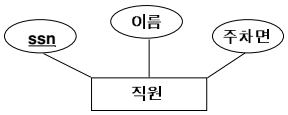
개체-관계 (Entity-Relationship)모델

Chapter 2

데이타베이스 설계 개관

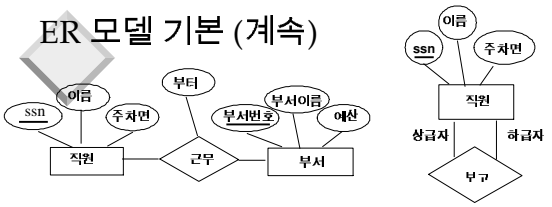
- 개념적 설계: (ER 모델이 이 단계에서 필요함.)
  - 해당 조직체에서 개체(entity)는 무엇이고 관계성 (relationship)은 무엇인가?
  - 개체와 관계에 대한 정보로는 무엇을 DB에 저장해야 할까?
  - 무결성 제약조건이나 사업 규칙으로는 어떤 것들이 있을까?
  - ER모델에서는 데이타베이스 '스키마'를 도표로 나타낼 수 있다. (ER 다이어그램)
  - ER 다이어그램을 관계 스키마로 변환할 수 있다.

ER 모델 기본



- 개체(Entity): 다른 객체와 구별되는 실세계의 객체. 개체는(DB내에서) 애트리뷰트(attribute) 집단을 통해 기술된다.
- 개체집합(Entity Set): 비슷한 개체들의 모임. E.g., 모든 직원.
  - 같은 개체집합에 속하는 개체들은 같은 애트리뷰트 집단을 가진다. (ISA계층에서는 예외)
  - 모든 개체집합마다 키(key)가 존재.
  - 애트리뷰트에는 도메인(domain)이 존재.

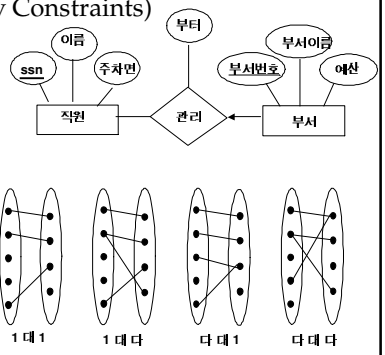
ER 모델 기본 (계속)



- 관계(Relationship):둘 이상의 개체간의 관련성. E.g., Attishoo 가 제약 부서에서 근무한다.
- 관계집합(Relationship Set): 비슷한 관계들의 모임.
  - n-진(n-ary) 관계 집합 R은 n개의 개체집합 E1 ... En간을 관련짓는다; R에 속하는 관계 하나는 개체 e1∈ E1, ..., en ∈ En에 관한 것이다.
  - 동일한 개체 집합이 여러 관계집합에 참여할 수 있으며, 동일한 관계 집합에서 다른 "역할"들을 할 수도 있다.

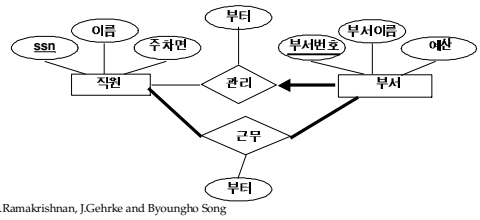
키 제약조건(Key Constraints)

- "근무"를 보자: 한 직원이 여러 부서에 근무할 수 있고, 한 부서에 여러 직원이 근무할 수 있다.
- 반대로 "관리"에는 키 제약조건이 있으므로 한 부서에 많아야 한 명의 부서장이 있다.



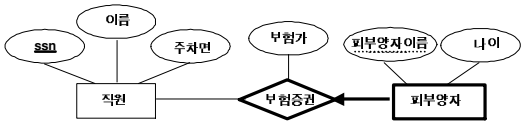
참여 제약조건(Participation Constraints)

- 모든 부서에 부서장이 있는가?
  - 그렇다면 참여 제약조건이 된다: "관리"에 대한 "부서"의 참여도를 전체적이라고 한다. (부분적도 있다).
    - 부서 테이블의 어떤 "부서번호"값도 빠 짐없이 관리 테이블에 한 행으로 나타나야 한다.(이때 ssn은 널이 될 수 없다)



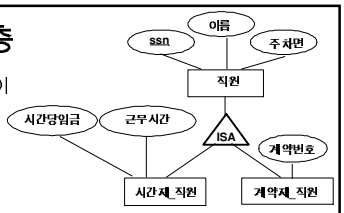
약 개체(Weak Entity)

- 약 개체는 다른(소유자) 개체의 기본 키까지 보아야 유일하게 식별할 수 있다.
  - 소유자 개체집합과 약 개체집합은 1대다 관계집합으로 연결된다. (소유자 하나에 여러 개의 약 개체)
  - 약 개체집합은 이 식별 (identifying) 관계집합에 전체적으로 참여한다.



ISA (‘is a’)계층

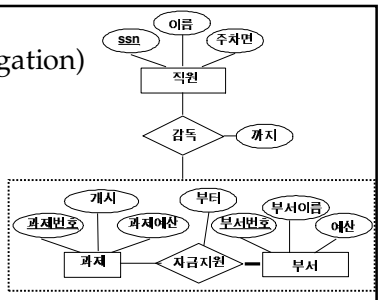
- C++처럼 애트리뷰트들이 계승된다.
- A ISA B 라고 선언하면, A개체는 모두 B개체로도 볼 수 있다.



- 중첩(overlap)제약조건: Joe가 시간제_직원이면서 계약제_직원도 될 수 있는가? (Allowed/disallowed)
- 포괄(covering)제약조건: 직원들은 모두 시간제_직원 아니면 계약제_직원인가? (예/아니오)
- ISA를 사용하는 이유:
  - 특정 세부부류에 설명용 애트리뷰트를 추가하고 싶어서.
  - 어떤 관계에 참여하는 개체들을 구별해 주기위해서.

집단화(Aggregation)

- (개체집합과) 관계집합이 참여하는 관계를 모델링 할 때 사용.
  - 집단화를 사용하면 관계집합을 일종의 개체집합으로 취급하여 또 다른 관계에 참여 시킬 수 있다.



- 집단화와 3진 관계:

* “감독”은 설명용 애트리뷰트가 있는, 독립된 관계.
* 각 “자금지원”건은 많아야 한 부서에서만 지원 가능하다는 것도 표현할 수 있다.

ER 모델을 이용한 개념적 설계

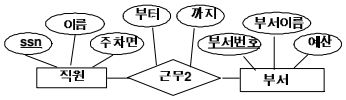
- 설계상의 선택들:
  - 이 개념을 개체로 모델링할 것인가, 애트리뷰트로 할 것인가?
  - 이 개념을 개체로 모델링할 것인가, 관계로 할 것인가?
  - 관계의 형태: 2진인가 3진인가? 집단화?
- ER모델의 제약조건 표현력:
  - 데이타의 의미들을 많이 표현할 수 있다 (또 해야 한다.)
  - 그러나 ER 다이어그램으로 표현할 수 없는 제약 조건들도 있다.

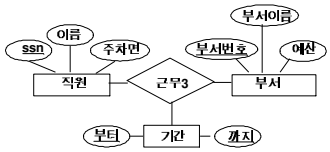
개체냐 애트리뷰트냐

- 주소를 "직원"의 애트리뷰트로 할까, 독립된 개체 (관계를 통해 직원과 연결)로 할까?
- 주소 정보의 용법과, 데이타의 의미에 따라 결정:
  - 한 직원이 여러 주소가 가능하면, "주소"는 개체로 (애트리뷰트에 집합값은 안되므로).
  - 주소의 구조(시, 구, 동)가 중요한 경우, "주소"는 개체로 (애트리뷰트는 원자값이므로).

개체냐 애트리뷰트냐(계속)

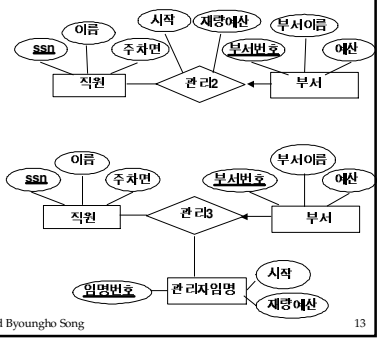
- "근무2에 따르면, 한 직원이 어떤 부서에 근무한 기간이 여럿 있을 수 없다.
- 한 직원에 여러 주소를 기록하고 싶을 때와 비슷한 문제: 이 관계에 속한 인스턴스에 여러 설명용 애트리뷰트 값들을 기록하고 싶다.





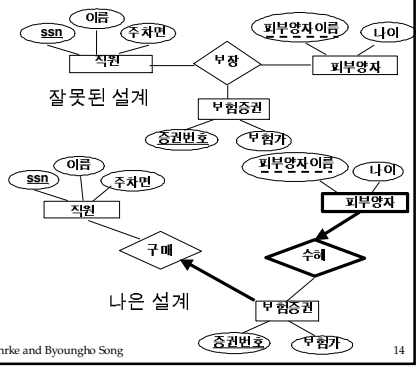
개체냐 관계냐

- 관리자가 부서마다 별도로 재량예산을 받는다면 위의 다이어그램은 OK.
- 관리자가 총액기준으로 재량예산을 받는다면?
  - 각 부서마다 같은 재량예산액을 중복 저장.
    잘못된 방향: 재량예산액을 부서에 붙여놓게 됨.



2진관계냐 3진관계냐

- 한 증권은 한 직원만 가질 수 있다면 :
  - "보험증권"에 키제약조건을 붙이면 증권이 피부양자 한명만 보장하게 됨!
- 두 번째 다이어그램에 추가된 제약조건은?



2진관계냐 3진관계냐 (계속)

- 앞의 예는 2진 관계 2개가 3진 관계 하나보다 나을 수 있음을 보여준다.
- "부품", "공급자", "부서"의 세 개체집합을 연관짓는 3진 관계 "계약"이 있고 설명용 애트리뷰트 "수량"이 있다. 어떤 2진 관계들을 조합해도 이를 대체할 수 없다.
  - S가 P를 공급 가능, D가 P를 필요, D가 S와 거래⇒D가 S로부터 P를 사기로 했다가 나오지 않음.
  - "수량"은 어디에 기록?

개념적 설계 요약

- 요구조건 분석 후에 개념적 설계 수행,
  - 저장할 데이타를 고수준에서 기술
- 개념적 설계에 ER 모델이 대중적.
  - 표현력이 풍부하고, 사람의 관점과 비슷
- 기본적인 구성요소: 개체, 관계, (개체나 관계의 애트리뷰트)
- 기타 구성요소: 약 개체, ISA 계층, 집단화 등.
- 주의: ER 모델에는 여러 변형이 있음.

ER 요약 (계속)

- ER 모델에서 표현할 수 있는 제약조건: 키 제약조건, 참여 제약조건, ISA 계층에 대한 중첩/포괄 제약조건. 관계집합 정의에는 일정한 외래 키 제약조건도 암시된다.
  - ER 모델로 표현할 수 없는 제약조건도 있다. (특히 함수 종속성)
  - 제약조건들은 최선의 데이타베이스 설계에 중요한 역할을 담당한다.

ER 요약 (계속)

- ER 설계는 주관적이다. 주어진 시나리오를 여러 가지로 모델링 할 수 있다! 대안들을 분석할 때에는 재주가 필요하다.

  일반적으로 다음 사항들을 선택하여야 한다.
  - 개체냐 애트리뷰트냐, 개체냐 관계냐, 2진이냐 3진이냐, ISA계층을 할까, 집단화를 할까.
- 좋은 DB 설계를 위해서: 결과된 관계 스키마를 더 분석해서 정제한다. 함수종속(FD)과 정규화 기술이 특히 유용하다.