## 데이터분석 전문가 가이드

# 과목 1. 데이터이해제 1장 데이터의이해

출처: 데이터분석 전문가 가이드, 한국데이터베이스진흥원

### 데이터분석 전문가 가이드

- ■데이터분석 전문가( <a href="http://www.dbguide.net/da.db?cmd=snb\_adp\_1">http://www.dbguide.net/da.db?cmd=snb\_adp\_1</a>) 자격증 취득을 위한 수 험서
- ■국가 공인 자격 : 공무원 또는 공공기관에 지원 할때 도움
- •기타 좋은점: 데이터분석에 대한 전반적인 과정을 이해할 수 있음

| 미데이터 분석 전문가의 직무 |   |  |
|-----------------|---|--|
| 직무              | 수행 내용   |  |
| 데이터 기획          | • 비즈니스 목표 달성을 위해 내부 업무 프로세스를 기반으로 다양한 분석기회를 발굴하여 분석의 목표를<br>정의하고, 분석대상 도출 및 분석 결과 활용 시나리오를 정의하여 분석과제를 체계화 및 구체화하는 빅데<br>이터 분석과제 정의, 분석로드맵 수립, 성과 관리 등을 수행한다.  |  |
| 데이터 분석          | • 분석에 대한 요건을 구체적으로 도출하고, 분석과정을 설계하고, 요건을 실무담당자와 합의 하는 요건정의, 모델링, 검증 및 테스트, 적용 등을 수행한다.  |  |
| 데이터 시각화         | <ul> <li>다양한 데이터들을 대상으로 어떤 요소를 시각화 해야 하는지 정보 구조를 분석하며 어떤 형태의 시각화<br/>모델이 적합한지 시각화에 대한 요건을 정의하고 시나리오를 개발하는 시각화 기획, 모델링, 디자인, 구축, 배포 및 유지보수 등을 수행한다.</li> </ul> |  |

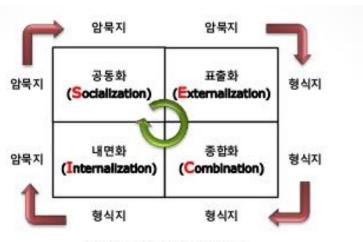
## 제 1절 데이터와 정보

- \* 데이터의 정의
- •라틴어인 dare(주다, to give)의 과거분사형으로 '주어진 것' 이란 의미.
- •옥스퍼드 대사전 => 추론과 추정의 근거를 이루는 사실 이는 데이터를 단순한 객체로서 가 치뿐만 아니라 다른 객체와의 상호관계 속에서 가치를 갖는 것으로 설명한 것
- •데이터는 "객관적인 사실이라는 존재적 특성을 갖는 동시에 추론/예측/전망/추정을 의한 근거"

#### [표 Ⅱ-1-31] 데이터 유형 변환의 종류

| 종류                         | 설명                                  |  |
|----------------------------|-------------------------------------|--|
| 명시적(Explicit)<br>데이터 유형 변환 | 데이터 변환형 함수로 데이터 유형을 변환하도록 명시해 주는 경우 |  |
| 암시적(Implicit)<br>데이터 유형 변환 | 데이터베이스가 자동으로 데이터 유형을 변환하여 계산하는 경우   |  |

## 제 1절 데이터와 정보



<지식전환의 4단계(노나카 이쿠지로, 1996)>

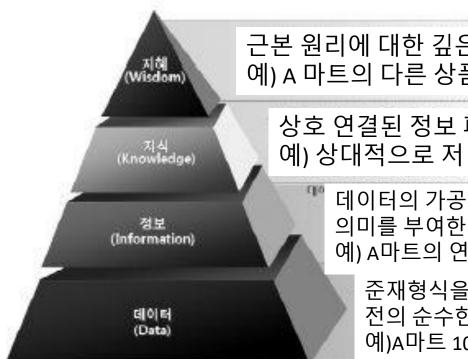
| 암묵지  | 형식지   |  |
|--|---|--|
| 말로 표현할 수 없거나 표현하기 어려운 지식     경험과 모감을 통해 얻은 직접적인 지식     몸으로 체득한 지식     개인적, 주관적, 정서적, 감성적     아날로그 지식     특정한 사람이나 장소, 대상 등에 한정 | <ul> <li>말로 명확히 표현되는 지식</li> <li>체계적인 지식</li> <li>매뉴얼화 된 지식</li> <li>사회적, 객관적, 이성적, 논리적</li> <li>디지털 지식</li> <li>정보기술을 통해 이동, 재활용 가능</li> </ul> |  |

- 1) 공동화(Socialization) 개인의 암묵지가 다른 개인의 암묵지와 만나 또 다른 암묵지를 창조하는 과정 2) 표출화(Externalization) 공동화 과정을 통해 생성된 암묵지를 형식지화 하는 과정 3) 연결화(Combination) 표출화되어 생성된 형식지들이 체계적으로 연결되는 과정

- 4) 내면화Internalization) 연결된 고부가가치 지식을 개인이 받아들여 개인의 암묵지와 연결하여 새로운 지식으로 내면에 쌓아두는 과정

※ 공동화 → 표출화 → 연결화 → 내면화를 지속적으로 반복하면서 지식의 깊이가 깊어짐

## 제 1절 데이터와 정보



근본 원리에 대한 깊은 이해를 바탕으로 도출되는 창의적 아이디어예) A 마트의 다른 상품들도 B마트보다 쌀 것이라고 판단

상호 연결된 정보 패턴을 이해하여 이를 토대로 예측한 결과물예) 상대적으로 저렴한 A마트에서 연필을 사야겠다.

데이터의 가공 및 상관관계간 이해를 통해 패턴을 인식하고 그의미를 부여한 데이터에) A마트의 연필가격이 더 싸다.

준재형식을 불문하고 타 데이터와의 상관관계가 없는 가공하기 전의 순수한 수치나 기호를 의미 예)A마트 100원에, B마트는 200원에 연필을 판매

(그림 1-2) 지식 피라미드

※ 출처 : 한국DB진흥원, DB산업 선진화 법제 개선 방안 연구, 2009.12, p.10

## 제 2절 데이터베이스 정의와 특징

\* 데이터베이스 정의

문자, 기호, 음성, 화상, 영상 등 상호 관련된 다수의 콘텐츠를 정보 처리 및 정보통신 기기에 의하여 체계적으로 수집/축적하여 다양한 용도와 방법으로 이용할 수 있도록 정리한 집합체

| 일반 특징    | 내용  |
|----------|---|
| 통합된 데이터  | 동일한 내용의 데이터가 중복되어 있지 않다는 것.<br>데이터 중복은 관리상의 복잡성을 초래                     |
| 저장된 데이터  | 컴퓨터가 접근할 수 있는 저장 매체(HDD)에 저장되는 것.<br>컴퓨터 기술을 바탕으로 한 것                   |
| 공용 데이터   | 여러 사용자가 서로 다른 목적으로 데이터베이스의 데이터를 공동<br>으로 이용한다는 것                        |
| 변화되는 데이터 | 데이터베이스의 상태는 새로운 데이터의 삽입, 기존 데이터의 삭제, 갱신으로 항상 변화하면서도 항상 현재의 정확한 데이터를 유지. |

## 제 2절 데이터베이스 정의와 특징

데이터 베이스는 다양한 측면에서 특성을 규정해 볼 수 있음.

| 다양한 측면         | 내용   |
|----------------|--|
| 정보의 축적 및 전달 측면 | 대량의 정보를 일정한 형식에 따라 컴퓨터 등의 정보처리기기가 읽고 쓸 수 있도록<br>하는 기계가독성과 다양한 방법으로 필요한 정보를 검색할 수 있는 검색 가능성, 그<br>리고 정보통신망을 통하여 원거리에서도 즉시 온라인으로 이용할 수 있는 원격조작<br>성을 갖음. |
| 정보 이용 측면       | 이용자의 정보 요구에 따라 다양한 정보를 신속하게 획득할 수 있고 원하는 정보를<br>정확하고 경제적으로 찾아낼 수 있다는 특성을 갖음.   |
| 정보 관리 측면       | 정보를 일정한 질서와 구조에 따라 정리/저장하고 검색/관리할 수 있도록 하여 방대한 양의 정보를 체계적으로 축적하고 새로운 내용 추가나 갱신이 용이   |
| 정보기술 발전의 측면    | 정보처리, 검색/관리 소프트웨어, 관련 하드웨어, 정보 전송을 위한 네트워크 기술 등의 발전을 견인  |
| 경제/산업적 측면      | 다양한 정보를 필요에 따라 신속하게 제공/이용할 수 있는 인프라로서 특성을 가지고 있어 경제, 산업, 사회 활동의 효율성을 제고하고 국민의 편의를 증진하는 수단으로서 의미  |