

소득소비 관리와 빅데이터 활용

02_빅데이터 개요 및 금융
적용_금융빅데이터 분석 필요성

정대현



부산대학교 SW중심대학

부산대학교 소프트웨어융합교육원
BUSAN NATIONAL UNIVERSITY SOFTWARE CONVERGENCE EDUCATION INSTITUTE

학습목표 및 학습내용

1

학습목표 및 학습내용

학습목표

1. 빅데이터와 핀테크 연계성에 관하여 생각해 본다
2. 데이터 계층 구조에 관하여 순서대로 나열할 수 있다.
3. 데이터 관리 방법에 관하여 설명할 수 있다.
4. 데이터베이스 종류에 대해 인지한다.
5. 디지털트윈 기술에 관하여 설명할 수 있다.

학습내용

1. 빅데이터와 핀테크
2. 데이터 계층 구조
3. 데이터와 정보의 차이점 파악
4. 참조무결성
5. 디지털트윈

Q1

데이터에 대한 정보를 담고있는 데이터를 지칭하는 용어는?

1

정보시스템

2

데이터베이스

3

빅데이터

4

메타 데이터

해설

정답 : 4

해설 : 접근 속도를 높이기 위하여 데이터에 대한 위치, 크기 등에 대한 정보를 기록한 데이터에 대한 정보 데이터를 지칭

1

Quiz

Q2

데이터의 추출, 변환, 통합 등 데이터 저장소를 지칭하는 용어는?

1

데이터웨어하우스

2

데이터베이스

3

데이터마트

4

메타 데이터

해설

정답 : 1

해설 : 데이터의 통합 저장소로써 데이터웨어하우징과 동일어

데이터

- data
- 의미 있는 단순한 사실들의 집합

정보

- information
- 해석할 수 있는 의미와 가치를 가진 자료



실제
데이터

- Sauce data

데이터
베이스

- Data base
- 한 조직의 응용 시스템들을 공용하기 위해 통합, 저장한 운영 데이터의 집합

- 파일의 위치, 크기, 소유자, 허가권

메타
데이터

- dB를 관리하기 위한 시스템

DBMS

OLTP

- on-line transaction processing
- 비즈니스 업무를 처리하는 과정에서 발생하는 데이터
- 단순한 거래 대부분의 회계시스템, 현황 파악이 목적

OLAP

- on-line analytical processing
- end-user가 다양한 분석을 하고 의사결정을 하기 위하여
요구하는 다차원 정보의 제공, DW에 저장

DW

- data warehouse, data warehousing
- 데이터 추출, 변환, 통합, 요약된 데이터베이스

DM

- data mart
- 소규모로 구축된 데이터웨어하우스

빅데이터 관리

4

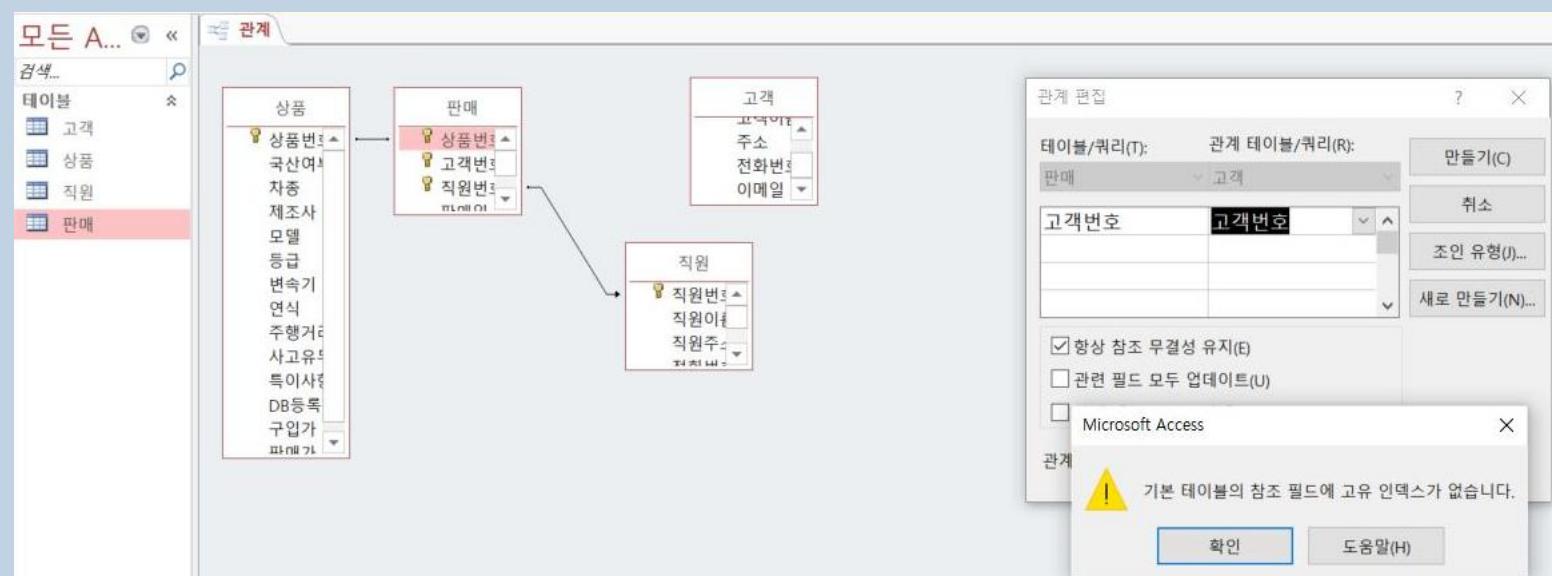
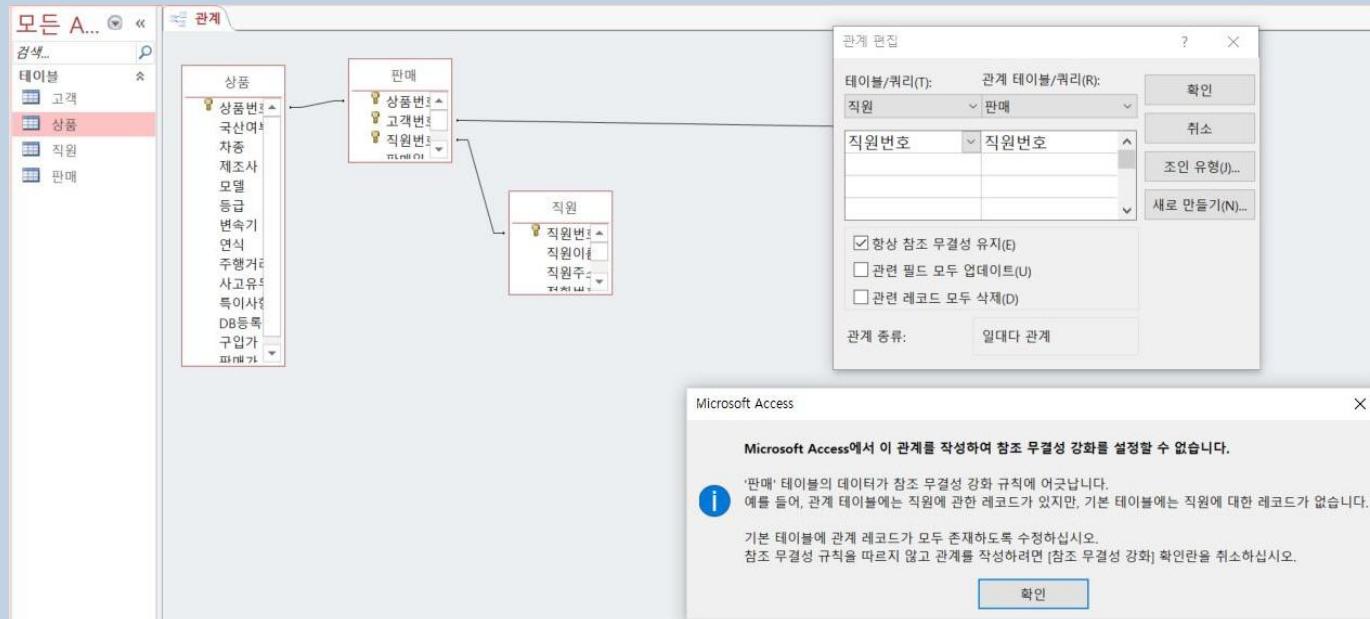
데이터 관리의 원칙

	A ID	B 시	C 구군	D 동	E 목적지
1					
2	11063661	부산	수영구	남천동	※ 좌동 대동아파트 513동
3	11063674	부산	북구	덕천동	양산신기주공아파트 103동1~2라인
4	11063684	부산광역시	금정구	부곡동	정관LH 휴먼시아4단지 411동
5	11063688	부산광역시	부산진구	부암동	※ 동삼도개공112동
6	11063720	부산시	사상구	괘법동	※ 아바니센트럴부산
7	11063749	부산시	해운대구	좌동	※ 수영로교회 교육관
8	1106 3751	부산	사상구	괘법동	※ 초량 신한은행(부산역지점)
9	1106 3793	부산	해운대구	송정동	우동 서앤장센텀내과의원
10	1106 3843	부산	부산진구	부전동	※ 부산역(KTX)
11	1106 3890	부산	동래구	명륜동	※ 덕포남영아파트 101동
12	1106 4003	부산광역시	부산진구	개금동	신평2동 621-7 초림
13	1106-4014	부산광역시	사상구	괘법동	※ 주례 보훈병원
14	1106-4048	부산시	남구	문현동	※ 대연사거리 하나은행 주차장
15	1106-4062	부산시	북구	구포동	※ 덕천동 약손지압원[구포전화국 맞은편]
16	1106-4078	부산시	기장군	정관읍	※ 우동 해운대부민병원
17	1106-4091	부산시	수영구	남천동	※ 남천2동 주민센터

빅데이터 관리

4

데이터 관리의 원칙



Legacy

- ✓ 자산, 유산
- ✓ 과거 오래된 기술, 데이터, 방법, 프로그램 (기존 시스템)

Integrity (무결성)

- ✓ 관계형 데이터베이스에서 데이터의 정확성과 일관성
- ✓ 결손과 부정함이 없음을 보증
- ✓ 주민등록번호

- 빠른 의사결정이 상대적으로 덜 요구
- 처리 복잡도가 높음
- 비정형 데이터 비중이 높음
- 처리 및 분석 유연성이 높음
- 동시 처리량이 낮음



연역적 추리

- ✓ deduction
- ✓ 일반적 원리로부터 특수한 명제(구체적 사실)를 유도하는 추론법
- ✓ 모든 사람은 죽는다 – 나도 사람이다 – 그러므로 나도 언젠가는 죽을 것이다
- ✓ 귀납을 통해 얻어진 : 대전제 – 소전제 – 결론

귀납적 추리

- ✓ induction
- ✓ 개별적 특수 사실로 미루어 보아 → 일반적 원리를 알아내는 추론법
- ✓ 전제가 모두 참일지라도, 그 결론이 반드시 참이라는 보장은 無
- ✓ A새는 난다 – B새도 난다 – C 새도 난다
- ✓ 따라서 모든 새는 날 수 있다
- ✓ 귀납논증은 ‘참, 거짓’의 문제가 아니라
- ✓ ‘설득력’이 얼마나 있느냐 없느냐를 따져야 할 문제 (논증의 강도)

비즈니스 모델 비즈니스 프로세스

- ✓ 수익 모델, 기업으로 하여금 수익을 유지하게 하는 일련의 활동
- ✓ Business Process, 고객가치 창출을 위한 일련의 경영활동

BI

- ✓ Business Intelligence
- ✓ 의사결정에 대한 지원에 사용되는 기술, 프로세스, 스킬, 응용프로그램 등을 모두 포함하는 포괄적인 용어
- ✓ 과거 성과를 분석하고, 미래를 예측할 수 있는 운영상의 데이터를 처리하는 시스템

학습 후 Quiz

1

Quiz

Q1

데이터와 정보는 동일어이다 (O, X)

X

Q2

관계형 데이터베이스에서 데이터의 정확성과 일관성을 뜻하는 용어는?

참조무결성

Q3

데이터 계층구조? bit – Byte – field – record – () – dB

파일 (file)

Q4

()에 적당한 정보의 진화 단계는? 데이터 – 정보 – () – 지혜

지식



1. 데이터는 의미 있는 단순한 사실들의 집합 (cf. 정보는 해석할 수 있는 가치를 가진 data 집합)

2. 참조무결성은 데이터베이스화에서 물리적 오류가 없는 신뢰할 수 있는 데이터

3. 빅데이터는 빠른 의사결정이 덜 요구되어지는 비정형 데이터 분석으로 가치를 창출

4. 소규모로 구축된 데이터웨어하우스를 데이터 mart라 칭한다

5. 데이터 관리의 효율성은 데이터베이스화를 통해 가능

6. 정보의 진화 단계 : 데이터 – 정보 – 지식 - 지혜

1

Quiz

Q1

언어로 명시할 수 있는 객관적 지식을 뜻하는 용어는?

1

형식지

2

암묵지

3

지식경영

4

지식공동체

해설

정답 : 1

해설 : 일반적 문서 형태의 형식적 지식을 지칭함

Q2

디지털 트윈 기술에 관한 설명으로 틀린 것은?

1

컴퓨터에 현실 속 쌍둥이를 만들어 미리 예측 기능을 수행

2

인공지능 기술을 활용

3

시뮬레이션과 동일어로 사용

4

의사결정 지원을 위한 시스템의 가상 모델

해설

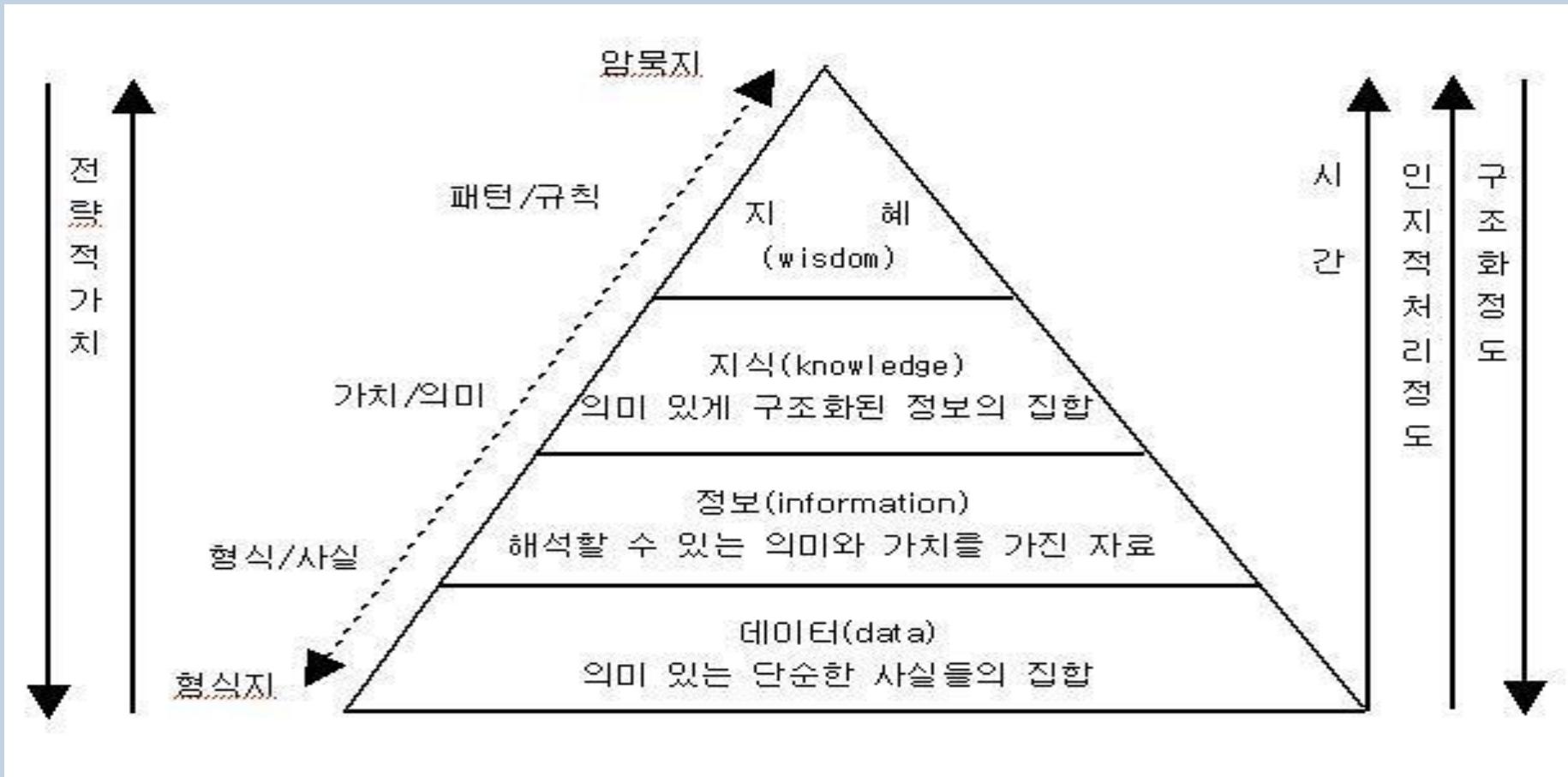
정답 : 3

해설 : 시뮬레이션의 일종이지만, 현실세계와 사실적으로 연동한다는 측면에서 별개의 기술

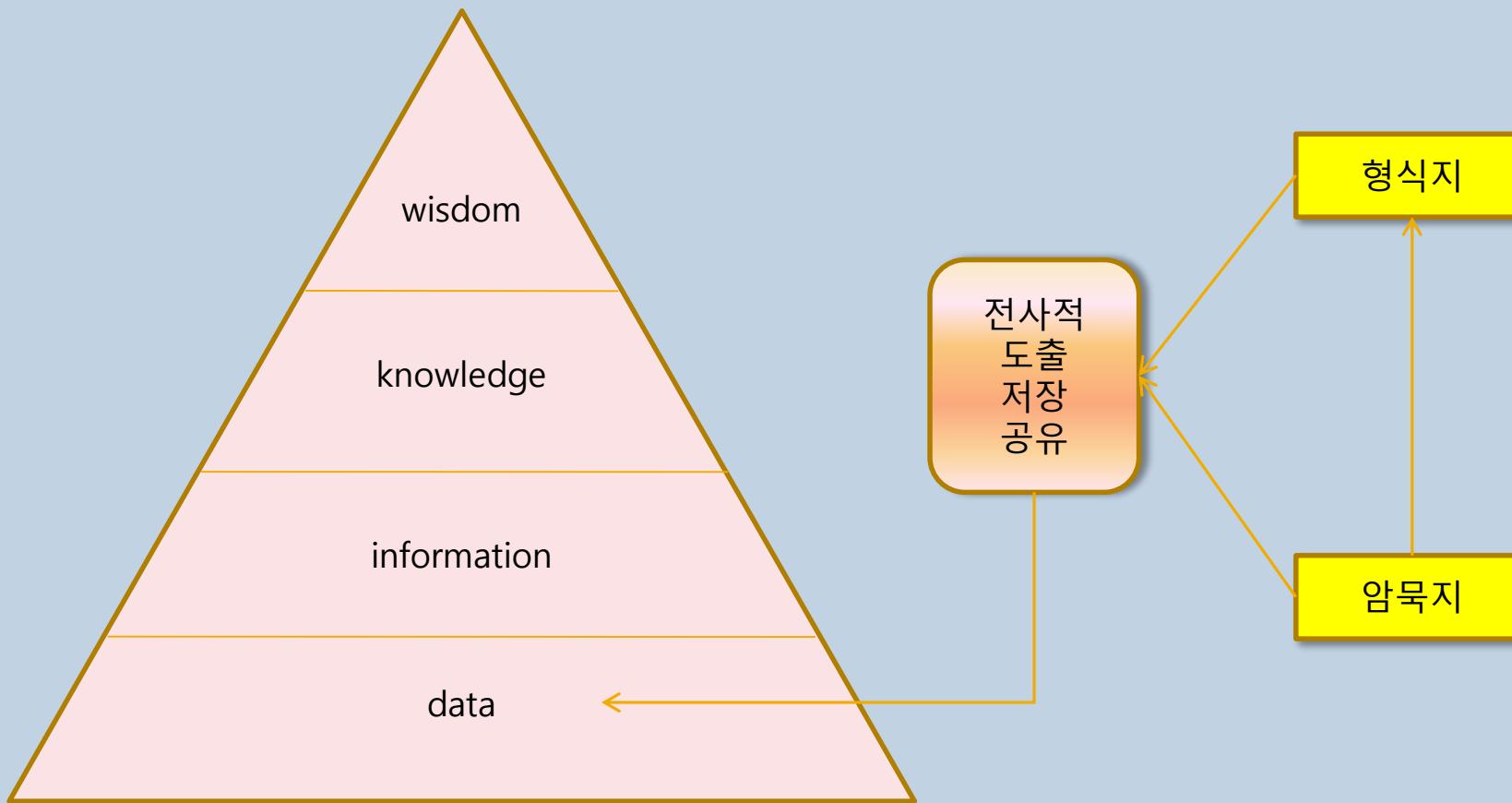
데이터 종류별 특성

1

데이터 종류와 단계



- 형식지(Explicit knowledge) : 언어로 명시할 수 있는 객관적이고 형식적인 지식
(데이터 정책, 절차, 소프트웨어, 문서 등)
- 암묵지(tacit knowledge) : 개인적인 경험에 의해 얻어지는 말로 표현하기 어려운 지식으로
매우 함축적인 노하우인 경험적 학습의 영역을 의미함



정의

- 라이프사이클 전반에 걸쳐 실시간 데이터를 사용하여 업데이트되고, 시뮬레이션, 머신 러닝, 추론을 통하여 의사결정을 돋는 시스템의 가상 모델

기술

- 인공지능 및 머신 러닝 기술 활용
- 실제 대상의 변화를 반영하도록 실시간 업데이트화

구분

- 시뮬레이션 : 일반적으로 하나의 특정 프로세스를 연구
- 디지털트윈 : 시뮬레이션의 일종, 현실세계의 변화와 사실적으로 연동 가능

효과

- 디지털전환은 정보화, 지능화, 자율동작 등을 통해 모든 것을 변화시킬 수 있다
- 이러한 변화와 전환을 사용자들에게 표출하는 수단으로써의 기능 수행

제품 설계 효율화

- 실제 환경의 데이터를 활용하여 제품 설계 시뮬레이션을 수행함으로써 효율성을 극대화

예지 보전

- 물리 시스템의 운영 상태를 지속적으로 파악하는 과정에 정상적 동작 모델을 벗어나는 상태를 파악함으로써 문제가 발생하기 전에 대응 조치 강구

실시간 모니터링과 사전 대응

- 예지 보전은 점진적 변화에 대한 사전 대응 조치를 지원하는 반면, 실시간 모니터링으로 돌발적인 상태 변화에 따른 긴급 대응 조치를 동작 모델을 제공

유지보수 효율성 개선

- 복잡한 시스템을 대상으로 유지보수를 수행할 때 작업자에게 디지털 트윈을 통해 정보를 제공함으로써 직관적 이해를 돋고 조치 방법 등 제반 정보 제공

프로세스 최적화

- 시스템의 동작 모델이 조건에 따른 시뮬레이션이 가능하며, 실제 동작 상태와 비교하면서 최적화 운영을 지원

원인 분석

- 시스템의 동작 상태에 대한 데이터를 이용하여 이미 발생한 문제에 대해 재현 시뮬레이션을 함으로써 원인 분석을 가능하게 함

다자간 의사결정

- 디지털 트윈의 연합적 연동을 통해 공동 문제 또는 복합 문제에 대한 원인을 보다 쉽게 파악하고, 산업 생태계 전반에 걸쳐 이해관계자 간에 협업을 지원함

학습 후 Quiz

1

Quiz

Q1

개인적 경험에 의해 획득한 매우 합축적이며 노하우인 경험적 지식을 뜻하는 용어는?

암묵지

Q2

암묵지를 형식지화하여 전자적으로 공유하는 것이 중요하다
(O, X)

O

Q3

컴퓨터에 현실 속 쌍둥이 모델을 만들어 미래를 예측 하는 기술을
뜻하는 용어는?

디지털 트윈

Q4

디지털 트윈 기술은 현실세계의 변화와 사실적으로 연동이
불가능하다 (O, X)

X



1. 데이터 종류는 크게 형식적 지식과 암묵적 지식으로 분류 한다
2. 암묵지를 형식지화하여 전사적 공유를 통해 데이터-정보-지식-지혜로 발전시킬 수 있다
3. 디지털 트윈기술은 의사결정을 돋는 시스템의 가상 모델이다
4. 디지털 트윈은 단순한 시뮬레이션의 일종이지만, 실제 대상의 변화를 반영하도록 실시간 분석가능
5. 디지털 트윈은 실시간 모니터링, 사전대응, 프로세스 최적화, 원인 분석, 다자간 의사결정 지원
6. 파괴적 혁신을 리더하기 위해 트렌드를 파악하고 주시하는 것이 중요하다