

논리 데이터베이스 설계

데이터베이스 정규화 개요



학습내용

- 정규화의 개념 및 특성
- 원자성과 함수적 종속성
- 정규화 대상 개체 도출

학습목표

- 정규화의 개념 및 특성을 설명할 수 있다.
- 원자성과 함수적 종속성의 개념을 설명할 수 있다.
- 정규화 대상 개체를 도출할 수 있다.

정규화의 개념 및 특성



1 데이터의 논리적 표현

1 이상(Anomaly)의 개념

- 관계 스키마의 잘못된 설계는 **데이터의 불필요한 중복**을 발생시킴
- 데이터의 중복으로 인해 **이상현상(Anomaly)** 발생

정규화의 개념 및 특성

1 데이터의 논리적 표현

2 이상(Anomaly)의 종류

삭제 이상

튜플을 삭제함으로써 유지되어야 하는 정보까지도 연쇄 삭제되는 정보의 손실

삽입 이상

어떤 데이터를 삽입하려고 할 때 불필요하고 원하지 않는 데이터도 함께 삽입해야만 삽입이 되는 현상

갱신 이상

중복된 튜플 중에서 일부 튜플의 애트리뷰트 값만을 갱신시킴으로써 정보의 모순성이 생기는 현상

정규화의 개념 및 특성

1 데이터의 논리적 표현

3 이상(Anomaly)의 종류 예제 (기본 키 : 학번, 과목번호)

학번	과목번호	성적	학년
100	C413	A	4
100	E412	A	4
200	C123	B	3
300	C312	A	1
300	C324	C	1
300	C413	A	1
400	C312	A	4
400	C324	A	4
400	C413	B	4
400	E412	C	4
500	C312	B	2

삭제 이상

만일 학번이 200인
학생이 과목
C123을 취소하여
이 튜플을 삭제할
경우 **학년 3이라는
정보까지 함께
삭제되는 문제**

정규화의 개념 및 특성

1 데이터의 논리적 표현

3 이상(Anomaly)의 종류 예제 (기본 키 : 학번, 과목번호)

학번	과목번호	성적	학년
100	C413	A	4
100	E412	A	4
200	C123	B	3
300	C312	A	1
300	C324	C	1
300	C413	A	1
400	C312	A	4
400	C324	A	4
400	C413	B	4
400	E412	C	4
500	C312	B	2

삽입 이상

만일 학번이 600인
학생이 2학년이라는
정보를 삽입하려고
할 때 **교과목을**
등록하지 않으면
삽입 불가능해지는
문제

정규화의 개념 및 특성

1 데이터의 논리적 표현

3 이상(Anomaly)의 종류 예제 (기본 키 : 학번, 과목번호)

학번	과목번호	성적	학년
100	C413	A	4
100	E412	A	4
200	C123	B	3
300	C312	A	1
300	C324	C	1
300	C413	A	1
400	C312	A	4
400	C324	A	4
400	C413	B	4
400	E412	C	4
500	C312	B	2

갱신 이상

만일 학번이 400인
학생의 학년을
4에서 3으로
변경하고자 할 때
**모두 4번의 갱신이
필요**하게 되는 문제

정규화의 개념 및 특성

1 데이터의 논리적 표현

4 이상(Anomaly)의 원인과 해결

이상의 원인

- 애트리뷰트들 간에 존재하는 여러 데이터 종속 관계를 무리하게 하나의 릴레이션으로 표현

이상의 해결

- 애트리뷰트들 간의 종속성을 분석해서 하나의 릴레이션에는 기본적으로 하나의 종속성 표현되도록 분해 (정규화(Normalization))

정규화의 개념 및 특성

1 데이터의 논리적 표현

5 스키마 변환의 세 가지 원리

정보 표현의
무손실

하나의 스키마에서 다른 스키마로
변환시킬 때 **정보의 손실이 있어서는
안 됨**(Nonloss Representation of
Information)

최소의
데이터 중복
허용

최소의 데이터 중복을 허용함으로써,
중복으로 인한 여러 가지 이상 제거
(Minimal Data Redundancy)

분리의 원칙

하나의 독립된 관계성은 **별도의
릴레이션으로 분리**시켜 표현
(Principle of Separation)

정규화의 개념 및 특성



2 정규화의 개념

- 1 애트리뷰트 간 종속성(Dependency)을 분석해서 하나의 종속성이 하나의 릴레이션으로 표현되도록 분해(Decomposition)하는 과정
- 2 릴레이션의 애트리뷰트, 엔티티, 관계성을 파악하여 데이터의 중복성을 최소화하는 과정
- 3 정규화를 통해 릴레이션을 분해하면 연산 시간 증가 (조인연산 필요)

정규화의 개념 및 특성

3 정규화의 특성

- 1 정규형을 만드는 목적은 데이터베이스 연산의 여러 가지 이상을 없애기 위함
- 2 데이터베이스의 물리적 구조나 물리적 처리에 영향을 주는 것이 아니라, **논리적 처리 및 품질에 큰 영향**을 미침
- 3 개체(Entity) 간의 관계를 쉽게 파악하여 **데이터 구조의 안정성 최대화**
- 4 **저장 공간 최소화**

원자성과 함수적 종속성

1 원자성의 개념

- 1 하나의 릴레이션에 저장하기 위해 **정보를 가장 작은 단위로** 쪼개는 것
- 2 원자적 데이터로 구성된 열은 오직 **하나의 데이터만** 가져야 함을 의미
- 3 데이터를 원자적으로 만드는 것이 릴레이션을 정규화하는 첫 단계(제 1 정규형)

원자성과 함수적 종속성

2 함수적 종속성의 개념

1 정의

- 어떤 릴레이션 R에서 X와 Y를 각각 R의 애트리뷰트 집합의 부분집합이라 하면, 애트리뷰트 X의 값 각각에 대해 시간에 관계없이 항상 애트리뷰트 Y의 값이 오직 하나만 연관되어 있을 때 **Y는 X에 함수 종속**이라 하고 **$X \rightarrow Y$** 로 표기
- $X \rightarrow Y$ 의 관계를 성립시키는 X를 **결정자**라고 하고, Y를 **종속자**라고 함

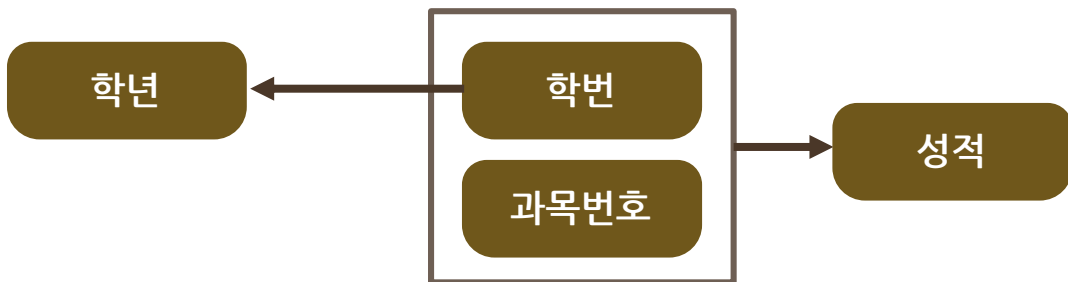
원자성과 함수적 종속성

2 함수적 종속성의 개념

2 특성

함수 종속
다이아그램
(FD diagram)

한 릴레이션에서 애트리뷰트들 간의
복잡한 함수 종속 관계를 쉽게
이해하기 위해 도식으로 표현한 것



원자성과 함수적 종속성

2 함수적 종속성의 개념

3 함수종속의 추론규칙(Armstrong's Axioms)

재귀/반사 규칙
(Reflectivity Rule)

$X \supseteq Y$ 이면 $X \rightarrow Y$ 이고
 $X \rightarrow X$ 이다.

증가/첨가 규칙
(Augmentation Rule)

$X \rightarrow Y$ 이면 $WX \rightarrow WY$ 이고
 $WX \rightarrow Y$ 이다.

이행 규칙
(Transitive Rule)

$X \rightarrow Y$ 이고 $Y \rightarrow Z$ 이면
 $X \rightarrow Z$ 이다.

유니온/결합 규칙
(Union Rule)

$X \rightarrow Y$ 이고 $X \rightarrow Z$ 이면
 $X \rightarrow YZ$ 이다.

분해 규칙
(Decomposition Rule)

$X \rightarrow YZ$ 이면
 $X \rightarrow Y$ 와 $X \rightarrow Z$ 이다.

가이행 규칙(Pseudo
Transitivity Rule)

만일 $W \rightarrow X$ 이고
 $XY \rightarrow Z$ 이면 $WY \rightarrow Z$ 이다.

정규화 대상 개체 도출

1 개체 도출 방법

1 개체와 관련된 **양식, 문서, 장표** 수집

2 양식, 문서, 장표로부터 **소스 자료**를 도출

- 소스 자료 도출 시 임의대로 해당 항목을 생략을 하거나, 제거해서는 안됨
- 빈 양식보다는 실제 데이터가 표현되어 있는 **활용 중인** 양식, 문서, 장표를 대상으로 도출
- 그 이유는 **실 데이터의 구조를 파악**하기에 유리할 뿐만 아니라 양식, 문서, 장표에 포함되는 **생략된 정보를 파악**하기에 유리함

정규화 대상 개체 도출

1 개체 도출 방법

예시

취미를 가진
사람들의 속성 분석

주민등록번호,
이름, 주소, 연락처,
취미코드, 취미명
등이 도출됨



취미, 사람으로
분류하여 배치

- 취미(취미코드,
취미명)
- 사람
(주민등록번호,
이름, 주소, 연락처)

정규화 대상 개체 도출

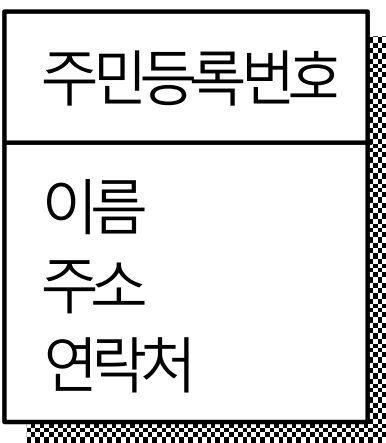
1 개체 도출 방법

예시

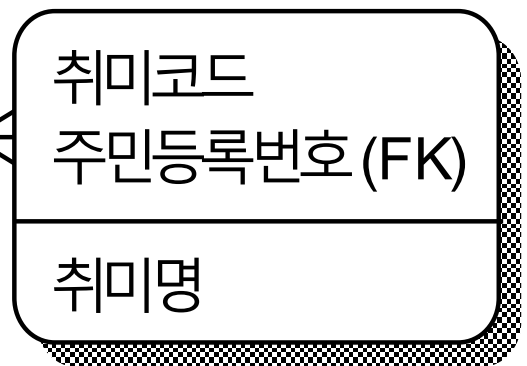
주민등록번호	이름	주소	연락처	취미 코드	취미명
123456-1234567	홍길동	서울	02-123-4567	H100	등산
				H200	골프
654321-7654321	홍길순	인천	032-123-4567	H100	등산
				H300	요리



사람

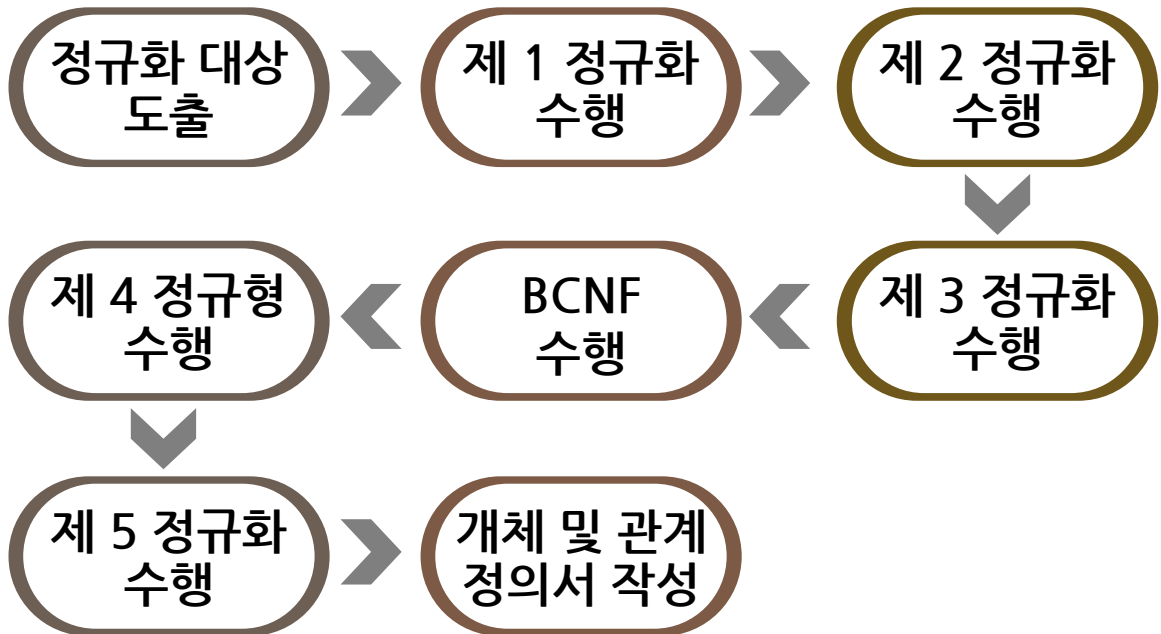


취미



정규화 대상 개체 도출

2 수행 순서



학습정리

1. 정규화의 개념 및 특성

- 이상(Anomaly)의 개념
 - 관계 스키마의 잘못된 설계는 데이터의 불필요한 중복을 발생시키며 이로 인해 이상현상(Anomaly) 발생
- 이상(Anomaly)의 종류
 - 삭제 이상, 삽입 이상, 갱신 이상
- 이상(Anomaly)의 해결
 - 애트리뷰트들 간의 종속성을 분석해서 하나의 릴레이션에는 기본적으로 하나의 종속성 표현되도록 분해(정규화(Normalization))
- 스키마 변환의 세 가지 원리
 - 정보 표현의 무손실(Nonloss Representation of Information), 최소의 데이터 중복(Minimal Data Redundancy) 허용, 분리의 원칙(Principle of Separation)
- 정규화
 - 애트리뷰트 간 종속성(Dependency)을 분석해서 하나의 종속성이 하나의 릴레이션으로 표현되도록 분해(Decomposition)하는 과정

학습정리

2. 원자성과 함수적 종속성

- 원자성
 - 하나의 릴레이션에 저장하기 위해 정보를 가장 작은 단위로 쪼개는 것
- 함수적 종속성
 - 어떤 릴레이션 R에서 X와 Y를 각각 R의 애트리뷰트 집합의 부분집합이라 하면, 애트리뷰트 X의 값 각각에 대해 시간에 관계없이 항상 애트리뷰트 Y의 값이 오직 하나만 연관되어 있을 때 Y는 X에 함수 종속이라 하고 $X \rightarrow Y$ 로 표기
- $X \rightarrow Y$ 의 관계를 성립시키는 X를 결정자(Determinant)라고 하고, Y를 종속자(Dependent)라고 함
- 함수종속의 추론규칙(Armstrong's Axioms)
 - 재귀/반사 규칙(Reflexivity Rule),
증가/첨가 규칙(Augmentation Rule),
이행 규칙(Transitive Rule),
유니온/결합 규칙(Union Rule),
분해 규칙(Decomposition Rule),
가이행 규칙(Pseudo Transitivity Rule)

학습정리

3. 정규화 대상 개체 도출

- 개체 도출 방법
 - 개체와 관련된 양식, 문서, 장표 수집 및 양식, 문서, 장표로부터 소스 자료를 도출
- 수행 순서
 - 정규화 대상 도출 → 제 1정규화 수행 → 제 2정규화 수행 → 제 3정규화 수행 → BCNF 수행 → 제 4정규화 수행 → 제 5정규화 수행 → 개체 및 관계 정의서 작성