통계데이터센터(SDC, Statistics Data Center)이용자들이 마이크로데이터서비스(MDIS, MicroData Integrated service)를 함께 이용할 수 있도록 7월초까지 관련 시스템을 개선하여 오픈할 예정이다.

- 마이크로데이터서비스(MDIS)의 제공 데이터는 통계조사 원자료에서 조사·입력오류 등을 수정한 개별단위(개인, 가구, 업체 등)자료로 많은 연구기관에서 보편적으로 이용하고 있는 자료이다.
- 마이크로데이터이용센터(RDC)의 제공 데이터는 마이크로데이터서비스(MDIS)에서 제공하는 데이터 중 가장 넓은 범위의 자료가 제공되며, 분석결과에 대해 승인을 통해 반출 서비스까지 제공되는 센터 이용 서비스이다.

통계데이터센터는 아래의 절차를 통하여 이용할 수 있다.



통계데이터센터에서 제공하는 공공 및 민간 데이터를 활용하여 통계생산, 정책결정을 위한 연구 등 다양한 분야의 분석을 시도할 수 있도록 분야별 데이터를 제공하고 있다.

특히, 최근 전 세계적으로 '코로나19' 발생에 따라 이와 관련된 데이터를 활용한 정책 수립이 사회적 이슈가 되었다. 예를 들면, 코로나19로 인한 경제성장 규모의 변화와 유동인구 변화 등이 이에 해당된다. 통계데이터센터에서는 이러한 변화에 맞춰 코로나19 방문인구정보 등과 같은 데이터를 통계데이터센터를 통해 빠르게 제공하였다.

코로나19 방문인구정보 데이터는 인구 이동을 중심으로 거주 행정동을 제외한 타 행정동에서 30분 이상 체류한 인구에 대한 행정동 단위의 집계자료로 유입지 시군구자료와 목적지 시군구자료, 연령대, 성별 정보들이 포함되어 있어 코로나19와 관련된 연구 및 정책수립에 기초자료로 활용할 수 있는 데이터이다.

이와 같은 코로나19와 관련된 데이터 외에 통계데이터센터에서는 '유통고객 구매상품 정보' 데이터와 평균소득·소비에 관련된 데

이터 등 다양한 민간데이터를 신규로 추가하여 연계·활용이 가능한 환경으로 고도화하는데 지속적인 힘을 쏟고 있다.

다음은 전국의 통계데이터센터에서 현재 제공하고 있는 데이터 종류에 대한 정보를 간략한 표로 작성한 내용이다.

행정통계자료			
구분	부문	자료명	시계열
행정통계자료	경제·사회	기업통계등록부DB(대표자, 사업자 기준)(연간)	2010~2019년
		기업통계등록부DB(대표자, 사업자 기준)(분기)	2018.4분기~2020.4분기
		사업장 기초DB(4대보험)	2009~2019년
		사업장 기초DB(법인)	2008~2019년
		종사자-기업계 연계DB	2015~2019년
	인구·가구	육아유치 사용현황DB	2015~2019년
		인구DB	2015~2019년
		가구DB	2015~2019년
		주택DB	2015~2019년
		농업DB	2015~2019년
농림·어업	임업DB	2015~2019년	
	어업DB	2015~2019년	

민간자료			
구분	부문	자료명	시계열
민간자료	인구·가구	성·연령별 유동인구(전국)	2015.1월~2020.12월
		시간대별 유동인구(전국)	2015.1월~2020.12월
		요일별 유동인구(전국)	2015.1월~2020.12월
		(코로나19)방문인구정보(전국)	2019.1월~2020.12월
		(코로나19)방문인구정보(유입지)(전국)	2019.1월~2020.12월
		성연령별 카드매출기(전국)(국내)	2018.10월~2020.12월
	경제·사회	성연령별 카드매출기(전국)(해외)	2018.10월~2020.12월
		신용통계정보(대출)(전국)	2018.10월~2020.12월
		신용통계정보(소속)(전국)	2018.10월~2020.12월
		신용통계정보(카드소비)(전국)	2018.10월~2020.12월
		인구이동정보(전입인구)(전국)	2019년 4분기~2020년 4분기
		인구이동정보(전출인구)(전국)	2019년 4분기~2020년 4분기
		인구이동정보(출근인구)(전국)	2019년 4분기~2020년 4분기
		구매상품정보(전국)(오프라인)	2019.12월~2020.11월
		모바일상품권 거래데이터(전국)	2018년
기초정보	온라인가격정보	2017.1월~2021.3월	
	소재지역별 1,2	1 : 2018년 2 : 2019년~(개편)	

통계기초자료			
구분	부문	자료명	시계열
통계기초자료	사업체	경제제조업조사, 경제용조사, 전국사업체조사 등	
		인구주택총조사, 사망원인통계, 인구동향조사 등	
	가구	가계금융복지조사, 경제활동인구조사 등	
		농림어업총조사, 농가경제조사 등	

통계청은 향후에도 지역센터 확충, 데이터 확보 및 시스템 등 지속적으로 분석환경을 개선하여, 이용자가 쉽고, 편리하게 통계데이터센터를 이용할 수 있도록 지원할 계획이다.



KEIS
 Korea Employment
 Information Service



INTERVIEW

김동규

한국고용정보원 미래직업연구팀 연구위원 / 경영학 박사

인공지능(AI) 및 빅데이터 시대,
 직업 세계는 어떻게 변화할까?



인공지능과 빅데이터가 무엇이고, 왜 중요한가요?

20세기 후반부터 시작된 정보화 혁명(3차 산업혁명)이 컴퓨터와 인터넷을 기반으로 시작되었다면, 21세기 초부터 시작된 초지능화 혁명(4차 산업혁명)은 인공지능과 빅데이터가 핵심 기반이 되었다고 볼 수 있습니다.

예를 들면, 로봇이 사람의 음성에 반응하여 움직이거나 자동차가 자율주행을 하려면, 음성이나 영상 이미지에 대한 방대한 데이터 축적과 머신러닝, 딥러닝 등의 인공지능 학습이 함께 필요합니다. 인공지능은 방대한 데이터를 통해 스스로 학습할 수 있고, 인공지능 기술은 빅데이터에서 유용한 정보를 발굴하는데 필요한 수단입니다. 즉, 인공지능과 빅데이터는 서로 긍정적 영향을 주는 한 세트와 같습니다.

물론, 두 기술의 발전에는 5G, 클라우드 컴퓨팅, 사물인터넷 등의 기술이 인프라로서 큰 역할을 하고 있습니다.

인공지능과 빅데이터가 현실에서 어떻게 활용되고 있나요?

현재, 인공지능과 빅데이터는 제조생산, 의료, 교통, 도시, 금융, 마케팅 등 다양한 영역에서 활용되어 두뇌 역할을 하고 있습니다.

인공지능과 빅데이터의 발전으로 협동로봇과 스마트팩토리, 자율주행차, 드론, 개인맞춤형 제조, 디지털 마케팅, 스마트시티와 스마트병원 등이 가능해졌다고 해도 과언이 아닙니다. 각 분야의 디지털 전환과 혁신을 이끌고 있는 핵심기술이 인공지능과 빅데이터라고 할 수 있습니다.



인공지능, 빅데이터 등의 혁신기술을 중심으로 4차 산업혁명 또는 디지털 전환이 경제·사회 전 분야에 큰 영향을 주고 있는데요, 직업세계 또는 노동시장에는 어떤 변화가 있을까요?

경제·산업은 물론 사회·문화 전반에 걸친 초지능화와 디지털화 그리고 자동화와 무인화는 직업세계와 노동시장에 큰 변화를 주고 있는데, 크게 네 가지로 말씀드릴 수 있을 것 같습니다.

첫 번째는 직업 구조와 일자리 구조의 변화입니다. 새로운 기술의 등장과 융·복합화는 새로운 직업과 일자리를 창출합니다. IT와 엔지니어링, 연구개발 등의 분야에서 많은 전문직 일자리가 생겨나고 있습니다. 특히 SW개발이나 인공지능, 빅데이터, 클라우드 컴퓨팅 등의 첨단 분야에서는 인력 공급이 산업계 수요를 따라가지 못하는 실정입니다. 반면에 제조·생산기능직이나, 판매원 및 계산원, 일반사무직 등의 직종에서는 일자리 감소가 불가피할 것입니다.

그리고 제조·생산 분야에서 생산 품종에 따라 희비가 엇갈릴 것입니다. 예를 들어, 전기차와 수소차가 내연기관차를 대체하는 속도가 가파르는데, 그렇게 되면 전기구동모터나 배터리팩, 고전압 전장품, 수소탱크, 자율주행 전장부품 등의 부품생산 업체에서는 일자리가 새로 생겨나겠지만, 반면에 엔진, 변속기 등의 부문에서 많은 일자리가 사라질 것입니다. 일자리 구조 변화에 미리 대응해야 하는 이유입니다.

두 번째는 직업 종사자들이 하는 일, 즉 직무(task)의 변화가 커지고 있는 것입니다. 기술발전은 근로자의 직무를 변화시키는 것이 당연하지만, 그 변화의 폭과 속도가 이전보다 더 크다는 뜻입니다. 변화의 속도와 폭이 크면 개인 차원에서 대응이 쉽지 않습니다. 앞으로 어떤 직종에서든 자신의 일자리를 유지하려면 도메인 날리지(domain knowledge)에 더하여 데이터와 SW, 디지털 도구에 대한 지식과 기술이 필요하게 될 것입니다. 또는 수행하던 직무가 아예 사라져 전직을 위한 직업훈련을 받아야 할지도 모릅니다. 정부와 기업, 노조가 협력하여, 근로자들에게 교육훈련 기회를 선제적으로 충분히 제공해야 하는 이유입니다.

세 번째는 인공지능, 빅데이터 등 첨단 분야일수록 직무가 더욱 전문화, 세분화되면서 특정 직무에 전문성을 갖춘 인재의 수요가 증가한다는 것입니다. 한편으로는 직무가 파편화되고 생산공정과 업무 과정이 휴먼스케일을 넘는 수준으로 매우 복잡화하면서 전체 시스템을 통찰하고 통제할 수 있는 고급 인재에 대한 필요성은 더욱 커질 것입니다.

마지막, 네 번째는 인사관리 변화입니다. 정규직, 정규근로시간, 직장출근 등 전통적인 고용형태와 일하는 방식이 더 이상 일반적이지 않게 될 것입니다. 플랫폼근로자, 프리랜서 등 비전형 고용형태, 시간선택제 근무와 재택근무 등 다양한 근로형태가 더욱 확산될 것입니다. 이와 같은 현상은 COVID-19에 의해 더욱 촉진되고 있습니다.

기업의 인력 채용에 있어서도 일괄 공채나 학벌 중심의 채용은 크게 줄어들 것입니다. 글로벌화로 기술발전 속도가 빠르고 기업 간 경쟁이 치열하기 때문에 기업들은 실제 업무에 곧바로 투입할 수 있는 실무형 인재를 채용하길 원합니다. 따라서 청년들에게 직무경험과 직업훈련을 제공하는 것은 기업의 인사관리 전략이 되어야 하는 동시에 사회적 책무로서 인식되어야 할 것입니다.

인공지능과 빅데이터 분야의 유망직업에 대해 소개 부탁드립니다.

어떤 것이 유망직업인지에 대한 다양한 의견이 있습니다만, 통상, 성장산업에 속하면서 일자리가 증가하고 높은 보수를 받는 직업을 유망직업이라고 합니다.

세계경제포럼(WEF)은 2020년에 발표한 보고서 「Jobs of Tomorrow Mapping Opportunity in the New Economy」를 통해 데이터 및 인공지능, 엔지니어링과 클라우드 컴퓨팅 분야 등 7개 고성장 분야에서 '신성장 직업(Emerging Jobs) 96선'을 발표한 바 있습니다. 그 중 몇몇 직업을 소개하면 다음과 같습니다.



- ▶ **인공지능전문가(Artificial Intelligence Specialist)** : 인공지능 알고리즘 또는 응용프로그램을 개발하는 일을 한다.
- ▶ **데이터 사이언티스트(Data Scientist)** : 머신러닝 등의 인공지능 기술을 활용하여 빅데이터를 분석하고 유용한 정보(서비스 이용자의 행위 예측 등)를 추출하여 제품 또는 서비스 개선에 활용하는 일을 한다.
- ▶ **데이터 엔지니어(Data Engineer)** : 데이터 분석에서 사용되는 혹은 필요한 데이터를 사용하기 좋도록 정제하여 서버에 수집·저장하고 유지·관리하는 일을 한다.
- ▶ **데이터 분석가(Data Analyst)** : 사업기획자나 서비스기획자가 인사이트를 얻도록 서비스들의 현황 데이터를 정제, 분석하고, 수치로 표현하여(시각화) 제공하는 일을 한다.
- ▶ **빅데이터 개발자(Big Data Developer)** : 빅데이터 기술을 사용하기 위해서 웹사이트나 서버, 데이터베이스 단의 환경을 구축하는 일을 한다. 빅데이터 분석 플랫폼을 개발·구축·운영한다.
- ▶ **비즈니스 인텔리전스(BI) 개발자(Business Intelligence Developer)** : 기업이 보유하고 있는 수많은 데이터를 정리·분석해 기업의 의사결정에 활용하는 일련의 프로세스인 BI 솔루션(타블로, 클릭뷰, 마이크로스트레티지, 오라클 OBIEE 등)을 개발하고, 컨설팅하는 일을 한다.
- ▶ **풀 스택 전문가(Full Stack Engineer)** : 웹사이트 또는 응용 프로그램의 프론트 엔드와 백 엔드 모두에서 작업하는 웹개발자 또는 엔지니어이다.
- ▶ **플랫폼 엔지니어(Platform Engineer)** : 플랫폼 아키텍처를 설계·구축·운영하는 일을 한다.
- ▶ **클라우드 엔지니어(Cloud Engineer)** : 클라우드 인프라(소프트웨어와 시스템)를 설계·구축·운영하는 일을 한다.



앞에서 현재 존재하는 유망직업에 대해 소개해 주셨는데, 앞으로 새로 등장할 유망직업들도 있을까요?

Cognizant(<https://careers.cognizant.com/>)의 직업미래센터(Center For The Future of Work)는 10년 내에 노동시장에 새롭게 등장할 유망직업들을 제시한 바 있는데, 현재 막 태동 단계에 있거나 앞으로 나타날 가능성이 높다고 생각되는 직업들을 몇 개 소개해 드리겠습니다.

- ▶ **인간-기계 협업 관리자(Man-machine Teaming Manger)** : 인간과 기계의 협업 증가에 따라 인간과 기계의 능력, 목표 등을 소통하고 조정하는 일을 한다. 새로 사용가능한 기술로 업그레이드 할 수 있는 작업, 프로세스, 시스템 및 경험을 식별하고, 새로운 접근 방식, 기술, 상호작용 및 구성을 바탕으로 인간과 기계가 함께 작업하기 위해 조정해야 하는 규칙을 설정한다.
- ▶ **IT 협업 환경 퍼실리테이터(Bring Your own IT Facilitator)** : 임직원들이 업무에 개인 도구들을 사용하는 'shadow IT' 현상이 증가함에 따라, 임직원들이 클라우드와 모바일을 기반으로 협업할 수 있는 환경을 만들어 주는 일을 한다.
- ▶ **재택근무 퍼실리테이터(Work from Home Facilitator)** : 재택근무를 원활히 수행할 수 있도록 필요한 원격 기술을 도입하여 디지털 협업 도구를 평가하고 마련하는 일을 한다.
- ▶ **알고리즘 편향성 감사(Algorithm Bias Auditor)** : 조직 내 모든 비즈니스 단위의 알고리즘이 공정하고 합법적이며 회사의 가치를 대표하는지 감사하고, 직원들이 이를 지킬 수 있도록 관련 지침과 규정을 수립하는 일을 한다.
- ▶ **사이버 재난 예보관(Cyber Calamity Forecaster)** : 사이버 불확실성에 대비하기 위해 사이버 재해를 모니터링, 탐지 및 예측하는 일을 한다.
- ▶ **기계 위험 관리자(Machine Risk Officer)** : 인공지능 기계의 행동이 의도하지 않은 결과를 초래할 경우 회사의 브랜드와 재정적 손실에 큰 악영향을 끼치기 때문에 인공지능 기계의 고장이나 오류로 인해 발생할 수 있는 잠재적 위험을 관리하는 일을 한다.
- ▶ **인공지능 비즈니스 개발 관리자(Artificial Intelligence Business Development Manager)** : 고객 및 파트너의 요구사항에 맞는 적절한 인공지능 서비스를 제안한다. 그리고 고객이 필요로 하는 인공지능 프로그램을 개발하고, 고객 맞춤형 인공지능 서비스와 프로그램 판매를 위한 영업 활동을 하기도 한다.
- ▶ **개인 정보 중개자(Personal Data Broker)** : 시민과 소비자의 개인 데이터 자산을 관리하기 위해 고객의 모든 형태의 데이터를 모니터링하고 데이터를 교환 또는 거래하는 일을 한다.
- ▶ **고속도로 컨트롤러(Highway controller)** : 자율주행차의 부상, 드론 배송 등에 따라 도로와 하늘 교통망 문제에 대해 각종 규제를 기반으로 통제하며, 안전한 통행을 보장하는 일을 한다.
- ▶ **증강현실 여행 제작자(Augmented Reality Journey Builder)** : 증강현실(AR) 기술을 기반으로 다양한 장소와 공간, 시간을 경험할 수 있도록 프로그램을 설정하고 분위기, 음악 등을 기획한다. 예를 들어 고객 취향에 맞춰 SF, 판타지 영화 속 환경을 경험할 수 있도록 한다.



디지털 전환 시대를 맞아, 근로자와 청(소)년들은 어떤 역량을 갖춰야 할까요?

분야에 따라 경중의 차이가 있지만, 정보통신, 제조, 의료, 문화예술 등 거의 모든 분야에서 디지털 문해력(Digital Literacy; 기본적인 사무용SW 활용, 기술이해 등), 산업특화 어플리케이션 활용, 데이터 분석 등의 기술기반역량(Tech Baseline Skills)은 필수역량이 되고 있습니다.

특히, 인공지능, 데이터사이언스, 인간-컴퓨터 인터페이스(HCI) 등 기술혁신역량(Tech Disruptive Skills)을 갖춘 인재에 대한 몸값은 천정부지로 치솟고 있습니다.



인공지능과 빅데이터의 발전으로 협동로봇과 스마트팩토리, 자율주행차, 드론, 개인맞춤형 제조, 디지털 마케팅, 스마트시티와 스마트병원이 가능해졌다고 해도 과언이 아닙니다. 각 분야의 디지털 전환과 혁신을 이끌고 있는 핵심기술이 인공지능과 빅데이터라고 할 수 있습니다.

이상의 기술역량 외에 소프트역량이 필요합니다. 기술과 지식의 발전 속도가 더욱 빨라지고 융·복합화되면서, 그 결과로 나타나는 현상도 매우 복잡해지고 있습니다. 따라서 분석적·비판적 사고, 복합문제 해결력, 창의력, 자기주도적 학습 역량 등의 소프트역량을 갖추고 리더십을 발휘하는 인재가 필요합니다.

마지막으로 말씀드리면, 기술역량과 소프트 역량을 탄탄히 뒷받침할 수 있는 '건전한 직업관과 윤리의식'에 대한 교육도 매우 중요합니다. 결국은 사람의 마음가짐이 근로자의 역량 또는 인재의 크기를 결정하기 때문입니다.



4차 산업혁명, 디지털 전환 시대를 맞아, 정부, 산업계, 노동계의 역할은 무엇이라고 생각하십니까?

산업현장에서 기술혁신에 따라 발생하는 법·제도의 미비를 보완하는 동시에 새로운 산업이 성장할 수 있도록 규제를 완화하거나, 신산업 R&D에 과감한 재정적 투입을 하는 등 시장에만 맡길 수 없는 정부의 역할이 많습니다.

새로운 첨단산업을 육성하고 기존 산업을 고도화하기 위해서 가장 기본이 되는 인재 육성에 대한 장기적 전략을 수립하고 추진해야 하는 것도 정부의 역할입니다. 디지털 기술혁신과 활용이 전 분야에서 성숙기에 접어들게 되면 인공지능과 데이터, SW개발, 클라우드 컴퓨팅, 블록체인 등 분야에서 전문가 수요가 폭발적으로 증가할 것입니다. 현재, 인력양성의 질과 양으로는 증가하는 수요를 충족하지 못할 것입니다. 4차 산업혁명 신기술 중심으로 교육훈련 분야의 구조조정이 필요한 이유입니다.

또한 디지털 기술혁신으로 일자리의 구조조정도 필연적으로 발생할 것입니다. 근로자 각자도 변화를 준비해야겠지만, 변화의 물결이 너무 커서 개인의 책임으로만 넘기기에는 한계가 있습니다. 따라서 정부를 비롯하여 산업계, 노동계가 힘과 지혜를 모아 준비를 해야 합니다.

정부는 '한국판 뉴딜'의 일환으로 인력양성, 취업지원 그리고 디지털 격차해소에 정책적 역량을 집중하고 있습니다. 고용노동부는 기업-대학-민간기업 및 훈련기관을 통해 대졸자 등 청년 구직자를 대상으로 인공지능, 빅데이터, 클라우드 등 디지털 및 신기술 훈련과정을 개발하여 제공하고 있습니다. 청년, 중장년 여성 구직자 등이 디지털 역량 부족으로 노동시장 진입 및 적응에 어려움을 겪지 않도록 '프로그래밍 첫걸음' 등과 같은 디지털기초역량훈련(K-Digital Credit) 사업을 운영하고 있습니다. 또, 중소기업 재직자들이 공동훈련센터를 통해 디지털 융합 훈련과정에 참여하도록 지원하고 있습니다('K-Digital Platform'). 그리고 실업자, 단시간 노동자 등 취약계층을 대상으로 실시하는 국가기간산업 및 전략산업직종 훈련 사업도 증강현실, 빅데이터분석 등 신기술 훈련종목들이 추가된 바 있습니다.

저희 한국고용정보원도 일자리알선 포털인 Work-Net(www.work.go.kr)과 직업훈련 포털인 HRD-Net(www.hrd.go.kr)을 통해 디지털 신기술 관련 일자리와 훈련정보를 제공하고 있습니다.



이것만 알면 나도 영상편집 전문가! 태블릿PC로 영상편집을 해보자

최근 제 주변에 태블릿PC를 이용하는 분들이 많아졌습니다. 핸드폰보다 화면도 크고 PC처럼 활용도 가능해서 연령대가 높은 분들도 쉽게 이용하는 것 같습니다. 사실 저도 최근 태블릿PC를 장만했는데요 생각보다 활용도가 높아서 굉장히 잘 사용하고 있습니다. 무엇보다 영상 편집할 때도 휴대전화보다 화면이 조금 더 큰 태블릿 PC가 훨씬 보기 편하고 사용성도 좋았습니다. 그래서 이번시간에는 태블릿PC의 'iMovie(아이무비)' 앱을 이용해서 보다 쾌적하게(?) 영상편집을 하는 방법을 알아보도록 하겠습니다.

*안드로이드 태블릿의 경우 영상편집 툴이 지난 시간 학습했던 VLL0를 참고하시면 되겠습니다.

정영국 디자이너



아이무비를 이용한 영상편집

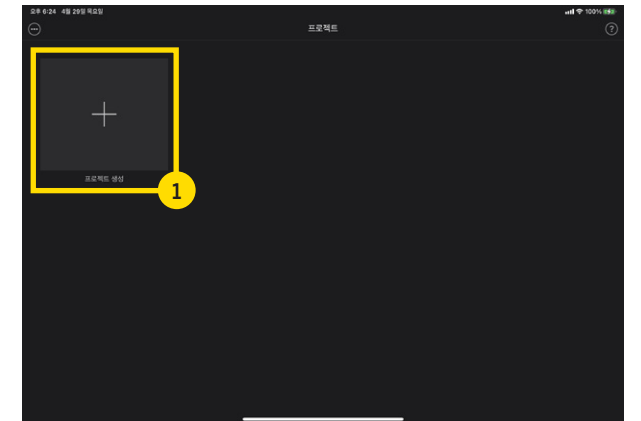
1 예고편 편집

템플릿을 활용하여 손쉽게 멋진 영상을 만들 수 있어요.

먼저 iOS의 태블릿PC(이하 아이패드)의 첫화면에 아이무비 [앱아이콘]을 선택합니다.

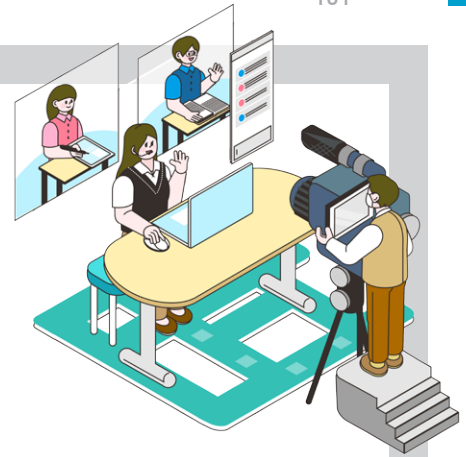
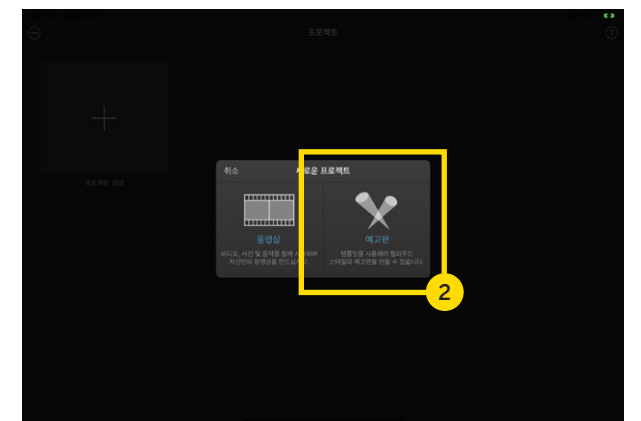
*아이무비는 기본앱이기 때문에 별도의 설치없이 이용할 수 있습니다.

1 프로젝트 생성을 선택하면 '새로운 프로젝트' 창에 [동영상]과 [예고편]을 선택할 수 있습니다.



2 [예고편]을 선택합니다.

*템플릿은 미리 완성해 놓은 예시 문서로 이해하면 됩니다.



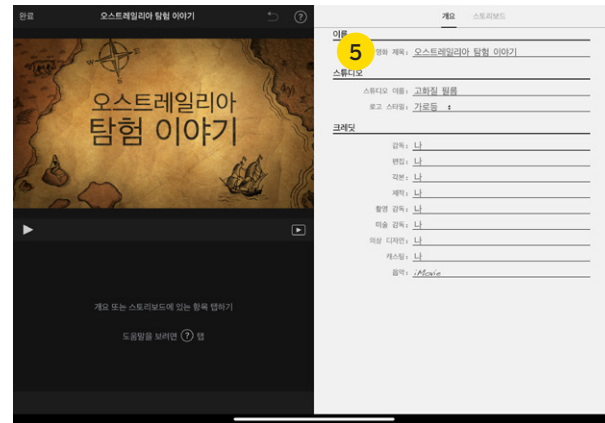
3 화면 하단에 미리 완성된 예시로 템플릿이 나열되어 있습니다. 하나하나 선택하고 재생버튼을 누르면 영상들이 보여집니다.



4 이중에 멋진 영상이 마음에 들었다면 오른쪽 상단에 [생성]을 선택 합니다. 참고로 저는 "모험활극" 템플릿으로 만들어보겠습니다.

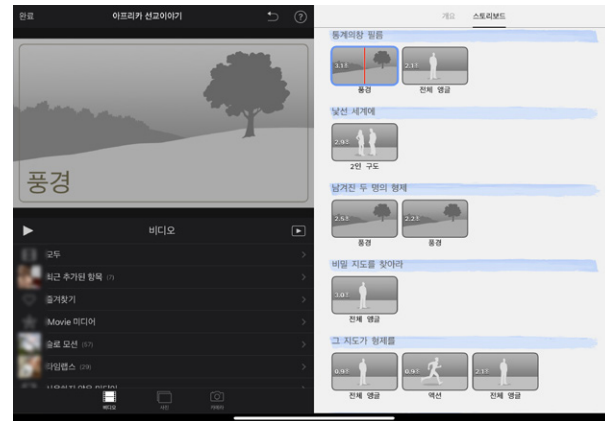
새로운 화면에 놀라거나 어려워하지마세요. 기본적인 영상 정보들을 적어주면 됩니다.

5 저의 경우 영상 제목만 적고 나머지는 그대로 두거나 지우도록 하겠습니다. 이것만으로 놀라게도 인트로와 엔딩크레딧이 완성되었습니다.



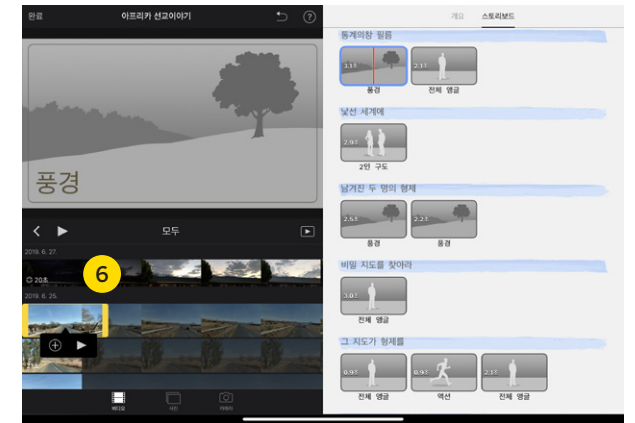
다음은 [스토리보드]를 선택합니다.

그러면 여러 화면들이 나열되어있습니다. 역시 어려워하지 마세요. 오히려 이 작은 화면들의 그림을 보고 그림과 유사한 사진 혹은 영상들을 내 앨범에서 찾아서 대체하면 됩니다. 각 작은 화면 아래에는 제목이 있는데요. 그에 맞는 이미지를 내 앨범에서 찾아보겠습니다.

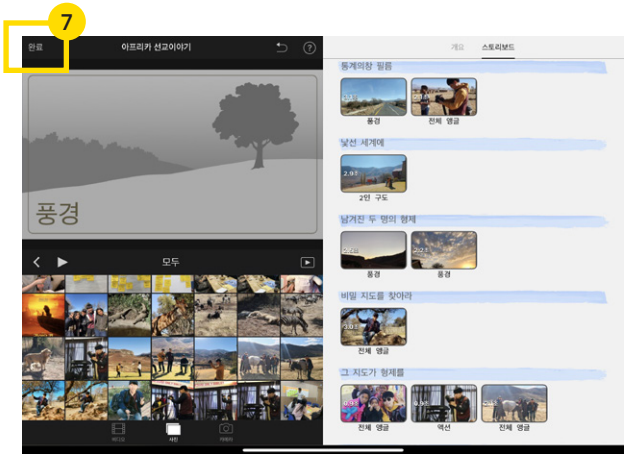


6 좌측 하단의 앨범에서 영상을 선택합니다. [+]버튼을 이용하면 삽입 가능합니다. 동일한 방법으로 모든 작은화면에 영상들을 하나씩 채워주세요.

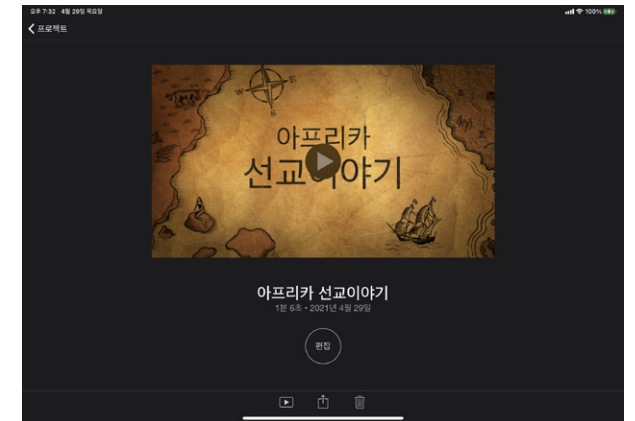
TIP 작은 화면의 위에 스토리 문장이 있습니다. 이 글들도 자유롭게 영상과 어울리게 수정해주시면 좋습니다.



7 영상과 사진을 모두 넣었다면 이제 끝났습니다! 왼쪽 상단의 [완료]버튼을 선택합니다. 굉장히 근사한 영상물이 완성되었습니다. 이제 즐기며 영상 감상을 해봐요!

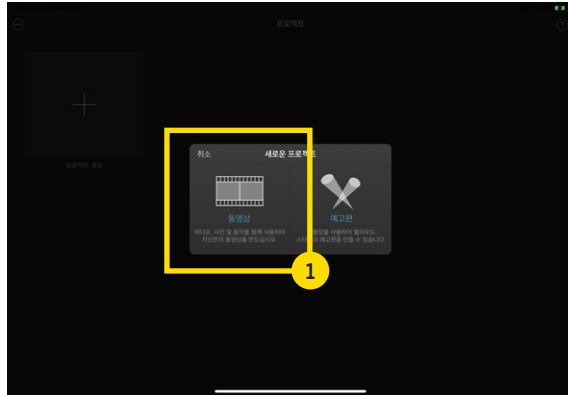


완성한 영상물에서 수정이 필요하다고 생각되면 [편집]을 선택해 다시 수정할 수 있습니다

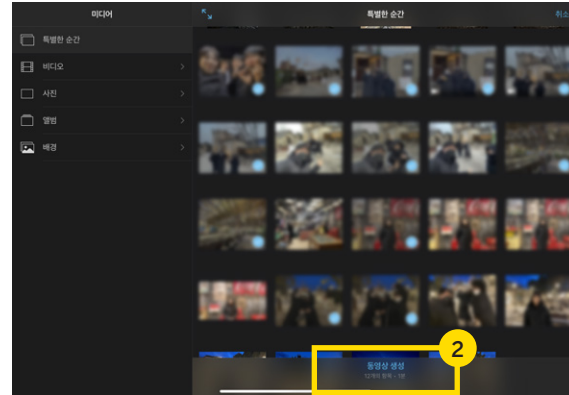


2 동영상 편집

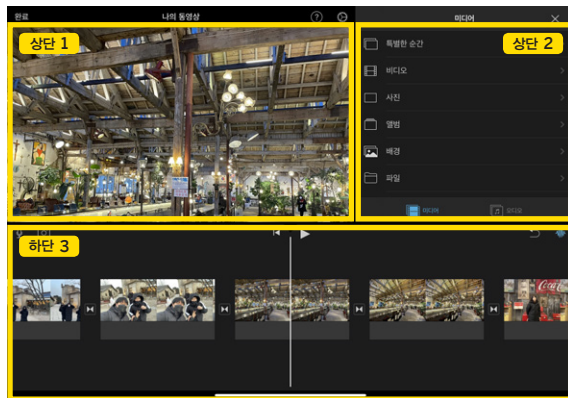
자신의 입맛에 맞게 자유롭게 영상을 만들 수 있어요.



1 프로젝트 생성 뒤 '새로운 프로젝트'창에서 [동영상]을 선택합니다.



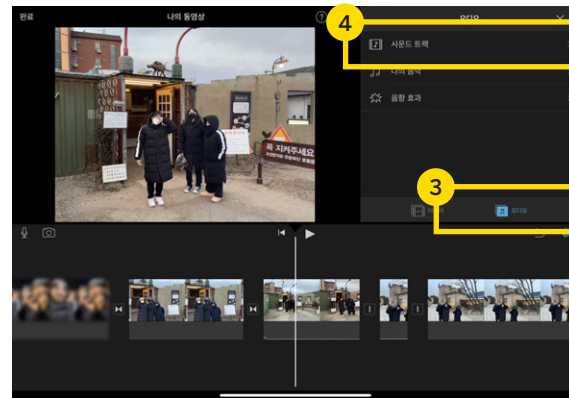
2 특별한 순간의 영상과 사진들이 보이는데요 마음에 드는 영상 또는 사진을 선택하고 [동영상 생성]을 합니다.



위의 화면처럼 나왔다면 잘 하신겁니다. 왼쪽 상단1은 편집중인 화면을 보여주고 오른쪽 상단2은 미디어와 오디오 소스를 불러올 수 있는 창이 있습니다. 하단3은 타임라인을 보여줍니다.

우선 불러온 영상 소스들이 순서대로 나열이 되어있는데요 타임라인을 살펴보면서 순서를 조정합니다. 조정방법은 직관적으로 영상클립을 손으로 길게 누르다음 원하는 위치에서 손을 떼면 됩니다.

자 이렇게 영상 소스를 올렸으니 음악도 올려야겠죠?



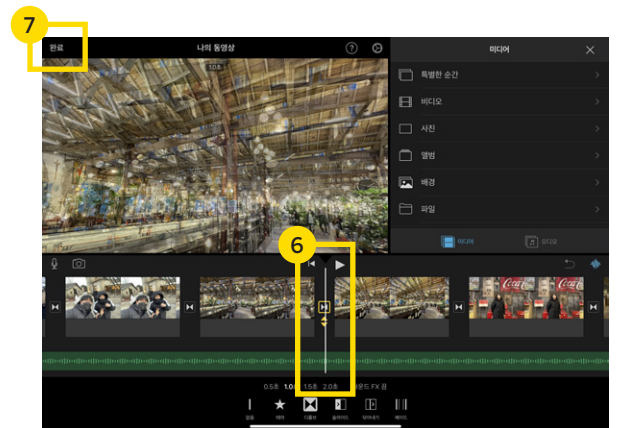
3 오른쪽 상단에 [오디오]탭을 선택합니다.

4 사운드 트랙을 선택하시면 애플에서 제공하는 음원들을 사용할 수 있습니다.



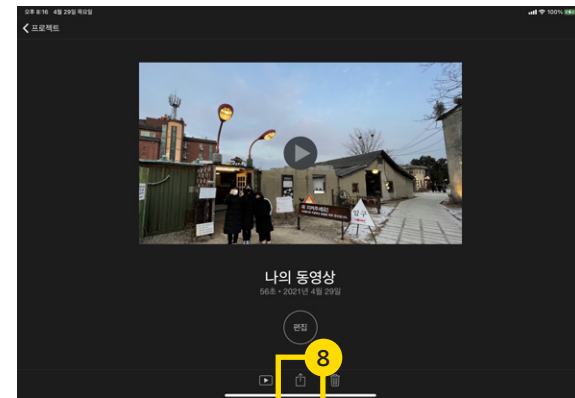
5 마음에 드는 음악을 골랐다면 [+] 버튼을 선택하여 오디오를 추가합니다.

TIP 음향효과도 적절하게 넣으시면 높은수준의 영상을 만들 수 있어요.



6 이제 처음부터 재생을 해보고 심심하거나 강조하고 싶은 부분에 트랜지션을 선택해 삽입해봅니다. 트랜지션의 종류가 많지 않은 것은 조금 아쉽지만 기본 디폴브만으로 충분히 깔끔한 영상을 만들 수 있습니다.

7 재! 이제 영상이 완성 되었다면 왼쪽 상단의 [완료] 버튼을 선택합니다.



8 완성된 영상을 재생해 보세요! 하단의 공유버튼으로 가족과 친구들과 함께 완성된 영상을 공유해보세요.

마무리

어떤가요? 태블릿PC를 이용하여 영상을 만들면 핸드폰보다 넓은화면에서 보다 쾌적하게 편집을 할 수있습니다. 그리고 PC보다 군더더기 없이 깔끔한 화면과 직관적인 조작방법으로 훨씬 쉽고 느껴집니다. 태블릿PC는 휴대가 쉽기때문에 여행중에도 이동중에도 손쉽게 영상편집하고 또 감상할 수 있습니다. 태블릿PC를 잘 활용하지 못한분이 계시다면 이번기회에 영상을 멋지게 만들어보세요!



8가지 방법

김여환 가정의학과전문의 / 의학박사

100살까지 뇌를 지키는



“아름다운 젊음은 우연한 자연현상이지만, 아름다운 노년은 예술작품이다.”

안나 엘리너 루스벨트(Anna Eleanor Roosevelt)

“배 아픈 것은 젊어 떠세요?”

“아이고, 친절도 하지. 이렇게 자주 찾아와서 고마워. 그런데 동사 무소에서 나왔어?” 의사가운을 입고 아침, 저녁으로 회전하는 나에게 천연덕스럽게 이렇게 되물었다. 이것이 치매다. 드라마나 영화에서나 나오는 이런 증상이 보이면 치매는 벌써 집에서 감당하기 힘든 중증단계이다.

65세 이상 노인 10명 중 1명은 치매를 앓고 있고, 80세가 넘으면 4명 중의 1명이다. 그러나 치매는 더 이상 노인의 병만도 아니다. 건강보험정책연구원에 따르면 40대 치매환자가 2001년 563명에서 2008년 862명, 50대가 1,901명에서 4,369명으로 늘었다. 2010년 65세 미만의 중년 치매 환자가 7년 전보다 3배가 늘었다. 정신적인 스트레스, 치매에 대한 가족력, 중금속 노출을 비롯한 유해 환경 노출과 나쁜 생활습관이 중년기 치매를 증가시키는 것으로 보고 있지만, 확실한 이유는 아직 모른다.

당신의 뇌는 건강하십니까?

신경학자 이영진 박사는 “간을 건강하게 하면 간 기능만 좋아지지만, 뇌를 건강하게 하면 뇌뿐만 아니라, 인생이 바뀐다.”라고 한다. 그러나 뇌를 지키는 방법은 그렇게 거창하지가 않다. 운동할 때 헬멧을 쓰는 것에서부터 이를 잘 닦고, 잠을 잘 자고, 삶의 목표를 가지라는 어쩌면 지극히 단순해 보인다. 하지만 이 대수롭지 않은 생활습관 때문에 30년 뒤 자식조차 기억 할 수 없는 참혹한 병에 걸릴 수 있다는 것이다.

건강한 노년은 살다보면 누구에게나 자연스럽게 찾아오는 선물이 아니라, 바쁘게 살아가는 중년이 사소한 것도 허투루 보지 않고 정성스럽게 만들어 낸 결과물이기 때문이다. 골든타임은 심장마비나 뇌출혈에 국한되는 말이 아니다. 당신의 뇌 건강, 인생이 바뀌는 골든타임은 바로 ‘오늘’이다.

치매와 뇌손상_뇌를 보호하고 뇌의 가소성을 키워라 권투선수 알리나 풋볼선수가 나이 50을 넘기면서 치매에 걸리는 것을 볼 때 뇌에 충격을 주는 것은 아주 위험하다. 자동차나 오토바이 사고같이 한 번의 손상이라도 큰 충격을 받는 경우뿐만 아니라, 권투나 격투기, 축구 등 경미하지만 반복적으로 머리에 충격이 가해지는 운동도 외상성 뇌손상으로 인한 치매를 발생 시킬 수 있다. 자동차 안전벨트를 맨다거나, 자전거를 포함한 다양한 스포츠 활동에서 반드시 헬멧을 착용해야한다.

또 과도한 정신적 충격, 만성통증과 같은 지속적인 작은 스트레스도 치매 발병률을 높인다. 지속적인 통증이 시작된 사람들은 통증이 없는 사람들에 비해 이후 10년간 기억력 테스트에서 하락 속도가 9.2%나 빠른 것으로 나타났다. 만성 통증은 스트레스 호르몬을 촉진시켜 인지력 감퇴를 초래할 수 있다. 견딜 수 없는 마음의 상처나 과도한 스트레스도 머리에 가해진 물리적 외상과 비슷하다. 지진이 지나간 집에 균열이 생기듯 뇌세포를 약해지게 만들 수 있기 때문이다.

이와는 반대로 긍정적인 삶의 목표는 뇌의 가소성을 높여 뇌를 건강하게 만든다. 새로운 지식을 잘 받아들이고 문화와 예술에 관심을 보이는 성향의 사람들이 뇌의 퇴화나 치매 위험이 적은 것으로 나타난다. 나이 들면서 고집이 세어진다면 치매를 조심해야 한다. 변화를 인정하고 새로운 목표를 추구할 때 뇌는 건강해진다.

근감소와 치매 당신의 영명이 근육이 뇌 건강을 결정한다 나는 회진을 할 때 환자와 악수를 꼭 한다. 악력이 가장 손쉽게 환자의 근력을 예측하는 방법이기 때문이다. 스웨덴 예네핑 대학은 20년 동안 700명을 대상으로 관찰한 결과 65세 노인의 악력이 급격히 줄면 언어능력, 기억력 등의 인지능력이 떨어져 치매 가능성이 높아진다고 했다. 그러나 일부러 악력으로 악력만 키우면 오히려 방아쇠 수지증후군 같은 병으로 고생한다. 전신 근력이 향상되어 자연스럽게 악력이 증가되어야 한다.

근육은 40세를 기점으로 해마다 1%씩 줄어든다. 80대에 이르면 30대 때의 근육의 절반 정도만 남는다. '근감소증'이 위험한 이유는 각종 질환의 도화선이 되기 때문이다. 근육은 팔다리를 움직이고 운동을 가능하게 하는 기관으로서의 역할뿐 아니라, 뼈와 혈관, 신경, 간, 심장, 췌장 등 신체 전반에 걸쳐 광범위한 영향을 미친다. 그렇다고 강제적인 근육 키우기에 목숨 걸 필요는 없다. 몸짱 열풍에 남용되는 아나볼릭 스테로이드(Androgenic-anabolic Steroids)를 장기간 사용하면 시간 공간적 뇌기능 영역이 크게 손상되어 치매 환자와 유사한 증상이 유발된다. 어깨강패를 만들기 위해 무분별한 아령운동도 관절손상을 유발시킨다. 무리하게 근육을 키우기보다는 젊었을 때 만큼의 근육을 유지할 때 뇌가 건강해진다.

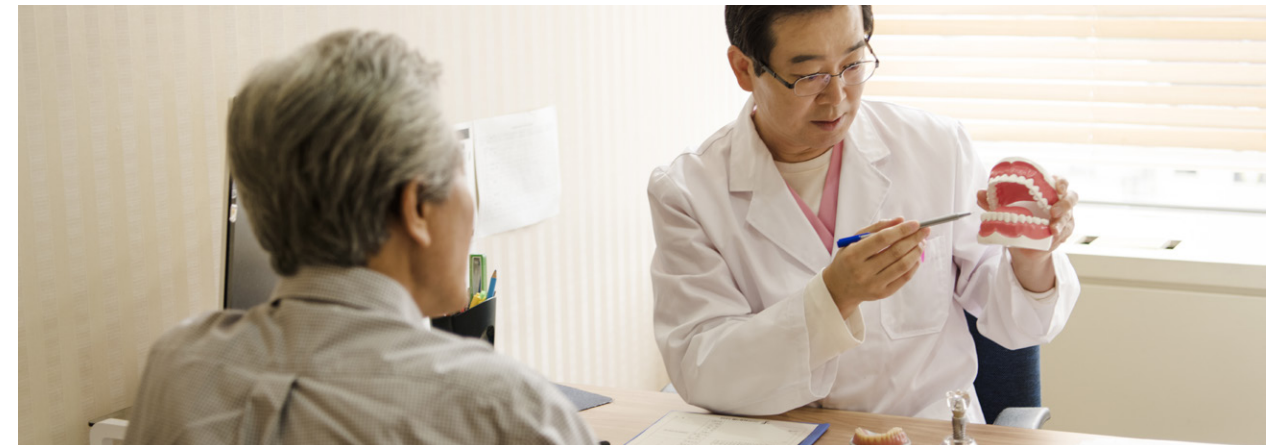


혈관과 치매 뇌세포의 생명줄, 혈관을 관리하라 뇌세포에 영양과 산소를 공급하는 뇌혈관이 막히거나 터지면 뇌 조직 손상으로 이어지면서 '혈관성 치매'가 발생할 수 있다. 알츠하이머 치매는 장기간에 걸쳐 천천히 진행되는 반면 혈관성 치매는 단 한차례의 뇌혈관 손상만으로도 편마비, 구음장애, 연하곤란, 시야장애 등의 심각한 합병증을 유발할 수 있다. 그래서 고혈압, 당뇨병, 고지혈증, 심장병, 흡연 등 혈관을 나쁘게 하는 질병을 치료해야 한다. 최근 한국지질동맥경화학회 발표에 의하면 20대 5명 중 1명이 이상 지질혈증으로 나타나 혈관 건강에 비상등이 켜졌다. 젊을 때부터 건강한 식습관과 운동이 심혈관 질환은 물론 치매까지 관리할 수 있다.

치아와 치매 잇몸이 보내는 치매신호 '설마 이 안 닦는다고 치매에 걸릴까?' 걸린다. 치주질환을 제때 치료하지 않고 방치할 경우 치매에 걸릴 확률이 높다. 음식을 씹는 저작활동 자체가 뇌의 섬유아세포 성장촉진인자(fibroblast growth factor) 분비를 촉진해 식욕을 조절하고 뇌세포 회복과 학습, 기억 형성을 촉진한다. 따라서 치아 수가 감소되어 씹기가 힘들어 뇌로 가는 혈류량이 줄어들면, 뇌 대사 활동의 감소, 전신적인 영양불량을 유발해 인지기능 저하에 영향을 미칠 수 있다. 또한 치주질환을 일으키는 구강세균도 치매를 유발한다. 10명의 알츠하이머병 환자 뇌 조직을 검사했을 때 4명에서 치주질환원균인 P. gingivalis에서 유래한 LPS라는 물질을 확인했고 이것은 치주질환원균이 뇌에 침입하여 지속적으로 감염시켜 점진적 치매, 뇌 위축, 아밀로이드 침착을 유발할 수 있다. 치아 상실은 단순히 구강기능의 저하뿐만 아니라, 전신적인 건강에 영향을 준다. 그래서 구강관리가 치매 예방의 시작이다.

청력과 치매 들리지 않으면 세상에서 멀어 진다 '노인성 난청'은 달팽이관을 비롯해 청신경의 세포 수가 감소하면서 발생한다. 하지만 안 들린다는 것은 단지 소리를 듣지 못하는 것에서 끝나지 않는다. 타인과 대화하는데 어려움을 겪다보니 대인기피증과 우울증이 생길 수 있다. 난청이 심해지면 소리가 들리지 않아 뇌에 언어 자극이 되지 않는다. 이로 인해 신경이 퇴화하게 되어 인지력이 떨어진다. 이것이 바로 난청 치료의 골든타임을 놓쳐서는 안 되는 중요한 이유다.

수면과 치매 잠이 보약이다 불면증은 치매 발병률을 50%까지 높인다. 뇌에 쌓이는 '베타-아밀로이드'가 원인이다. 깨어있는 동안 뇌는 활동하면서 베타-아밀로이드를 발생시키는데, 우리 몸은 잠을 자는 동안 이를 몸 밖으로 배출시킨다. 그러나 잠을 깊게 자



지 못하면 뇌에 베타-아밀로이드가 축적돼 알츠하이머 치매 같은 퇴행성 질환 발병 위험을 높인다. 건강보험심사평가원에 따르면, 2019년 불면증(수면장애)으로 진료받은 환자는 2015년 대비 28.4% 증가했다. 50세 이상 중·장년층이 가장 많다.

잠을 잘 자려면 잘 깨야 한다. 몇 시에 잠들었는지 상관없이 매일 일정한 시간에 기상하는 습관을 들인다. 밤 10시~새벽 2시 사이에는 멜라토닌이 활발하게 분비되므로 이 시간에는 최대한 깊은 잠을 자는 게 좋다. 낮에 햇볕을 쬐고 운동하는 것도 부교감신경을 활성화해 숙면에 도움을 준다.

뇌 건강에 좋은 음식 직접 요리한 오메가3를 먹어라 치매를 예방하기 위해서 WHO는 '지중해 식단'이라 부르는 고기는 적고 채소와 요구르트, 올리브가 많은 음식을 먹으라고 권고한다. 특히 고등어, 참치, 연어 같은 생선류와 호두, 들기름 등 오메가3지방산이 많은 식품을 섭취하는 것이 효과적이다. '오메가3'는 뇌 세포막과 뇌 신경계를 구성하는 주요 성분일 뿐 아니라 베타아밀로이드가 뇌에 쌓이는 것을 막아주기 때문이다.

설거지, 요리, 청소 같은 집안일을 매일 하는 것도 치매예방에 효과적이다. 그래서 오메가3가 든 알약을 먹는 것보다 들기름 막국수를 직접 해먹는 것이 뇌 건강에 좋다.

뇌 건강에 좋은 운동 오늘 한 운동, 25년 후 치매를 막는다 중년에 신체 활동을 많이 할수록 노년기 뇌 건강이 좋아진다. 영국 알츠하이머 연구소는 중년기에 1주일에 150분 이상 운동한 그룹을 장기간 추적한 결과 운동량이 적거나 없는 그룹보다 뇌혈관이 건강하고, 치매 등 뇌 질환을 나타내는 지표도 더 적게 나타났다고 밝혔다. 일주일에 3일 50분씩 빠른 걸음으로 걷기만 해도 체력뿐만 아니라 정신력이 개선된다. 운동을 하고 몸을 움직이는 것은 신경 성장인자 분비를 촉진해 스트레스를 예방하고 뇌에 혈액과 산소, 영양분의 공급을 원활하게 하기 때문이다.



들기름 막국수



준비물: 들깨가루, 들기름, 삶은 계란, 메밀국수100g, 천일염, 국간장, 사골육수, 김가루

- ① 메밀국수를 삶아서 찬물에 행군다.
- ② 물기를 뺀 메밀국수에 천일염, 국간장으로 간을 하고 들깨가루를 넣고 조물조물 묻힌다.
- ③ 국수를 그릇에 담고 고명으로 반 자른 삶은 계란, 김가루, 들기름을 올린다.
- ④ 먹기 전에 국수가 약간 잠길 정도로 사골육수를 붓는다.

통계로 바라보는 세상이야기...

01

롤린 롤린 역주행 중심에 MZ세대!

한국언론진흥재단이 조사한 미디어이슈 자료를 살펴 보면 유튜브 이용률은 전 세대에 걸쳐 높게 나타나지만, 그중에서도 MZ세대에 해당하는 20대~30대는 다른 연령대에 비해 이용행태가 특히 두드러지는 양상을 보였는데, 20대가 78.7%, 30대가 67.8%로 그 수치가 높게 나타났습니다. 2020년 나스미디어가 조사한 「인터넷 이용자 조사」에 따르면, 전체 SNS 이용자들은 인스타그램, 페이스북, 밴드, 카카오톡, 트위터 순으로 사용하고 있었는데요. 이들은 자신들이 원하는 형태로 새로운 콘텐츠를 만들고 공유하며 트렌드를 견인하고 있습니다. MZ세대는 1980년초~2000년대 초 출생한 밀레니얼 세대와 1990년대 중반~2000년대 초반 출생한 Z세대를 통칭하는 말입니다.

03

우리나라 신한류 콘텐츠, K-웹툰!

한국콘텐츠진흥원의 '2020 웹툰 사업체 실태조사'의 연도별 신규 작품 현황을 보면, 2016년 2,315건에서 2019년 2,767건으로 증가하는 추세이며, 최근 3년간은 연간 2,500건 이상의 신규 웹툰 작품이 나오고 있습니다. 대표적인 웹툰 플랫폼인 NAVER에 따르면, 네이버 웹툰의 글로벌 월간 순 사용자가 꾸준한 상승세를 보이는 물론, 2020년 8월 월간 이용자 수가 6,700만 명까지 크게 늘어난 것을 볼 수 있습니다. 2019년 8월부터 2020년 3월까지 7개월간 200만 명 증가한 것에 비해 2020년 3월부터 8월까지 5개월 만에 500만 명이나 이용자 수가 증가했고, 웹툰이 아시아를 넘어 북미와 남미, 유럽 등 전 세계에서 지속적으로 성장하고 있다고 전했습니다.

02

장애인의 일할 권리, 어디까지 왔나?

한국장애인고용공단의 '한눈에 보는 2020 장애인통계'에서 장애인 고용률은 대부분 상승세를 보였지만, '1인 이상 기업체 장애인고용총괄(2019년 기준)'을 살펴보면 전체 사업체 중 장애인고용기업체 수는 3.9%이며, 특히 상시근로자 수는 전체 상시근로자 중 1.5%에 그치는 것으로 나타났습니다. 올해 3월 장애인에 대한 주요 정책이 바뀌었는데 '소득 및 일자리 지원'을 보면, 월 급여가 '20년 179.5만원에서, '21년 182.2만원으로 인상되었고, 전년 대비 복지일자리가 11.2% 증가하여 올해 2만 4,896개로 늘었으며, 앞선 급여 및 일자 확대와 더불어 근로능력이 부족한 장애인을 위해 장애인 연금 기초급여액을 월 25만원에서 30만원으로 인상하였습니다.

04

트렌드로 알아보는 HMR!

재료를 손질한 후 어느 정도 조리가 된 상태로 포장되어, 간단히 끓이거나 데우기만 하면 요리가 완성되는 식품이 HMR(Home Meal Replacement, 가정간편식)입니다. 농림축산식품부의 2019 가공식품소비자태도조사에 따르면, 소비자가 HMR을 구입하는 이유는 "조리가 번거롭고 귀찮아서(21.6%)", "재료를 사서 조리하는 것보다 저렴해서(17.1%)", "시간이 없어서(16.7%)", "맛이 있어서(14.8%)" 순으로 나타났고, HMR을 찾을 때 "유통기한(24.4%)", "브랜드(17.0%)", "가격(11.9%)", "용량(10.6%)", "원산지(9.6%)" 순으로 중요하게 생각하는데, 과거 HMR은 건강에 좋지 않다는 인식이 있었지만, 지금은 신선함과 가격, 건강함까지 행기는 제품으로 발전했습니다.

05

중고거래, 20조 규모 시대

한국개발연구원이 분석한 중고거래 앱 월별 이용자 현황에 따르면 '20년 1월 4,377만 명이었던 앱 이용자 수는 '21년 1월 1억 588만 명까지 늘었고, 이용시간은 '20년 1월 1,194만 시간에서 '21년 1월 2,900만 시간으로 1,706시간 증가했습니다. 트렌드모니터가 조사한 '2020년 중고거래 및 중고거래 플랫폼 관련 인식 조사'를 보면, 전체 응답자의 67.1%가 '이전에 비해 중고물품에 대한 거부감이 덜한 것 같다.'고 답했는데, 사용한 물건을 재사용 하는 것에 대한 우려와 불신이 낮아진 것을 알 수 있었습니다. 와이즈앱 와이즈리테일의 중고거래 앱별 사용자 규모는 2021년 1월 기준, 당근 1,325만 명, 번개 284만 명, 중고 74만 명 순으로 나타났습니다.

07

우리시대 문제해결, 데이터로부터!

지난해 통계청은 조용민 구글메니저와 함께 빅데이터 시대에 더욱 중요해진 통계 데이터 해석에 대해 이야기를 나누었습니다. 우리가 당연한 사회적 문제나 기업의 가치 창출 등 다양한 부문에서 국가 통계, 통계적 데이터를 활용한 '데이터 마이닝'으로 해결이 가능하다는데, 이는 데이터 베이스 내에서 어떠한 방법에 의해 관심 있는 지식을 찾아내는 과정이며, 기대했던 정보뿐만 아니라 기대하지 못했던 정보까지 찾을 수 있는 기술을 말합니다. 구글은 사람의 입술 모양을 분석해 어떤 말을 하는지 분석할 수 있는 '머신러닝'도 소개했는데, 머신러닝은 인공지능의 한 분야로 컴퓨터에 명시적인 프로그램 없이 배울 수 있는 능력을 부여하는 연구 분야입니다.

06

통계조사도 언택트 시대가 열렸다!

우리가 일반적으로 얘기하는 비대면은 언택트, 즉 얼굴을 마주보지 않는 것을 의미합니다. 통계청에서는 현재 비대면조사 방법으로 컴퓨터를 이용한 자기기입식 조사(CASI), 스마트폰을 이용한 자기기입식 조사(SASI), 컴퓨터를 이용한 전화조사(CATI), E-mail, Fax를 활용하고 있습니다. 비대면조사 확대 필요성은 세 가지 관점에서 접근할 수 있어요. 첫째, 통계조사 비용을 절감할 수 있고, 둘째, 코로나19 등 신종 감염병 또는 가축전염병이 발생하는 상황이니 더욱 조심해야하며, 셋째, 2019년 기준 우리나라 전체 가구의 30.2%가 1인 가구, 46%가 맞벌이 가구일 뿐 아니라 아파트나 기업 등 조사대상처의 보안 강화로 응답자를 접촉할 기회 자체가 축소되고 있기 때문입니다.

08

과학의 날, 일상에서 만나는 인공지능!

2016년 이세돌 9단이 알파고와 바둑 대결을 펼치는 장면이 전 세계로 생중계되면서, 인공지능이 사람들에게 알려지기 시작했습니다. 지난 4월 21일 '과학의 날'을 맞아, 인공지능 이야기를 해보려고 합니다. 인공지능은 '인간의 지능이 가지는 학습, 추리, 적응, 논증과 같은 기능을 갖춘 컴퓨터 시스템'으로, 의료 기술과 같은 전문적인 분야부터, 사물인터넷(IoT)처럼 일상적인 부분까지 널리 쓰이고 있습니다. 방송통신위원회의 "2020 방송매체이용행태조사"에 따르면, 사람들이 사용하는 음성인식 인공지능 기기에서 음악 재생(35.1%)이 가장 많이 쓰였고, 정보 검색(30.5%)과 TV 프로그램 재생(22.7%)이 그 다음으로 많았습니다. 앞으로 어떤 새로운 과학기술이 우리를 더욱 풍요롭게 만들까요?

09

홈코노미, 이젠 집에서 다 할 수 있다?!

홈코노미는 홈(home)과 이코노미(economy)의 합성어로,...

10

4월 5일 식목일, 3월로 앞당긴다고?!

매년 4월 5일은 식목일입니다. 현재까지는 4월에 식목일이 지정되어 있지만, 3월로 앞당기는 방안을 검토하고...

13

국가통계포털(KOSIS) '챗봇 서비스' 개시!

통계청(청장 류근관)은 지난 4월 22일부터 국민들이 보다 손쉽게 통계 정보에 접근할 수 있도록 대화형 인공지능 플랫폼 'KOSIS chatbot(이하 코봇)'을 국가통계포털(KOSIS, https://kosis.kr)을 통해 서비스하고 있습니다.

14

완성도 높은 문서를 만드는 트리맵!

트리맵은 큰 사각형 안에 작은 사각형이 들어있는 형태로, 사각형을 순차적으로 표시하여 계층 구조와 데이터의 비율을 표시하는 차트를 말합니다.

11

세계 물의 날, 슬기로운 물 절약생활!

인구와 경제 활동의 증가로 수질이 오염되고 먹는 물이 부족해지자 1993년부터 UN이 경각심을 일깨우기 위해 매년 3월 22일을 세계 물의 날로 지정했습니다.

12

결혼 5년차, 유자녀 유주택 맞벌이 높아!

2019년 기준 혼인 5년차 초혼부부 212,287쌍을 대상으로 조사한 결과, 혼인신고 후 5년 동안 국내에 거주하면서 혼인 관계를 유지 중인 초혼부부는 21만 2천 쌍인데, 이들 초혼부부 중 아내는 1980년대생이 80.3%(초반 36.3%, 후반 44.0%)로 가장 많았고, 5년간 맞벌이를 유지한 부부는 25.6%로 5년간 외벌이 유지 부부 18.5%보다 높았으며, 5년차 신혼부부의 40.7%는 5년간 주택을 소유하지 않았고, 1차년에 주택을 소유하여 5년 동안 주택 소유를 유지한 부부는 28.7%로 나타났습니다.

15

플라스틱, 쓰는데 5년 썩는데 500년

지난 1월 15일 세계적인 환경단체 그린피스에 2020년 10월 18일부터 10월 24일까지 일반가정 260가구 구성원을 대상으로 플라스틱 폐기물을 조사한 '우리 집 플라스틱 어디서 왔니' 보고서를 공개했습니다.



2021년도 통계교육원 교육훈련계획

집합과정 - 9개 과정 147회 4,290명

구분	과정명	교육대상	교육일수	기당인원	교육횟수	교육일정	
A 기본교육	1	9급 신규자 기본교육(공채)	9급 공채 신규임용예정자	15	60	1	2.22-3.12
	2	9급 신규자 기본교육(경채)	9급 경채 신규임용예정자	10	-	-	-
	3	7급 신규자 실무교육(신설)	7급 신규자	4	20	1	6.22-6.25
	4	4급 승진후보자 역량향상	4급 승진후보자	5	-	-	-
	5	5급 승진후보자 역량향상	5급 승진후보자	5	25	2	4.12-4.16/8.16-8.20
	6	6급 승진후보자 역량향상	6급 승진후보자	4	30	2	5.11-5.14/10.26-10.29
	7	7급 승진후보자 역량향상	7급 승진후보자	4	30	2	2.17-2.19/7.12-7.14
	8	지방청 조사관리자 역량강화1	지방청 통발자	3	30	1	3.24-3.26/6.7-6.9
	9	지방청 조사관리자 역량강화2	지방청 통발자	3	30	1	5.12-5.14
B 전문교육	B-1 국가통계정책						
	B-1-1 통계정책						
	10	국가통계의 이해	제한없음	3	40	2	4.21-4.23/9.8-9.10
	11	통계와 정책	통계청	3	20	2	3.17-3.19/6.30-7.2
	12	정책과정과 통계의 역할	제한없음	3	30	1	9.29-10.1
	13	국가승인통계관리	통계작성기관	3	30	2	3.22-3.24/9.15-9.17
	14	통계학 원리	제한없음	3	30	2	2.17-2.19/7.12-7.14
	15	정책자료 작성 방법론	제한없음	2	20	1	9.9-9.10
	16	통계기초 및 활용	제한없음	5	30	2	5.31-6.4/11.1-11.5
	B-1-2 통계기초						
17	한국표준사업분류	제한없음	3	30	1	10.18-10.20	
18	한국표준직업분류	제한없음	3	30	1	6.9-6.11	
19	한국표준종별·사업분류의 이해	제한없음	1	30	1	6.18	
20	한국표준종별·사업분류의 활용	제한없음	1	30	1	11.12	
B-2 국가통계이해							
B-2-1 경제통계							
21	경제통계의 이해	제한없음	3	20	1	11.1-11.3	
22	국민계정	제한없음	3	20	1	9.4-9.8	
23	재무제표	제한없음	3	30	2	6.21-6.23/8.23-8.25	
B-2-2 사회통계							
24	노동통계의 이해	제한없음	3	30	1	9.6-9.8	
25	인구통계의 이해	제한없음	3	20	1	4.14-4.16	
26	북한통계의 이해	제한없음	3	30	1	10.11-10.13	
27	소득분배지표의 이해(신설)	제한없음	2	20	1	8.5-8.6	
28	고용통계의 이해(신설)	제한없음	2	20	1	6.28-6.29	
B-2-3 기타							
29	빅데이터와 행정자료의 이해	통계청	3	30	2	3.22-3.24/9.13-9.15	
30	지속가능발전목표(SDGs)의 이해(신설)	제한없음	1	20	1	6.2	
B-3 국가통계작성							
B-3-1 조사기획							
31	국가통계실무(조사설계 및 조사표설계 등)	통계청	4	40	2	3.8-3.11/8.24-8.27	
32	국가통계실무(자료분류 및 수집)	통계청	4	40	2	4.5-4.8/9.27-9.30	
33	조사설계 및 조사표설계	제한없음	3	20	1	4.19-4.21	
34	표본실무	제한없음	3	20	1	7.14-7.16	
B-3-2 자료수집 처리 및 분석							
35	통계조사관 직무연수	통계청	3	30	1	8.25-8.27	
36	지역통계실무	제한없음	3	20	1	3.3-3.5	
37	국가통계실무(자료수집 처리 및 분석)	통계청	4	40	2	5.11-5.14/10.25-10.28	
38	자료수집 처리 및 분석	제한없음	3	30	1	11.15-11.17	
39	계정조사실무	통계청	2	30	1	1.11-1.12	
40	데이터 에디팅	제한없음	4	20	1	7.5-7.8	
41	시계열분석	제한없음	4	20	1	4.12-4.15	
42	지수이론	제한없음	3	20	1	7.7-7.9	
B-3-3 공표 및 관리							
43	국가통계실무(통계작성·공표 등)	통계청	4	40	2	6.15-6.18/11.23-11.26	
44	국가통계정보의 활용	제한없음	3	20	1	4.26-4.28	
45	통계데이터 비밀보호의 이해	제한없음	3	20	1	7.14-7.16	
46	통계보고서 작성	제한없음	3	30	3	3.3-3.5/6.28-6.30/10.4-10.6	
B-4 데이터 분석							
B-4-1 R							
47	R 초급 통계분석	제한없음	4	30	2	2.23-2.26/8.17-8.20	
48	R 데이터 시각화	제한없음	3	30	2	4.7-4.9/8.9-9.10	
49	R 중급 통계분석	제한없음	3	30	2	6.28-6.30/10.4-10.6	
50	R 고급 통계분석	제한없음	3	30	2	7.14-7.16/11.17-11.19	
B-4-2 SAS							
51	SAS 통계실무(신설)	통계청	3	20	1	3.29-3.31	
52	SAS 통계분석 I	제한없음	5	20	2	4.12-4.16/10.11-10.15	
53	SAS 통계분석 II	제한없음	5	20	2	7.12-7.16/11.8-11.12	
B-4-3 SPSS							
54	SPSS 통계분석 I	제한없음	5	30	2	5.10-5.14/8.23-8.27	
55	SPSS 통계분석 II	제한없음	5	30	2	6.14-6.18/11.22-11.26	
B-4-4 엑셀							
56	엑셀 데이터 분석 및 활용	제한없음	3	30	3	3.3-3.5/6.21-6.23/11.3-11.5	
57	엑셀 통계작성	제한없음	3	30	2	4.26-4.28/10.13-10.15	
B-4-5 빅데이터							
58	행정자료 통계작성	제한없음	3	20	1	7.7-7.9	
59	빅데이터 프로그래밍 수행	제한없음	4	20	1	7.19-7.22	
60	하둡 기반 빅데이터 통계분석	제한없음	3	20	2	6.2-6.4/11.1-11.3	
61	파이썬 통계분석 I	제한없음	3	20	2	3.17-3.19/8.16-8.18	
62	파이썬 통계분석 II(신설)	제한없음	3	20	2	4.28-4.30/10.18-10.20	
B-5 전문가 양성							
63	경제시계열분석 및 지수이론	통계청	14	10	-	-	
64	국민계정	통계청	12	10	2	-	
65	인구통계분석	통계청	12	10	2	-	
66	빅데이터 분석(신설)	통계청	14	10	1	-	
67	통계기획(신설)	통계청	13	10	-	-	
68	표본설계 및 추정	통계청	13	10	-	-	
69	무응답 자료처리 및 분석	통계청	12	10	-	-	
70	통계데이터 비밀보호(신설)	통계청	12	10	-	-	
B-6 기타 통계교육							
B-6-1 맞춤형 통계교육							
71	맞춤형 통계교육(기관)	통계작성기관	2	25	16	-	
B-6-2 통계세미나							
72	통계세미나	제한없음	1	40	4	-	
B-6-3 학생교육							
73	사랑나눔 통계교실	초등학교 5-6학년	3	25	2	7.26-7.28/8.9-8.11	
74	어린이 통계캠프	초등학교 5-6학년	3	30	2	6.28-6.30/7.5-7.7	
75	중학생 통계이카데미	중학생	3	30	2	7.14-7.16/7.21-7.23	
76	고등학생 통계이카데미	고등학생	2	30	1	8.5-8.6	
B-6-4 교사교육							
77	초등학교 교사 통계연수	초등학교 교사	2	30	2	1.18-1.19/8.9-8.10	
78	중학교 교사 통계연수	중학교 교사	2	30	2	1.27-1.28/8.9-8.10	
79	고등학교 교사 통계연수	고등학교 교사	2	30	2	1.21-1.22/8.2-8.3	
80	중등 교사 통계연수(신설)	중·고등학교 교사	3	30	2	1.18-1.20/7.26-7.28	
81	실용통계 지도교사 통계연수	고등학교 수학 교사	4	30	2	1.25-1.26/7.26-7.29	
82	시도교육청 교사 통계연수	초·중·고 수학교육원 교사	2	30	3	-	
83	통계를 활용한 통합사회 지도교사 연수(신설)	고등학교 사회과 교사	2	30	1	8.2-8.3	
B-6-5 외국어교육							
84	KOICA 위탁 통계연수	외국공무원	17	20	2	-	
85	UNSIAP 공조 통계연수	외국공무원	5	20	2	-	
B-7 정보화							
86	오피스를 활용한 데이터 시각화	제한없음	3	40	4	3.22-3.24/6.7-6.9/9.13-9.15/11.8-11.10	
87	디지털 영상 및 이미지 활용	제한없음	3	30	1	10.20-10.22	
C 기타교육							
C-1 사책교육							
88	소셜미디어 활용	제한없음	2	20	1	6.10-6.11	
C-2 일반소양							
89	생활 속의 음악	제한없음	3	60	2	8.11-8.13/11.3-11.5	
90	현장조사 스트레칭 관리	통계청	3	30	3	5.26-5.28/6.30-7.2/10.27-10.29	
91	공무직 은퇴설계	통계청	3	30	1	9.29-10.1	

이러닝(e-Learning) 과정 - 112개(2.1~12.10.) 중 실시 운영

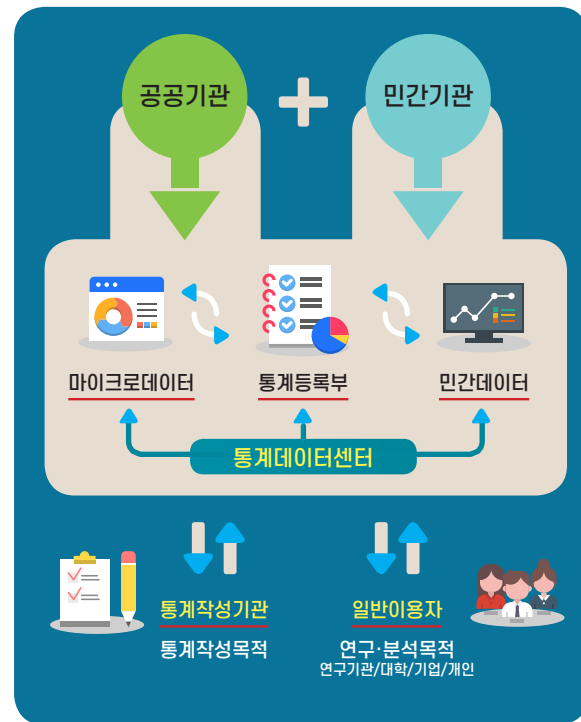
구분	과정명	수준	교육대상	인정시간	
A 기본교육	1	5급 승진후보자 역량평가의 이해	초급	통계청	3
					301
					68
					57
					1
					1
					4
					11
					3
					2
B 전문교육	B-1 국가통계정책				13
	B-1-1 통계정책				3
	2	국가통계 및 제도(신설)	초급	제한없음	2
	3	상위통계의 이해(신설)	초급	제한없음	2
	4	5급 통계작성기관 관리	초급	제한없음	2
	5	지역정보 및 통계의 활용	초급	제한없음	4
	6	지역정책과 통계 활용	초급	제한없음	11
	7	통계 및 통계	초급	제한없음	3
	8	통계기초 및 활용(신설)	초급	제한없음	2
	9	통계기초 및 활용	초급	제한없음	2
10	통계기초	초급	제한없음	2	
11	통계기초 필수 지식	초급	제한없음	3	
12	통계작성기관을 위한 통계DB시스템 사용법	초급	제한없음	2	
13	통계작성기관을 위한 통계DB시스템 사용법	초급	제한없음	3	
14	통계작성기관(신설)	초급	제한없음	1	
15	통계학의 이해	초급	제한없음	7	
B-1-2 통계기초				11	
16	한국표준사업분류	초급	제한없음	6	
17	한국표준직업분류	초급	제한없음	5	
B-2 국가통계이해				67	
B-2-1 경제통계				19	
18	국민계정통계작성	초급	통계청	4	
19	서비스업통계작성	초급	통계청	3	
20	소비자물가지수	초급	통계청	5	
21	주요인수통계작성	초급	통계청	2	
22	재무제표 이해	초급	제한없음	5	
B-2-2 사회통계				48	
23	노동통계작성	초급	통계청	3	
24	가계통계작성	초급	통계청	2	
25	경제활동인구조사	초급	통계청	3	
26	경제활동인구조사 사례집	초급	통계청	2	
27	노동통계작성	초급	통계청	3	
28	통계작성기관 관리(신설)	초급	제한없음	4	
29	노동통계작성(신설)	초급	통계청	6	
30	노동통계작성(신설)	초급	통계청	2	
31	사회통계의 이해	초급	제한없음	6	
32	산자물가지수	초급	통계청	2	
33	양식통계작성(신설)	초급	통계청	2	
34	인구통계작성	초급	통계청	2	
35	이동통계작성	초급	통계청	2	
36	이동통계작성	초급	통계청	3	
37	이민통계작성 및 고령조사	초급	제한없음	2	
38	인구통계작성	초급	통계청	3	
39	지역통계작성	초급	통계청	3	
40	축산물통계작성(신설)	초급	통계청	3	
B-3 국가통계작성				58	
B-3-1 조사기획				10	
41	조사방법 기초	초급	제한없음	3	
42	조사방법의 이해	초급	제한없음	7	
B-3-2 자료수집 처리 및 분석				26	
43	시계열자료의 분석과 실무	초급	제한없음	5	
44	자료수집(신설)	초급	제한없음	2	
45	표본조사 방법론	초급	제한없음	6	
46	표본조사 방법론(신설)	초급	제한없음	4	
47	현장조사 이해(신설)	초급	제한없음	4	
48	통계작성의 이해와 사례	초급	제한없음	7	
B-3-3 공표 및 관리				22	
49	통계작성 방법론(신설)	초급	통계청 공무원	2	
50	MDIS 활용	초급	제한없음	2	
51	NARA-PC 활용	초급	제한없음	6	
52	SGIS 연동	초급	제한없음	4	
53	SGIS 온라인 활용	초급	제한없음	2	
54	국가통계포털(KOSIS) 활용	초급	제한없음	2	
55	통계를 활용한 보고서 작성방법	초급	제한없음	4	
B-4 데이터 분석				69	
B-4-1 R				24	
56	R 기초	초급	제한없음	12	
57	R 활용	초급	제한없음	12	
B-4-2 SAS				6	
58	예제론 SAS	초급	제한없음	6	
B-4-3 SPSS				9	
59	SPSS 초급 통계분석	초급	제한없음	5	
60	SPSS 중급 통계분석	초급	제한없음	4	
B-4-4 엑셀				4	
61	엑셀을 배우는 통계분석(신설)	초급	제한없음	4	
B-4-5 빅데이터				23	
62	빅데이터를 활용한 통계작성	초급	제한없음	6	
63	공공데이터를 활용한 통계작성	초급	통계청	6	
64	데이터 분석을 위한 SQL 입문(신설)	초급	제한없음	3	
65	빅데이터와 통계	초급	제한없음	3	
66	파이썬 데이터분석 초급(신설)	초급	제한없음	3	
B-4-6 기타				3	
67	통계작성 학습을 위한 필수 통계지식	초급	제한없음	1	
68	통계 교육용 공학도구 토크라이 활용	초급	제한없음	2	
B-5 기타 통계교육				35	
B-5-1 학생교육				27	

행정통계자료와 민간자료를 한곳에!

통계데이터 서비스

통계데이터센터가 새로운 서비스로
정보화 사회를 선도합니다.

행정자료를 수집하여 가공한 행정통계자료(통계등록부),
통계청이 제공하는 승인된 통계기초자료(마이크로데이터) 등
통계자료뿐만 아니라 민간자료까지 한 곳에서 분석이 가능한 통계데이터센터(SDC)



제공서비스

분석센터 이용서비스 <ul style="list-style-type: none"> 데이터분석 플랫폼 제공 (분석시스템, 통계패키지) 통계자료 및 민간자료, 이용자 반입 자료 등 연계·분석 	주문형 분석서비스 <ul style="list-style-type: none"> 시간 및 거리상 센터 방문이 어려운 이용자를 위한 서비스 이용자가 원하는 형태로 결과 제공(비식별화처리된 형태)
교육지원 <ul style="list-style-type: none"> 대학생을 대상으로 통계데이터 센터의 실제 데이터를 통한 다양한 분석 체험 기회 제공 통계학과 등 관련학과와 정규 전공 수업 실습 분석 지원 및 이용 지원 	전문가분석 지원 <ul style="list-style-type: none"> 분석 경험이 없는 이용자를 위한 분석 지원 및 상담
명부서비스 이용 <ul style="list-style-type: none"> 통계작성기관이 승인통계 작성 목적으로 요청할 때 조사표본 명부 제공 	
빅데이터를 활용한 서비스 (홈페이지 내 제공)	

제공자료

행정통계자료

구분	부문	자료명	시계열
행정통계자료	경제·사회	기업통계등록부DB(대표자, 사업자 기준)(연간)	2010~2019년
		기업통계등록부DB(대표자, 사업자 기준)(분기)	2018. 4분기~2020. 4분기
		사업장 기초DB(4대보험)	2009~2019년
		사업장 기초DB(법인)	2008~2019년
		중사자-기업체 연계DB	2015~2019년
		육아휴직 사용현황DB	2015~2019년
인구·가구		인구DB	2015~2019년
		가구DB	2015~2019년
		주택DB	2015~2019년
농림·어업		농업DB	2015~2019년
		임업DB	2015~2019년
		어업DB	2015~2019년

민간자료

구분	부문	자료명	시계열
민간자료	인구·가구	성·연령별 유동인구(전국)	2015.1월~2020.12월
		시간대별 유동인구(전국)	2015.1월~2020.12월
		요일별 유동인구(전국)	2015.1월~2020.12월
		[코로나19]방문인구정보(전국)	2019.1월~2020.12월
		[코로나19]방문인구정보(유입지null)(전국)	2019.1월~2020.12월
		성연령별 카드매출2(전국)(국내)	2018.10월~2020.12월
	경제·사회	성연령별 카드매출2(전국)(해외)	2018.10월~2020.12월
		신용통계정보(대출)(전국)	2018.10월~2020.12월
		신용통계정보(소득)(전국)	2018.10월~2020.12월
		신용통계정보(카드소비)(전국)	2018.10월~2020.12월
		인구이동정보(전입인구)(전국)	2019년 4분기~2020년 4분기
		인구이동정보(전출인구)(전국)	2019년 4분기~2020년 4분기
	기초정보	인구이동정보(통근인구)(전국)	2019년 4분기~2020년 4분기
		구매상품정보(전국)(오프라인)	2019.12월~2020.11월
		모바일상품권 거래데이터(전국)	2018년
		온라인가격정보	2017.1월~2021.3월
		소지역정보1,2	1 : ~2018년 2 : 2019년~(개편)

통계기초자료

구분	부문	자료명
통계기초자료	사업체	광업제조업조사, 경제총조사, 전국사업체조사 등
	인구	인구주택총조사, 사망원인통계, 인구동향조사 등
	가구	가계금융복지조사, 경제활동인구조사 등
	농림·어업	농림어업총조사, 농가경제조사 등

통계청, 정부부처, 지방자치단체, 연구기관 등
모든 기관의 마이크로데이터를 한 곳으로...

보다 심도 있고 다양한 분석을 원한다면
지금 바로 **MDIS**를 클릭해보세요

■ 서비스 소개 (2020년 11월 기준)

가. 서비스명 : 마이크로데이터통합서비스(MDIS, mdis.kostat.go.kr)

나. 제공통계수 : 22개 주·시·도별 총 290종 통계 제공(통계청 48종 및 통계작성기관 242종)

다. 제공형태 : 마이크로데이터(통계에 따라 사람, 사업체, 가구 기반 자료)

기준	주요 통계
인구 · 가구	경제활동인구조사, 가계동향조사, 국내인구이동통계, 사망원인통계, 가계금융복지조사, 지역별고용조사, 인구주택총조사, 인구동향조사, 생활시간조사, 사회조사 외 5종
통계청 사업체 · 농어가	전국사업체조사, 광업·제조업조사, 농가경제조사, 기업활동조사, 농업총조사, 농산물생산비조사, 경제총조사, 어가경제조사, 운수업조사 외 16종
행정통계 및 기타	귀농귀촌인통계, 영리법인기업체행정통계, 신혼부부통계, 주택소유통계, 중장년층행정통계, 퇴직연금통계, 일자리행정통계, 기업생멸행정통계
통계작성기관	전국다문화가족실태조사, 가족실태조사, 자동차주행거리통계, 직종별사업체노동력조사, 보육실태조사, 기상관측통계, 국민여가활동조사, 외래관광객실태조사, 한부모가족실태조사, 청소년종합실태조사 외 207종

■ 서비스 내용

가. 구분 : 자료의 민감성 정도에 따라
공공용, 인가용, 특수목적용으로 구분 운영

나. 수수료

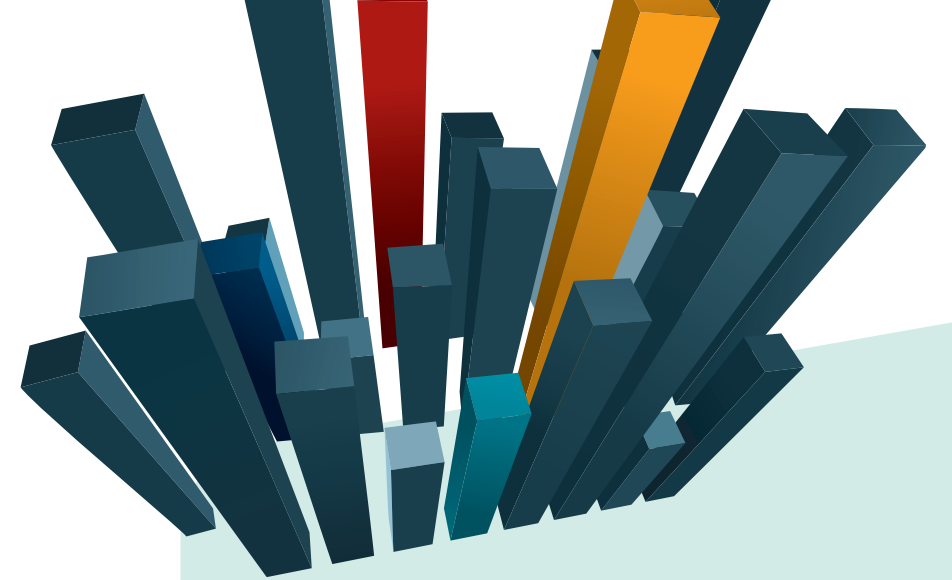
- 무료 : 공공용 자료
- 인가용/특수목적용 : 선택제 수수료 부과

다. 서비스 방법

- 추출·다운로드 : MDIS 포털에서 직접 무료 다운로드
- 원격접근서비스 : 승인 후 이용자가 집·사무실 등에서 통계청 서버 접속 후 활용
- 이용센터 : 승인 후 지정된 장소를 방문 · 활용

■ 문의

- 연락처 : 재단법인 한국통계진흥원
- 전화 : (02) 512-0167 FAX : (02) 515-0240
- 주소 : (우) 06097 서울특별시 강남구 선릉로 612, 6층
- E-mail : MDIS@stat.or.kr



공공자료의 개방 및 공유 확대

MDIS
[mdis.kostat.go.kr]

통계청 46종 및 통계작성기관 230종의
통계자료로 제공확대



통계청에서 국가통계를 활용하세요!

통계청은 통계개발·활용에 필요한 모든 정보와 도움을 제공합니다.
 다양한 국가통계정보 제공사이트를 활용하세요.

원하는 자료를 직접 분석 및 요청

MDIS
 [mdis.kostat.go.kr]

온라인으로 추출/다운로드 선택 시
 공공용 마이크로데이터를
 무료로 분석 활용 가능



국가통계 쉽게 찾기

KOSIS
 [kosis.kr]

국내, 국제, 북한의 주요 통계를
 한 곳에 모아 알기 쉽게 분류해 제공



행정통계자료와 민간자료를 한곳에

통계데이터센터
 [data.kostat.go.kr]

행정통계자료(통계등록부), 민간자료의
 연계·융합이 가능한 데이터 플랫폼



지도 위 통계정보 살펴보기

SGIS
 [sgis.kostat.go.kr]

인구, 가구, 주택, 사업체 통계 등 각종
 통계를 지도(GIS) 위에서 한눈에 파악



국내 유일의 국가통계교육 전문기관

통계교육원
 [sti.kostat.go.kr]

통계작성 및 활용 전문통계과정,
 기관맞춤형과정, e-러닝 과정



통계청
 통계교육원