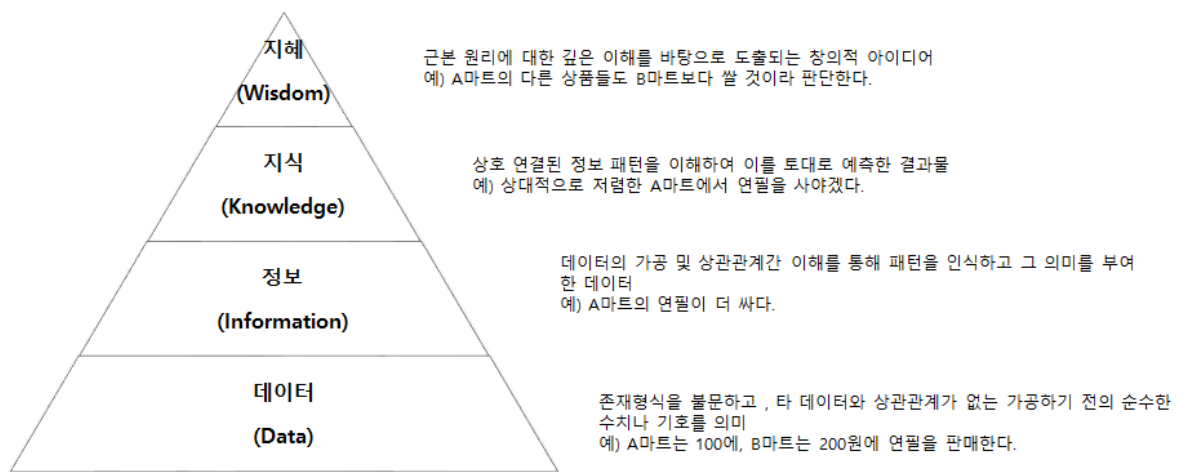


빅데이터 분석기사

빅데이터 특징

- 1.데이터 정의 - 현실 세계로부터 관찰되거나 측정되어 수집된 사실 또는 값
의미 있는 정보를 가진 모든 값
객관적 사실을 의미(당위적 특징)

2.DIKW 피라미드



*지식의 종류

암묵지 - 개인에게 축적된 내면화된 지식 (ex.노하우)

형식지 - 언어나 문자로 표준화 및 형상화된 지식 (ex.교재)

암묵지와 형식지의 상호작용



공동화 (암묵지 → 암묵지) : 개인 혹은 집단의 경험 공유
표준화 (암묵지 → 형식지) : 개인의 지식을 공유하기 위해 문자나 매체로 표현
내면화 (형식지 → 암묵지) : 교육, 훈련등으로 형식지를 개인이 취득
연결화 (형식지 → 형식지) : 개인 혹은 집단의 형식지를 상호 결합

3.빅데이터 정의

일반적으로 관리할 수 있는 규모를 넘어선 데이터를 의미
가트너 - 높은 시사점을 제공하고 더 좋은 의사결정을 위해 사용하는 정보자산
매킨지 - 일반적인 DB가 저장,관리,분석 가능한 범위를 넘어서는 규모의 데이터
IDC - 다양하고 대규모의 데이터에서 저비용으로 가치를 추출하고 데이터의 수집과 발굴을 초고속으로 처리할 수 있는 차세대 기술 및 아키텍처

빅데이터의 특징

V3 - Volume(크기), Variety(다양성), Velocity(속도)

V5 - Veracity(신뢰성(노이즈 제거)), Value(가치)

빅데이터가 만드는 변화

사전처리 → 사후처리 : 필요한 데이터만 수집 → 가능한 많은 데이터를 수집 후 조합
표본조사 → 전수조사 : 데이터 처리 비용감소 → Sampling 활용
질 → 양 : 데이터가 지속적으로 추가될 경우 양질의 정보가 증가
인과관계 → 상관관계 : 상관관계 분석이 미래 예측을 압도

빅데이터 가치

- 1.빅데이터 활용 및 가치
- 2.빅데이터 가치 선정 - 다양한 활용방식, 새로운 가치, 기술발전으로 가치 선정이 어려움
- 3.빅데이터의 영향
 - 기업 - 비즈니스 모델 혁신
 - 정부 - 미래 대응 전략 수립
 - 개인 - 기회비용 절약(생활의 스마트화)
- 4.빅데이터 위기요인과 통제방안
 - 사생활 침해 - 동의제 → 책임제

책임 훼손 원칙 - 기존 책임 원칙 강화
데이터 오용 - 분석 알고리즘 접근 허용

데이터 산업의 이해

1. 데이터 산업과 분석 인사이트

제조 : 제품 생산 분량률 개선
쇼핑 : 구매 분석을 통한 소비 예측
물류 : 물류 관리 최적화
의료 : 헬스케어 서비스를 통한 의료 복지 향상

2. 데이터 사이언스

다양한 유형의 데이터로부터 의미 있는 정보를 추출하는 분야

3. 데이터 사이언스의 구성요소

분석 + IT + 비즈니스 분석(Hard Skill - 분석 기술, 지식 / Soft Skill - 대화, 통찰력 있는 분석)

빅데이터 조직 및 인력

1. 빅데이터 업무 프로세스

빅데이터 도입 단계 - 도입 기획, 기술검토, 도입 및 조직 구성, 예산 확보
빅데이터 구축 단계 - 요구사항 분석, 설계, 구현, test 진행
빅데이터 운영 단계 - 빅데이터 플랫폼, 분석 모델, 운영 조직, 운영 예산

2. 빅데이터 조직의 구성

데이터 분석의 가치를 발굴하고 이를 활용하여 비즈니스를 최적화하는 목표 구상

3. 빅데이터 조직 구조

집중구조 - 부서별 분석 진행 & 전담 분석팀 보유(이원화, 이중화 가능성 존재)
기능구조 - 부서별 분석 진행(핵심 분석의 한계점이 명확)
분산구조 - 분산 조직을 협업부서로 배치

빅데이터 기술

1. 데이터 수집

2. 데이터 저장 : 수집된 데이터를 목적에 맞는 형태로 저장

3.데이터 분석 : 통계분석, 머신러닝, 딥러닝, 데이터 마이닝

4.데이터 활용 : 시각화, 분석 리포트, 응용 프로그램 연계

*데이터 마이닝 : 현상 분석 / 머신러닝 : 현상 파악 후 미래 예측

빅데이터 플랫폼

데이터의 수집, 저장, 분석, 활용 등 분석 프로세스를 지원하는 구격화된 빅데이터 프로세스 기술

1.방대하고 복잡한 데이터를 처리하기 위해 다양한 빅데이터 플랫폼이 개발되었다.

2.오픈 소스 기반의 분산처리 환경에는 하둡 분석플랫폼이 있다.

빅데이터 에코시스템

특정 기술이나 솔루션에 국한되지 않고 수집, 변환, 적재, 분석, 시각화 여러 단계를 거치면서 사용되는 여러 가지 기술을 이용해 플랫폼을 구축하는 과정을 의미

빅데이터 분석 프로세스

수집 → 저장 및 관리 → 처리 → 분석 → 활용 → 폐기

개인정보 법

- 1.개인식별 정보는 비식별화 조치 후 사용 가능
- 2.빅데이터 처리 사실, 목적 등의 공개를 통한 투명화
- 3.개인정보 재식별 시, 즉시 파기 및 비식별화 조치
- 4.민감정보 및 통신 비밀 수집, 이용, 분석 등 처리금지
- 5.수집된 정보의 저장, 관리 시 기술적 관리 보호조치 시행

개인정보 비식별 조치 가이드라인

데이터 활용이 증가함에 따라 개인정보 보호 강화에 대한 요구가 지속되어 개인정보 보호를 보장하면서 데이터를 활용하기 위해 만들어졌으며, 개인정보를 이용 또는 제공할 때 준수해

야 할 조치

1.가명처리 - 대체값 적용 (ex.홍길동)

2.총계처리 - 개인정보에 통계값을 적용

총합, 부분합, 라운딩, 재배열

3.데이터 삭제 - 특정 데이터 값 삭제

4.데이터 범주화 - 그룹의 대표값, 구간값 변환 (ex.30대)

5.데이터 마스킹 - 전체 또는 부분을 대체값 적용

임의 잡음 추가, 공백 대체