

2020 멀티캠퍼스 청년취업 아카데미

빅데이터 활용을 위한 DB관리 전문가 과정

Apr. 2020

multicampus

Copyright by Multicampus Co., Ltd. All right reserved

CONTENTS

2020 멀티캠퍼스 청년취업 아카데미

01 빅데이터 개요

02 빅데이터 활용 사례

03 빅데이터 시장

04 취업시장에서의 전망

05 과정 소개

06 과정 체계

07 커리큘럼

08 활용 장비

01_ 빅데이터 개요

빅데이터는 기존 데이터보다 너무 방대하여 기존의 방법이나 도구로 수집/저장/분석 등이 어려운 정형 및 비정형 데이터들을 의미합니다.

오늘날에는 데이터 수집/저장/관리/분석하는 역량 및 대용량 데이터를 처리하는 아키텍처라고 정의합니다.

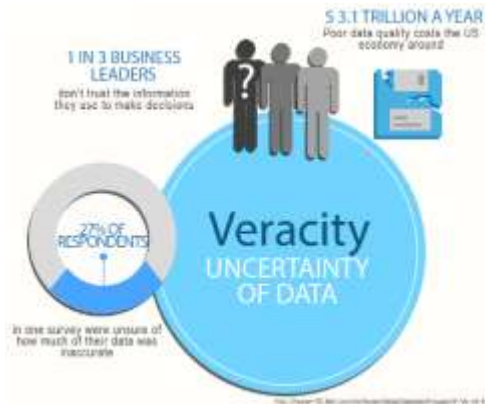
1 빅데이터의 공통적 속성 3V

빅데이터의 특징으로는 크기(Volume), 속도(Velocity), 다양성(Variety)을 들 수 있다.



2 빅데이터의 새로운 V - 정확성(Veracity)

데이터가 많아질수록 엉터리 데이터도 커질 가능성이 높아지기 때문이다, 따라서 빅데이터를 분석하는 데 있어 기업이나 기관에 수집한 데이터가 정확한 것인지, 분석할 만한 가치가 있는지 등을 살펴야 하는 필요성이 생겼고 이러한 측면에서 빅데이터의 새로운 속성인 정확성(Veracity)이 제시되고 있다.



3 빅데이터의 새로운 V - 가변성(Variability)

최근 소셜미디어의 확산으로 자기 의견을 웹을 통해 자유롭게 게시하는 것이 쉬워졌지만 실제로 자신의 의도와는 달리 자기 생각을 글로 표현하게 되면 맥락에 따라 자신의 의도가 다른 사람에게 오해를 불러일으킬 수 있다. 이처럼 데이터가 맥락에 따라 의미가 달라진다고 하여 빅데이터의 새로운 속성으로 가변성(Variability)이 제시되고 있다.



02_ 빅데이터 활용사례

이미 국내외 빅데이터 분석 대표 기업들이 지속적으로 노하우를 축적하고 고도화하면서 빅데이터를 활용해 높은 성과를 달성하고 있는 사례가 증가하고 있습니다.

- 1분동안 구글에서는 200만 건의 검색, 유튜브에서는 72시간의 비디오, 트위터에서는 27만건의 트윗이 생성됨.



- Growth of Bigdata Technology landscape 2018
빅데이터 적용 기업 2011년 150개 -> 2019년 7,040개



02_ 빅데이터 활용사례 (국내)

이미 국내외 빅데이터 분석 대표 기업들이 지속적으로 노하우를 축적하고 고도화하면서 빅데이터를 활용해 높은 성과를 달성하고 있는 사례가 증가하고 있습니다.

국내 빅데이터 활용사례

| 분야 | 업체명 | 내용 |
|----|---------------|--|
| 유통 | 신세계 인천 명품관 | 고객 구매이력 역추적해 핵심고객에 타겟마케팅으로 구매 유도 |
| | GS삼 | 홈쇼핑이력 기반한 추천 프로그램으로 구매 전환율 높여 |
| | 쿠팡 | 구매이력 외 홈피 채류, 검색 경험한 사람도 관심고객 재분류해 매출 상승 |
| 금융 | 현대카드 MY MENU | 맛집 선정 때 성별, 연령, 결제인도 등의 통합자료 제공 |
| | 삼성카드 m모켓 | 250만개 가맹점 통합 데이터 제공해 전자지갑 소비 촉진 |
| IT | 삼성전자 | 소셜데이터를 분석해 차세대 반도체 SSD 마케팅에 활용 |
| | SK텔레콤 | 기자국 정보와 제휴 카드사 데이터로 상권 분석에 활용(지오비전) |
| | SK플래닛 | 전국 5만여대 택시, 상용차량 GPS 장치 통해 정확한 교통 정보 제공(T맵) |
| | CJ E&M 계열사업본부 | 15개 핵심 지표 매일 모니터링, 예측치 벗어나는 이상 현상 발견 시 즉시 담당자에게 보고 |

< 국내 공공기관 빅데이터 사례 >

| 데이터명 | 출처 | 주요 항목 및 특징 |
|------------------------|---------------------|--|
| 구조 구급활동 및 지령시스템 원시 데이터 | 전북소방본부 | • 119종합상황실에 접수된 재난·재해별 출동 이력 원시정보 • 재난지점주소, 재해유형, 사망·사고자수, 출동인원수 등 |
| 119 구급차량 이동데이터 | | • 119 구급차량 이동(GPS) 정보 • 일자, 시간, 좌표(x,y) 등 |
| 119안전센터 시설 및 인력 현황 | | • 도내 구조구급 관련 현황 정보 • 119 안전센터, 구급차 유형별 수, 시설·분야별 인력규모 등 |
| 소통정보 | 교통정보 공개서비스 (국토교통부) | • 일별 5분 단위 교통상황 정보 • 집계일자, 속도, 교통량, 지체시간, 차량 대기길이 등 |
| 표준노드링크 | ITS국가교통정보센터 (국토교통부) | • 전국 교통망에 단일화된 ID체계를 적용한 표준교통망 DB • 도로망, 최고 제한 속도, 통과제한하중, 통과제한높이 등 |
| 주요시설 POI위치 데이터 | 유관기관공공데이터 | • 터널, 지하상가, 문화재, 소방차 불통지역, 시장, 건축물 등 |
| 인구 센서스 데이터 | 통계청 | • 5세 단위 성·연령별인구수, 가구, 주택 정보 등 |
| 사업체 조사 데이터 | | • 업종별 사업체수, 종사자수 등 |
| 표준지공시지가 | 국토교통부 | • 지번, 기준일, 공시지가, 용도지역, 토지이용상황 등 |

자료: 행정안전부와 한국정보화진흥원, '행정을 스마트하게 바꾸다'



02_ 빅데이터 활용사례 (해외)

이미 국내외 빅데이터 분석 대표 기업들이 지속적으로 노하우를 축적하고 고도화하면서 빅데이터를 활용해 높은 성과를 달성하고 있는 사례가 증가하고 있습니다.

<해외 기업의 빅데이터 분석 활용 주요 사례>

| 구분 | 기업 | 내용 |
|----------|---------------|--|
| 예측 | 포드자동차 | 운전자의 이동이력 분석, 운전자의 목적지 예측, 최적 연료배분 제안 |
| | 구글 | 검색어 분석으로 정부보다 2주 빨리 독감 유행을 예측 |
| | 다음소프트 | 북카페현상, 서울시장 보궐선거 트위터유용 분석, 트위터를 통해 본 한 국민들의 기분분석, 방송프로그램에 대한 평판분석 |
| | 아마존 | 고객의 온라인상 흔적 토대 고객취향 분석, 상품구매 등의 행동 예측 |
| 고객감성 분석 | 기아자동차 | 미국 슈퍼볼 광고 후 웹상의 고객의견을 분석, 경기도중 시청자들이 자동차에 대해 나누는 대화에서 자사비중이 4%에서 9%로 상승, 고객들의 긍정적인 신뢰성도 4%에서 18%로 상승 |
| 생산성 향상 | 네슬러 | 페이스북 등에서 고객이 원하는 맛을 선택하게 하는 투표를 실시, 개별제품 수를 한정, 수요가 적은 제품의 생산비용 절감 |
| | 월마트 | 매주 2억개 이상의 고객거래 처리, 리테일 링크라는 재고관리시스템 도입, 점포별 재고량, 시간당/일당 판매량 등의 세세한 정보 확인 가능 |
| 개인화, 맞춤화 | ING | 데이터 통합으로 더욱 효율화된 맞춤화 |
| | 아후, 다음 | 소셜미디어 분석과 맞춤화 |
| 위치정보 | 프로그래시브 (보험회사) | 차량내 운행기록장치를 통해 운전행태를 보험회사에 전송 → 위험수준에 근거한 운전자 등급화 → 가입자 맞춤서비스 제공 |
| | NTT 도코모 | 모바일 데이터를 활용해 시간단위로 인구변화 추적, 네트워크 고지 |
| 의사결정 | P&G | 200테라바이트에 달하는 전사 빅데이터를 실시간 분석하여 글로벌 시장상황 즉각 파악, 의사결정으로 연결 |
| | Zara (패션기업) | 전 세계 매장에서 본사로 유입되는 판매데이터를 바탕으로 최적화된 알고리즘을 이용해 어떤 시점에 어떤 상품이 어떤 매장에 진열되어야 하는지 실시간으로 분석 |
| 비즈니스 창출 | 스파코드 | 네델란드에서 소에 센서를 장착하여 식사량, 운동량 파악 |
| | 온스타와 GMAC보험 | 원격측정데이터를 이용하여 이동거리가 짧은 운전자에게 낮은 보험프리미엄을 부과하는 선택프로그램 개발 |
| | 구글 | [무인자동차] 차량운행 정보, 실시간 교통상황, 이용자의 운행패턴 수집 |
| | 애플 | [시리(Siri)] 센서데이터를 기반으로 단말이 개인비서 역할(음성검색) |
| | 필스-K | 기업이 자사 이미지에 부정적 영향을 실시간 파악해 원인 분석 |

주요 국가별 빅데이터 활용 전략

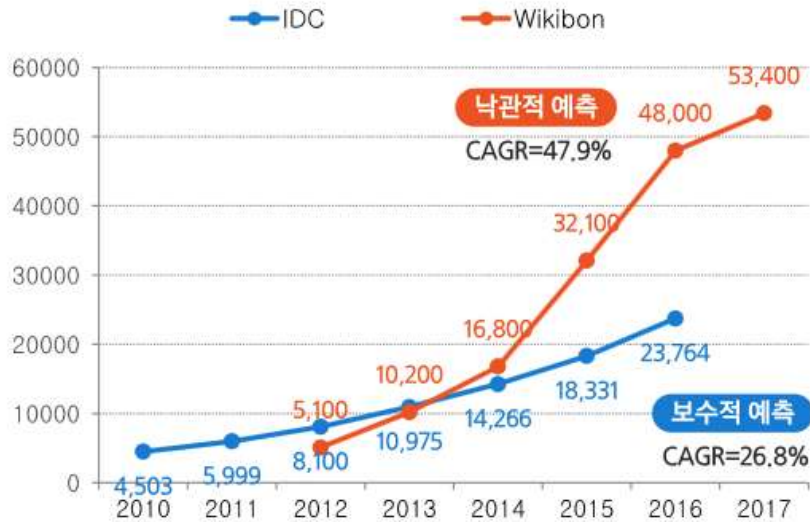
| 국가 | 전략 |
|----|--|
| 미국 | 공공데이터 축적한 사이트(www.data.gov)운영 빅데이터 이니셔티브 통해 2억 달러 규모 투자 진행 미국의 대표적인 R&D 프로그램인 NITRD 일환으로 빅데이터 연구 |
| 영국 | 오픈데이터 전략(Open Data Strategy)으로 데이터 활용 전략 통합 오픈데이터 연구소 주도로 세부과제 연구 정부 혁신 전략 중 4번째 목표로 데이터 활용 개선 선정 |
| 일본 | 내각관방 주도로 전자행정 오픈데이터 전략 추진 데이터 활성화(Active Data) 전략 일본재흥전략 2016 통해 데이터 주도 사회 실현을 목표로 제시 |
| 중국 | 2015년 빅데이터 발전 촉진을 위한 행동요강 발표 빅데이터 산업발전 계획(2016~2020) 통해 2020년까지 빅데이터 관련 제품 및 매출 1조위안 확대 목표 제시 빅데이터 응용 기술 표준 및 통계 표준 제정 |

03_ 빅데이터 시장

세계 빅데이터 시장은 연평균 26~48% 이상 성장하고 있으며, 오는 2020년까지 시장 규모가 238조원에 이를 전망입니다. 국내 빅데이터 및 분석 시장도 은행, 조립 및 공정 제조, 통신, 공공 분야 등 다양한 산업에서 지속적인 성장세가 예상되며 2020년에는 9680억원 규모에 달할 것으로 예상됩니다.

세계 빅데이터 시장규모 전망

단위: 백만 달러



빅데이터 주요 세그먼트별 국내 시장규모 전망

단위: 백만 달러



※ 출처: IDC, KISTI

04_ 취업 시장에서의 전망

데이터 분석 수요는 느는데, 빅데이터 **분석가** 가 없다



국내 데이터 직무 관련 현재 인력 대비 필요인력 비교. 한국정보화진흥원 제공



빅데이터 공급기업과 수요기업이 선정한 부문별 선호인력 순위 출처:2015년 빅데이터 시장현황조사

05_과정 소개

본 과정은 빅데이터 분야로 취업을 희망하는 ICT 전공 학생 대상으로 단기간 동안 ICT의 이해와 빅데이터 개념을 학습하고 빅데이터 분야의 직무를 체험하여 **빅데이터 전문가**로 성장할 수 있습니다.

1

학습 목표

- 응용소프트웨어 개발에 사용되는 프로그래밍 언어의 기초 문법을 적용하고, 언어의 특징과 라이브러리를 활용하여 기본 응용소프트웨어를 구현할 수 있다.
- 빅데이터 관리 및 분석의 의미를 설명할 수 있으며, 파이썬을 이용한 데이터 처리와 빅데이터 분석 알고리즘을 활용할 수 있다.
- 머신러닝을 이해하고, Tensorflow를 활용한 빅데이터를 분석할 수 있다.
- 관계형 데이터베이스에서 SQL을 사용하여 목적에 적합한 데이터를 정의하고, 조작하며, 제어할 수 있다.
- 설계된 데이터베이스 모델을 적용하기 위해 DBMS(Data Base Management System)를 설치하고, 데이터베이스와 데이터베이스 오브젝트를 계획, 설계, 생성할 수 있다.

2

학습 강점

- 빅데이터 실무를 간접 체험 해볼 수 있는 프로젝트와 취업특강으로 구성
- 연수생 주도의 빅데이터 분석 프로젝트를 수행하여 빅데이터 기술의 이해도를 높임
- 과정 수료 후 빅데이터 분석이 가능한 전문가가 될 수 있도록 기초부터 심화과정을 학습 할 수 있도록 과정을 설계
- 빅데이터 분석의 초기 작업 단계에서 중요한 데이터 수집, 저장 및 관리(Data Store)와 관련된 부분을 중점적으로 반영하여, 특히 연수생들의 전공과 융합 및 활용을 통해 향후 빅데이터 전문가로서의 커리어를 밟아 나갈 수 있는 기반을 제공

06_과정 체계

01



관계형 데이터베이스 활용 및 관리

정형화 데이터 처리를
위한 SQL 활용

데이터베이스
설계 및 구현

02



빅데이터 수집 및 처리

빅데이터 핵심 기술

R, 파이썬을 이용한
데이터 전처리

03



빅데이터 분석 및 머신러닝 활용

R을 이용한 데이터
분석 및 시각화

파이썬을 이용한
머신러닝 이해

빅데이터 분석 프로젝트

07_ 커리큘럼

| 데이터를 관리를 중심으로 빅데이터 전문가로 입문하기 위한 기본 기술 및 빅데이터 분석 기술 학습 및 습득합니다.

| 주차 | 교과목명[NCS모듈] | 세부내용[NCS교육내용] | 세부내용[실 교육내용] | 교육시간 |
|----------|--------------------|---|--|-------|
| 1주 ▶ | 특강 | <ul style="list-style-type: none"> 비NCS | <ul style="list-style-type: none"> 취업특강 OPIC특강 노동인권특강 | 20시간 |
| 1주 ~4주 ▶ | 관계형 데이터베이스 활용 및 관리 | <ul style="list-style-type: none"> SQL활용 / 응용 데이터베이스 구현 응용SW기초기술활용 논리 데이터베이스 설계 | <ul style="list-style-type: none"> SQL 주요 문법과 실행 원리 단일행 함수와 복수행 함수 테이블과 인덱스 부와 시퀀스 프로그래밍 언어 활용 공통 모듈, 서버, 배치 프로그램 구현 DBMS설치, DB오브젝트 계획,설계, 생성 | 144시간 |
| 4주 ~5주 ▶ | R, 파이썬 프로그래밍 | <ul style="list-style-type: none"> 프로그래밍 언어 활용 | <ul style="list-style-type: none"> R 기초 R을 이용한 통계 분석 파이썬 기초 파이썬 데이터 분석 패키지 이해 | 66시간 |
| 6주 ~7주 ▶ | 빅데이터 핵심기술 | <ul style="list-style-type: none"> 빅데이터 분석 및 머신러닝 활용 빅데이터 분석 프로젝트 | <ul style="list-style-type: none"> 빅데이터의 의의와 주요 기술 이해 하둡, NoSQL 데이터베이스 이해 파이썬을 이용한 데이터 처리 머신러닝 기반 데이터 분석 딥러닝 및 Tensorflow 활용 미니 프로젝트 | 70시간 |

08_활용 장비(1/2)

| 교과목명 | 활용 장비 | 사진 | 설명 |
|-----------------------|-------------|--|--|
| 관계형 데이터베이스 활용 및 관리 | Oracle DBMS |  | <ul style="list-style-type: none"> Oracle DBMS 11 express |
| R, 파이썬 프로그래밍 | Phthon |  | <ul style="list-style-type: none"> Python 3.x |
| | R |  | <ul style="list-style-type: none"> R 3.x |
| 빅데이터 핵심기술 | Hadoop |  | <ul style="list-style-type: none"> hadoop-1.2.1 |

08_활용 장비(2/2)

| 교과목명 | 활용 장비 | 사진 | 설명 |
|-----------|---------|---|---|
| 빅데이터 핵심기술 | MongoDB |  | <ul style="list-style-type: none">▪ MongoDB |
| | NoSQL |  | <ul style="list-style-type: none">▪ Oracle NoSQL Database |

Thank you

www.multicampus.com

서울 강남구 언주로 508 (주)멀티캠퍼스

multicampus

Copyright by Multicampus Co., Ltd. All right reserved

