

논리 데이터베이스 설계



데이터베이스 정규화 기법

학습내용

- 정규화의 필요성
- 정규화 기법

학습목표

- 정규화의 필요성을 설명할 수 있다.
- 정규화 기법을 이해하고, 정규화를 수행할 수 있다.

정규화의 필요성



1 정규화 모델의 특징

- 1 개체 관계의 **정확성**
- 2 데이터의 **일치성**
- 3 데이터 모델의 **단순성**
- 4 개체에 존재하는 속성의 **비중복성**

정규화의 필요성



2 관계형 DB 설계의 가이드라인

- 1 서로 다른 개체 타입(Entity Type)들을 하나로 혼합하지 말 것
- 2 중복이나 갱신 부작용이 발생하지 않게 할 것
- 3 널 값(Null Value) 발생을 될 수 있는 한 피할 것
- 4 위조 튜플(Spurious Tuple)들의 발생을 피할 것

정규화 기법

1 제 1 정규형

정의

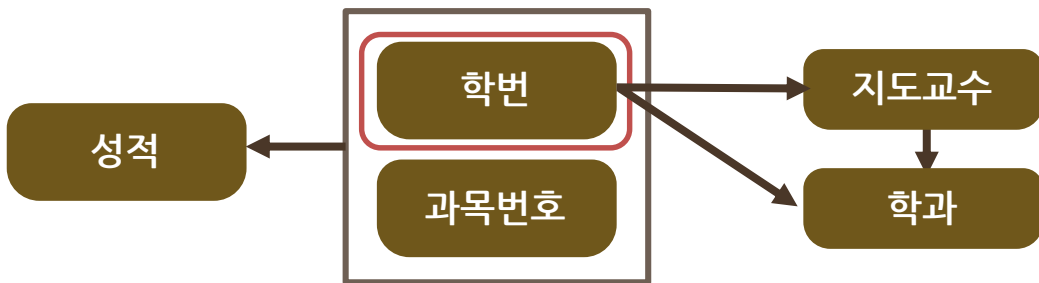
- 어떤 릴레이션 R에 속한 모든 도메인이 **원자 값(Atomic Value)** 만으로 구성
- 모든 **정규화 릴레이션**은 제 1 정규형에 속함
- 제 1 정규형에서 이상 현상의 원인은 **기본 키에 부분 함수 종속된 애트리뷰트가 존재**하기 때문임

정규화 기법

1 제 1 정규형

예시

- 수강지도 (학번, 지도교수, 학과, 과목번호, 성적)
 - 기본 키 : (학번, 과목번호)
 - 함수 종속 : (학번, 과목번호) \rightarrow 성적
 - 학번 \rightarrow 지도교수
 - 학번 \rightarrow 학과
 - 지도교수 \rightarrow 학과



정규화 기법

1 제 1 정규형

수강지도	학번	지도교수	학과	과목번호	성적
	100	P1	컴퓨터	C413	A
	100	P1	컴퓨터	E412	A
	200	P2	전기	C123	B
	300	P3	컴퓨터	C312	A
	300	P3	컴퓨터	C324	C
	300	P3	컴퓨터	C413	A
	400	P1	컴퓨터	C312	A
	400	P1	컴퓨터	C324	A
	400	P1	컴퓨터	C413	B
	400	P1	컴퓨터	E412	C

삽입 이상

학생의 지도교수
삽입 시 과목번호 필요

삭제 이상

과목 취소 시 지도교수
정보 연쇄 삭제

갱신 이상

400번 학생 지도교수
변경 시 4번 변경

정규화 기법

2 제 2 정규형

정의

- 어떤 릴레이션 R이 1NF이고, 키(기본)에 속하지 않은
애트리뷰트 모두가 **기본 키에 완전 함수 종속**
(부분 함수 종속 제거)

- 제 2 정규형은 부분 함수 종속을 제거하여
완전 함수 종속만 남도록 하는 것
- 2NF에서는 함수 종속 관계 $A \rightarrow B$, $B \rightarrow C$ 이면 $A \rightarrow C$ 가
성립하는 **이행적 함수 종속(Transitive FD)**이
존재하는데, 이는 이상 현상의 원인이 됨

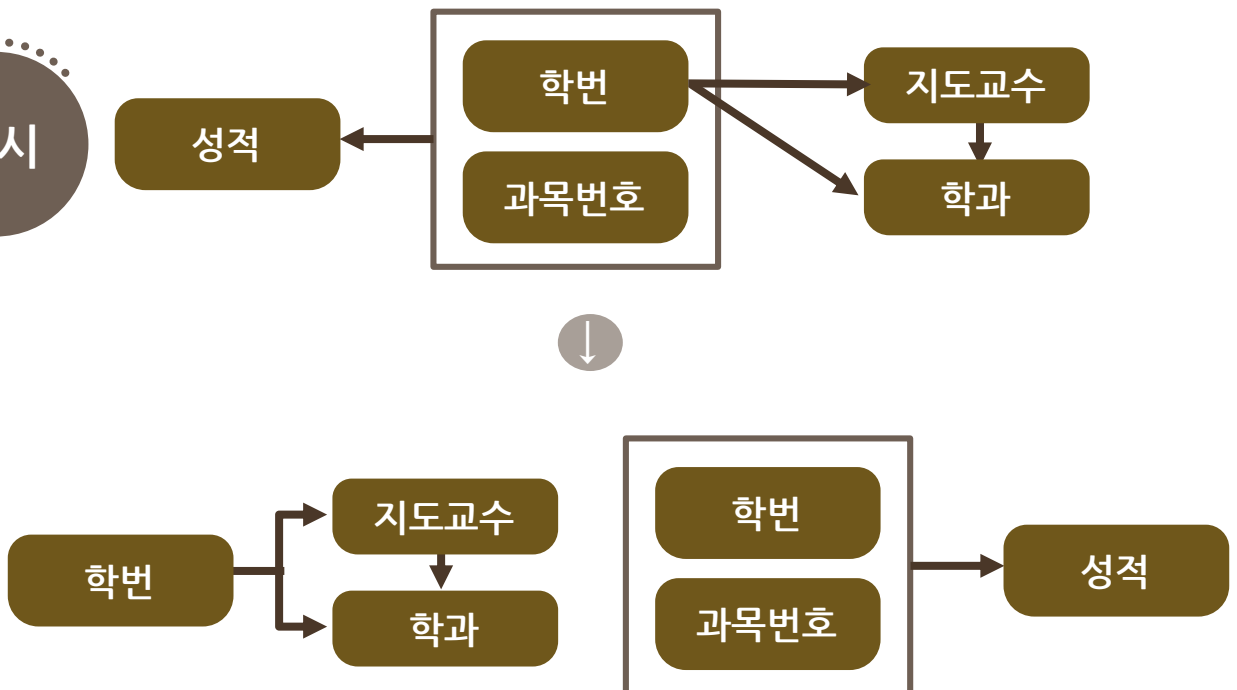
정규화 기법

2 제 2 정규형

예시

- 지도(학번, 지도교수, 학과)
 - 학번 → 지도교수
 - 학번 → 학과
 - 지도교수 → 학과
- 수강 (학번, 과목번호, 성적)
 - (학번, 과목번호) → 성적

예시



정규화 기법

제 2 정규형

지도	학번	지도교수	학과	수강	학번	과목번호	성적
	100	P1	컴퓨터		100	C413	A
	200	P2	전기		100	E412	A
	300	P3	컴퓨터		200	C123	B
	400	P1	컴퓨터		300	C312	A
					300	C324	C
					300	C413	A
					400	C312	A
					400	C324	A
					400	C413	B
					400	E412	C

삽입 이상

교수정보 삽입 시
학번 필요

삭제 이상

학생 정보삭제 시
교수 정보 연쇄 삭제

갱신 이상

P1 교수정보 변경 시
2번 변경

정규화 기법

3 제 3 정규형

정의

- 어떤 릴레이션 R이 2NF이고, 키(기본)에 속하지 않은 모든 애트리뷰트들이 기본 키에 이행적 함수 종속이 아닌 경우(이행 종속 제거)

- 제 3 정규형은 이행 함수 종속 제거

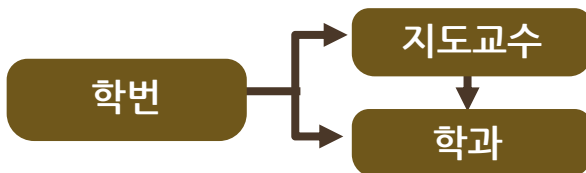


정규화 기법

3 제 3 정규형

예시

- 지도(학번, 지도교수, 학과)
 - 기본 키 : {학번}, 외래 키 : {지도교수}, 참조 : 지도교수학과
 - 학번 → 지도교수
- 지도교수학과 (지도교수, 학과)
 - 기본 키 : {지도교수}
 - 지도교수 → 학과



학생지도

학번	지도교수
100	P1
200	P2
300	P3
400	P1

지도교수 학과

지도교수	학과
P1	컴퓨터
P2	전기
P3	컴퓨터

정규화 기법

3 제 3 정규형

- 제 1 정규형부터 제 3 정규형까지는 Codd의 원본적 정의로서 모두 하나의 후보 키, 즉 하나의 기본 키만 가진 것으로 가정

릴레이션 : 3NF를 만족하면서 BCNF를 만족하지 않는 경우

- 복수의 후보 키를 가지고 있고
- 후보 키들이 두 개 이상의 애트리뷰트들로 구성되고
- 후보 키의 애트리뷰트가 서로 중첩되는 경우에는 적용할 수가 없음
- 이런 경우를 위해서 보이스/코드 정규형 (Boyce/Codd Normal Form : BCNF)이 제안됨

정규화 기법

4 BCNF(Boyce Codd Normal Form)

정의

릴레이션 R의 모든 결정자(Determinant)가 후보 키(Candidate Key)로 지정

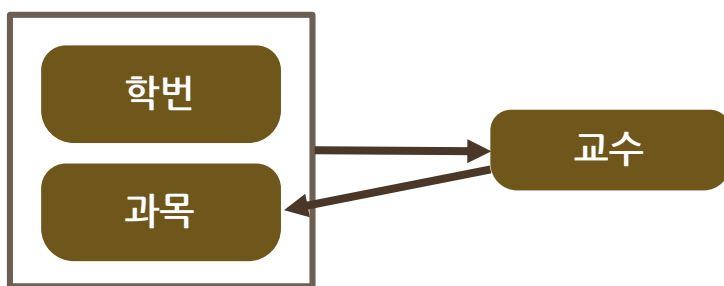
- 1NF, 2NF, 기본 키, 이행 종속 등의 개념을 이용하지 않고 직접 후보 키를 이용하여 정의될 수 있기 때문에 개념적으로 3NF보다 간단
- 제 3 정규형보다 강력하다고 볼 수 있으므로 “강한 제 3 정규형(Strong 3NF)”이라고도 함

정규화 기법

4 BCNF (Boyce Codd Normal Form)

예시

- 3NF이지만 BCNF가 아닌 경우



정규화 기법

4 BCNF(Boyce Codd Normal Form)

수강과목	학번	과목	교수
	100	프로그래밍	P1
	100	자료구조	P2
	200	프로그래밍	P1
	200	자료구조	P3
	300	자료구조	P3
	300	프로그래밍	P4

변경 이상의 원인은 사실상 애트리뷰트 교수가 결정자이지만 **후보 키로 취급하고 있지 않기 때문임**

삽입 이상

교수정보 삽입 시
학번 필요

삭제 이상

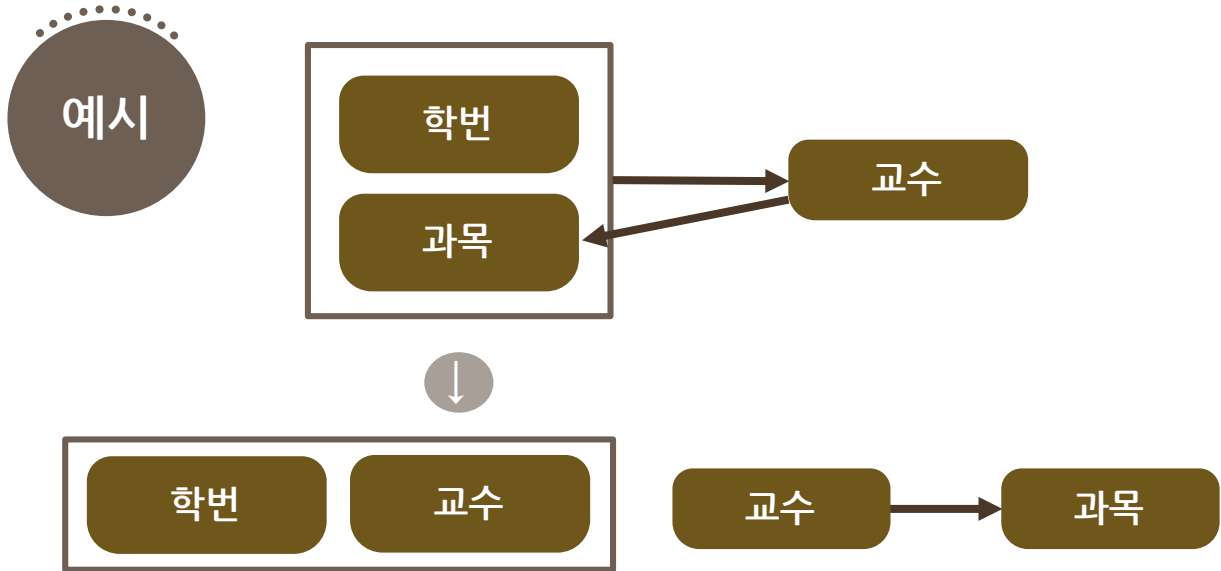
학생 정보삭제 시
교수 정보 연쇄 삭제

갱신 이상

P1 교수정보 변경 시
2번 변경

정규화 기법

4 BCNF (Boyce Codd Normal Form)



정규화 기법

5 제 4 정규형

정의

- 릴레이션 R에 MVD $A \twoheadrightarrow B$ 를 만족하는 애트리뷰트 부분집합 A, B가 존재할 때 R의 모든 애트리뷰트들이 A에 함수 종속
- R의 모든 애트리뷰트 X에 대해 $A \rightarrow X$ 이고 A가 후보 키
- 다치 종속 제거

정규화 기법

5 제 4 정규형

다치 종속(MVD : Multivalued Dependency)

- A,B,C를 릴레이션 R의 애트리뷰트의 부분집합이라 할 때
 - 애트리뷰트 쌍(A, C)값에 대응되는 B값의 집합이 A값에만 종속되고 C값에는 독립이면 B는 A에 다치 종속이라 하고 $A \twoheadrightarrow B$ 로 표기

정규화 기법

5 제 4 정규형

1 다치 종속 예시

개설과목	과목	교수	교재	BCNF
	자료구조	P1	T1	
	자료구조	P1	T2	
	자료구조	P2	T1	
	자료구조	P2	T2	
	데이터베이스	P1	T1	
	데이터베이스	P1	T3	
	데이터베이스	P1	T4	

(∵키는 모든 애트리뷰트를 포함한 {과목, 교수, 교재}가
되어 이외에는 어떤 결정자도 없음)

기본 키 : (과목, 교수, 교재)

- 과목»교수에서 과목의 자료구조는 교수 애트리뷰트 값의 집합 {P1,P2}와 대응하고,
- 과목»교재에서 과목의 자료구조는 교재 애트리뷰트 값의 집합 {T1,T2}와 대응한다는 의미

정규화 기법

5 제 4 정규형

1 다치 종속 예시

개설과목	과목	교수	교재
	자료구조	P1	T1
	자료구조	P1	T2
	자료구조	P2	T1
	자료구조	P2	T2
	데이터베이스	P1	T1
	데이터베이스	P1	T3
	데이터베이스	P1	T4

이상 현상은 교수와 교재가 무관한데 **하나의 릴레이션에 표현**했기 때문임

삽입 이상

데이터베이스 과목을 P4
교수가 담당한다면
교재수만큼 입력

갱신 이상

교수 정보 변경 시 교재
수만큼 변경

정규화 기법

5 제 4 정규형

2 제 4 정규형 예시

개설과목	과목	교수	교재
	자료구조	P1	T1
	자료구조	P1	T2
	자료구조	P2	T1
	자료구조	P2	T2
	데이터베이스	P1	T1
	데이터베이스	P1	T3
	데이터베이스	P1	T4



4NF

과목 교수	과목	교수	과목 교재	과목	교재
	자료구조	P1		자료구조	T1
	자료구조	P2		자료구조	T2
	데이터베이스	P1		데이터베이스	T1
				데이터베이스	T3
				데이터베이스	T4

정규화 기법

6 제 5 정규형

정의

- 릴레이션 R에 존재하는 모든 조인 종속이 릴레이션 R의 후보 키를 통해서만 성립, PJ/NF(Projection - Join Normal Form), (조인 종속 제거)

정규화 기법

6 제 5 정규형

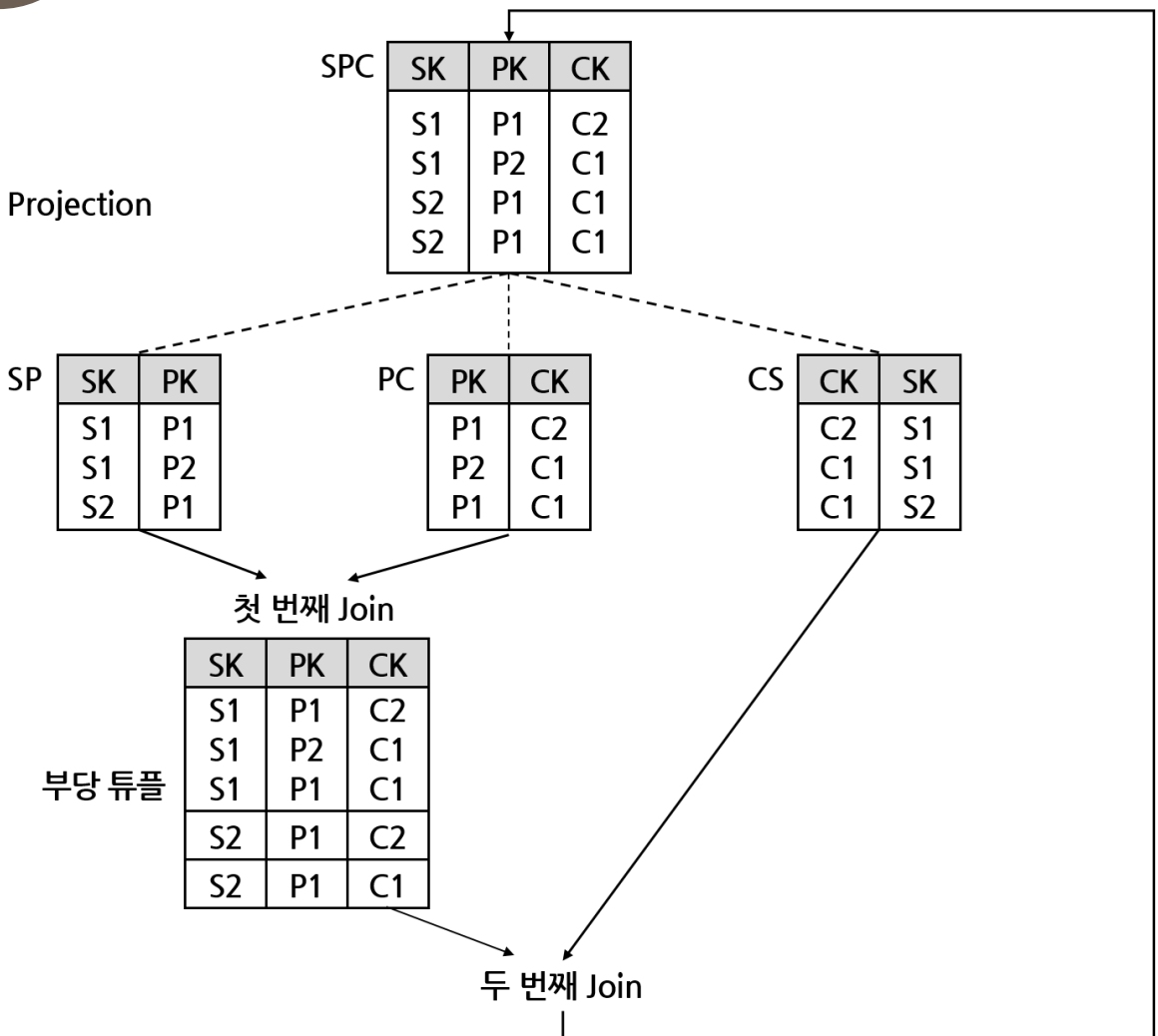
조인 종속(JD : Join Dependency)

- 어떤 릴레이션 R의 애트리뷰트 부분 집합 A, B, \dots, Z 가 있는 경우
 - 이때 만일 릴레이션 R이 그의 프로젝션 A, B, \dots, Z 를 모두 조인한 것과 똑같다면 R은 조인 종속* (A, B, \dots, Z) 을 만족시킴

정규화 기법

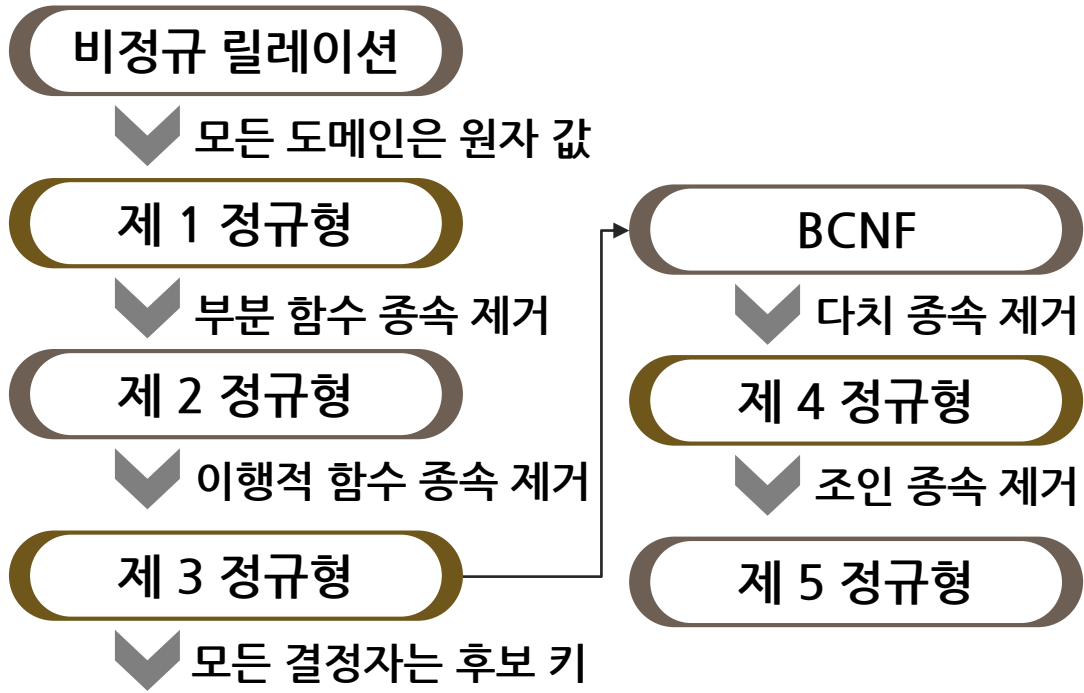
6 제 5 정규형

예시



정규화 기법

7 정규화 순서



학습정리

1. 정규화의 필요성

- 정규화 모델의 특징
 - 개체 관계의 정확성, 데이터의 일치성, 데이터 모델의 단순성, 개체에 존재하는 속성의 비중복성
- 관계형 DB 설계의 가이드라인
 - 서로 다른 개체 타입(Entity Type)들을 하나로 혼합하지 말 것, 중복이나 갱신 부작용이 발생하지 않게 할 것, 널 값(Null Value) 발생을 될 수 있는 한 피할 것, 위조 튜플(Spurious Tuple)들의 발생을 피할 것

학습정리

2. 정규화 기법

- 제 1 정규형
 - 어떤 릴레이션 R에 속한 모든 도메인이 원자 값(Atomic Value) 만으로 구성
- 제 2 정규형
 - 어떤 릴레이션 R이 1NF이고, 키(기본)에 속하지 않은 애트리뷰트 모두가 기본 키에 완전 함수 종속 (부분 함수 종속 제거)
- 제 3 정규형
 - 어떤 릴레이션 R이 2NF이고, 키(기본)에 속하지 않은 모든 애트리뷰트들이 기본 키에 이행적 함수 종속이 아닌 경우 (이행 종속 제거)
- BCNF
 - 릴레이션 R의 모든 결정자(Determinant)가 후보 키(Candidate Key)로 지정
- 제 4 정규형
 - 릴레이션 R에 MVD $A \twoheadrightarrow B$ 를 만족하는 애트리뷰트 부분집합 A,B가 존재할 때 R의 모든 애트리뷰트들이 이 A에 함수 종속(즉, R의 모든 애트리뷰트 X에 대해 $A \rightarrow X$ 이고, A가 후보 키)(다치 종속 제거)

학습정리

2. 정규화 기법

- 다치 종속(MVD : Multivalued Dependency)
 - A,B,C를 릴레이션 R의 애트리뷰트의 부분집합이라 할 때 애트리뷰트 쌍 (A, C)값에 대응되는 B값의 집합이 A값에만 종속되고, C값에는 독립이면 B는 A에 다치 종속이라 하고 $A \twoheadrightarrow B$ 로 표기
- 제 5 정규형
 - 릴레이션 R에 존재하는 모든 조인 종속이 릴레이션 R의 후보 키를 통해서만 성립,
PJ/NF (Projection - Join Normal Form), (조인 종속 제거)
- 조인 종속(JD : Join Dependency)
 - 어떤 릴레이션 R의 애트리뷰트 부분 집합 A,B,...,Z가 있다고 하자. 이때 만일 릴레이션 R이 그의 프로젝션 A,B,...,Z를 모두 조인한 것과 똑같다면 R은 조인 종속*(A,B,...,Z)을 만족시킴