Ch2. 관계 데이터 모델

숭실대학교 스마트시스템소프트웨어 학과
Department of Smart Systems Software

목차

1. 관계 데이터 모델의 개념
2. 무결성 제약조건
3. 관계대수

01. 관계 데이터 모델의 개념

- 릴레이션
- 릴레이션 스키마와 인스턴스
- 릴레이션의 특징
- 관계 데이터 모델

1.1 릴레이션

■ **릴레이션(relation) : 행과 열로 구성된 테이블**

표 2-1 릴레이션과 관련된 한글 용어

용어	한글 용어	비고
relation	릴레이션, 테이블	"관계"라고 하지 않음
relational data model	관계 데이터 모델	
relational database	관계 데이터베이스	
relational algebra	관계대수	
relationship	관계	

1.1 릴레이션

■ 릴레이션이란?

도서 1, 축구의 역사, 굿스포츠, 7000
도서 2, 축구아는 여자, 나무수, 13000
도서 3, 축구의 이해, 대한미디어, 22000
도서 4, 골프 바이블, 대한미디어, 35000
도서 5, 피겨 교본, 굿스포츠, 8000

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

그림 2-1 데이터와 테이블(릴레이션)

도서번호 = {1,2,3,4,5}
도서이름 = {축구의 역사, 축구아는 여자, 축구의 이해, 골프 바이블, 피겨 교본}
출판사 = {굿스포츠, 나무수, 대한미디어}
가격 = {7000, 13000, 22000, 35000, 8000}

→ 첫 번째 행(1, 축구의 역사, 굿스포츠, 7000)의 경우 네 개의 집합에서 각각 원소 한 개씩 선택하여 만들어진 것으로 이 원소들이 관계(relationship)를 맺고 있다.

1.1 릴레이션

관계(relationship)

1. 릴레이션 내에서 생성되는 관계 : 릴레이션 내 데이터들의 관계
2. 릴레이션 간에 생성되는 관계 : 릴레이션 간의 관계

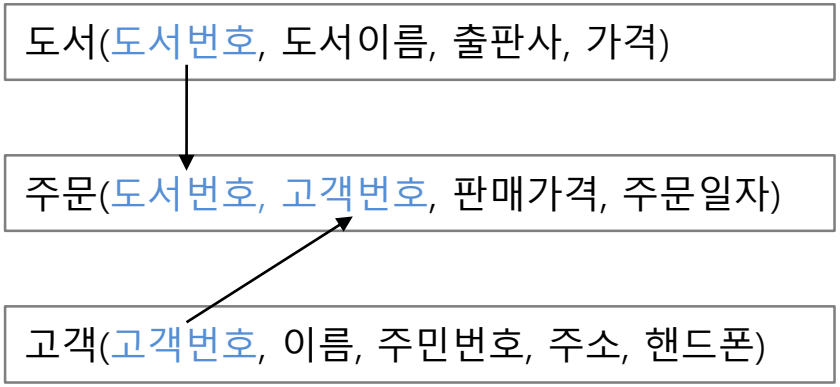

그림 2-2 릴레이션 간의 관계

1.2 릴레이션 스키마와 인스턴스

속성(애트리뷰트),
열(column) 이라고도 함
(차수=4)

도서

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

스키마(내포)
Schema

인스턴스(외연)
Instance

투플(tuple),
행(row) 이라고도 함
(카디널리티=5)

그림 2-3 도서 릴레이션

1.2.1 릴레이션 스키마

스키마의 요소

- 속성(attribute) : 릴레이션 스키마의 열
- 도메인(domain) : 속성이 가질 수 있는 값의 집합
- 차수(degree) : 속성의 개수

스키마의 표현

- 릴레이션 이름(속성1 : 도메인1, 속성2 : 도메인2, 속성3 : 도메인3 ...)

  EX) 도서(도서번호, 도서이름, 출판사, 가격)

1.2.2 릴레이션 인스턴스

인스턴스 요소

- 투플(tuple) : 릴레이션의 행 → 투플이 가지는 속성의 개수는 릴레이션 스키마의 차수와 동일하고, 릴레이션 내의 모든 투플들은 서로 중복되지 않아야 함
- 카디날리티(cardinality) : 투플의 수

표 2-2 릴레이션 구조와 관련된 용어

릴레이션 용어	같은 의미로 통용되는 용어	파일 시스템 용어
릴레이션(relation)	테이블(table)	파일(file)
스키마(schema)	내포(intension)	헤더(header)
인스턴스(instance)	외연(extension)	데이터(data)
투플(tuple)	행(row)	레코드(record)
속성(attribute)	열(column)	필드(field)

1.3 릴레이션의 특징

- **속성은 단일 값을 가진다**

  각 속성의 값은 도메인에 정의된 값만을 가지며 그 값은 모두 단일 값이여야 함.

- **속성은 서로 다른 이름을 가진다**

  속성은 한 릴레이션에서 서로 다른 이름을 가져야만 함.

- **한 속성의 값은 모두 같은 도메인 값을 가진다**

  한 속성에 속한 열은 모두 그 속성에서 정의한 도메인 값만 가질 수 있음.

- **속성의 순서는 상관없다**

  속성의 순서가 달라도 릴레이션 스키마는 같음.

  예) 릴레이션 스키마에서 (이름, 주소) 순으로 속성을 표시하거나 (주소, 이름) 순으로 표시하여도 상관없음.

- **릴레이션 내의 중복된 투플은 허용하지 않는다**

  하나의 릴레이션 인스턴스 내에서는 서로 중복된 값을 가질 수 없음. 즉 모든 투플은 서로 값이 달라야 함.

- **투플의 순서는 상관없다**

  투플의 순서가 달라도 같은 릴레이션임. 관계 데이터 모델의 투플은 실제적인 값을 가지고 있으며 이 값은 시간이 지남에 따라 데이터의 삭제, 수정, 삽입에 따라 순서가 바뀔 수 있음.

1.3 릴레이션의 특징

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구 아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000
5	피겨 교본	굿스포츠	8000
6	피겨 교본, 피겨 기초	굿스포츠	8000

동일한 투플이 중복되면 안 됨

속성의 값은 단일 값이어야 함

그림 2-4 릴레이션의 특징에 위배된 경우

1.4 관계 데이터 모델

- **관계 데이터 모델은 데이터를 2차원 테이블 형태인 릴레이션으로 표현함. 릴레이션에 대한 제약조건(constraints)과 관계 연산을 위한 관계대수(relational algebra)를 정의함.**

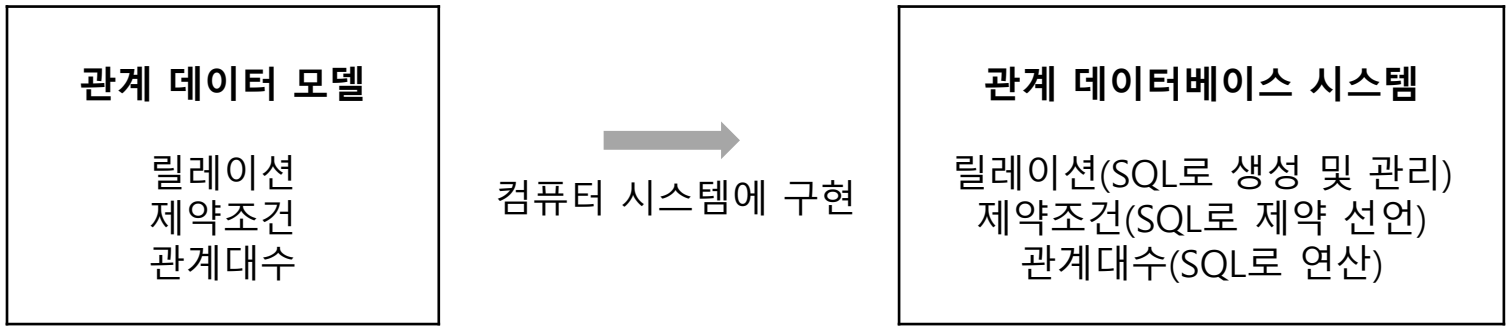

그림 2-5 관계 데이터베이스 시스템

02. 무결성 제약조건

- 키
- 무결성 제약조건
- 무결성 제약조건의 수행

2.1 키

- 특정 투플을 식별할 때 사용하는 속성 혹은 속성의 집합임.
- 릴레이션은 중복된 투플을 허용하지 않기 때문에 각각의 투플에 포함된 속성들 중 어느 하나(혹은 하나 이상)는 값이 달라야 함. 즉 키가 되는 속성(혹은 속성의 집합)은 반드시 값이 달라서 투플들을 서로 구별할 수 있어야 함.
- 키는 릴레이션 간의 관계를 맺는 데도 사용됨.

그림 2-6 자동차 1 대당 키는 단 하나

2.1 키

고객

고객번호	이름	주민번호	주소	핸드폰
1	박지성	810101-1111111	영국 맨체스타	000-5000-0001
2	김연아	900101-2222222	대한민국 서울	000-6000-0001
3	장미란	830101-2333333	대한민국 강원도	000-7000-0001
4	추신수	820101-1444444	미국 클리블랜드	000-8000-0001

도서

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

주문

고객번호	도서번호	판매가격	주문일자
1	1	7000	2014-07-01
1	2	13000	2014-07-03
2	5	8000	2014-07-03
3	2	13000	2014-07-04
4	4	35000	2014-07-05
1	3	22000	2014-07-07
4	3	22000	2014-07-07

그림 2-7 마당서점 데이터베이스

2.1.1 슈퍼키

- **투플을 유일하게 식별할 수 있는 하나의 속성 혹은 속성의 집합**

  투플을 유일하게 식별할 수 있는 값이면 모두 슈퍼키가 될 수 있음

- **(고객 릴레이션 예)**
  - 고객번호 : 고객별로 유일한 값이 부여되어 있기 때문에 투플을 식별할 수 있음.
  - 이름 : 동명이인이 있을 경우 투플을 유일하게 식별할 수 없음.
  - 주민번호 : 개인별로 유일한 값이 부여되어 있기 때문에 투플을 식별할 수 있음.
  - 주소 : 가족끼리는 같은 정보를 사용하므로 투플을 식별할 수 없음.
  - 핸드폰 : 한 사람이 여러 개의 핸드폰을 사용할 수 있고 반대로 핸드폰을 사용하지 않는 사람이 있을 수 있기 때문에 투플을 식별할 수 없음.

- **고객 릴레이션은 고객번호와 주민번호를 포함한 모든 속성의 집합이 슈퍼키가 됨.**

  EX) (주민번호), (주민번호, 이름), (주민번호, 이름, 주소), (주민번호, 이름, 핸드폰),
  (고객번호), (고객번호, 이름, 주소), (고객번호, 이름, 주민번호, 주소, 핸드폰) 등

2.1.2 후보키

- **투플을 유일하게 식별할 수 있는 속성의 최소 집합**
  (주문 릴레이션 예)
  - 고객번호 : 한 명의 고객이 여러 권의 도서를 구입할 수 있으므로 후보키가 될 수 없음. 고객번호가 1인 박지성 고객은 세 번의 주문 기록이 있으므로 투플을 유일하게 식별할 수 없음.
  - 도서번호 : 도서번호가 2인 '축구아는 여자'는 두 번의 주문 기록이 있으므로 투플을 유일하게 식별할 수 없음.

- **주문 릴레이션의 후보키는 2개의 속성을 합한 (고객번호, 도서번호)가 됨. 참고로 이렇게 2개 이상의 속성으로 이루어진 키를 복합키(composite key)라고 함.**

2.1.3 기본키

- **여러 후보키 중 하나를 선정하여 대표로 삼는 키**
- **후보키가 하나뿐이라면 그 후보키를 기본키로 사용하면 되고 여러 개라면 릴레이션의 특성을 반영하여 하나를 선택하면 됨.**

- **기본키 선정 시 고려사항**
  - 릴레이션 내 투플을 식별할 수 있는 고유한 값을 가져야 함.
  - NULL 값은 허용하지 않음.
  - 키 값의 변동이 일어나지 않아야 함.
  - 최대한 적은 수의 속성을 가진 것이라야 함.
  - 향후 키를 사용하는 데 있어서 문제 발생 소지가 없어야 함.

- **릴레이션 스키마를 표현할 때 기본키는 밑줄을 그어 표시함**

  릴레이션 이름(<u>속성1</u>, 속성2, ... 속성N)

  EX) 고객(<u>고객번호</u>, 이름, 주민번호, 주소, 핸드폰)

  도서(<u>도서번호</u>, 도서이름, 출판사, 가격)

2.1.6 외래키

- **다른 릴레이션의 기본키를 참조하는 속성을 말함. 다른 릴레이션의 기본키를 참조하여 관계 데이터 모델의 특징인 릴레이션 간의 관계(relationship)를 표현함.**

- **외래키의 특징**
  - 관계 데이터 모델의 릴레이션 간의 관계를 표현함.
  - 다른 릴레이션의 기본키를 참조하는 속성임.
  - 참조하고(외래키) 참조되는(기본키) 양쪽 릴레이션의 도메인은 서로 같아야 함.
  - 참조되는(기본키) 값이 변경되면 참조하는(외래키) 값도 변경됨.
  - NULL 값과 중복 값 등이 허용됨.
  - 자기 자신의 기본키를 참조하는 외래키도 가능함.
  - 외래키가 기본키의 일부가 될 수 있음.

2.1.6 외래키

고객

고객번호	이름	주민번호	주소	핸드폰
1	박지성	810101-1111111	영국 맨체스타	000-5000-0001
2	김연아	900101-2222222	대한민국 서울	000-6000-0001
3	장미란	830101-2333333	대한민국 강원도	000-7000-0001
4	추신수	820101-1444444	미국 클리블랜드	000-8000-0001

기본키

도서

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

기본키

참조

외래키

참조

주문

주문번호	고객번호	도서번호	판매가격	주문일자
1	1	1	7000	2014-07-01
2	1	2	13000	2014-07-03
3	2	5	8000	2014-07-03
4	3	2	13000	2014-07-04
5	4	4	35000	2014-07-05
6	1	3	22000	2014-07-07
7	4	3	22000	2014-07-07

기본키

그림 2-9 릴레이션 간의 참조 관계

2.1.6 외래키

- **외래키 사용 시 참조하는 릴레이션과 참조되는 릴레이션이 꼭 다른 릴레이션일 필요는 없음. 즉 자기 자신의 기본키를 참조할 수도 있음.**

참조

기본키 외래키

선수번호	이름	주소	멘토번호
1	박지성	영국 맨체스타	NULL
2	김연아	대한민국 서울	3
3	장미란	대한민국 강원도	4
4	추신수	미국 클리블랜드	NULL

그림 2-10 멘토 릴레이션

03. 관계대수

- 관계대수
- 셀렉션과 프로젝션
- 집합연산
- 조인
- 디비전
- 관계대수 예제

3.1 관계대수

관계대수(relational algebra, 關係代數)

릴레이션에서 원하는 결과를 얻기 위해 수학의 대수와 같은 연산을 이용하여 질의하는 방법을 기술하는 언어

관계대수와 관계해석

- **관계대수** : 어떤 데이터를 어떻게 찾는지에 대한 처리 절차를 명시하는 절차적인 언어이며, DBMS 내부의 처리 언어로 사용됨
- **관계해석** : 어떤 데이터를 찾는지만 명시하는 선언적인 언어로 관계대수와 함께 관계 DBMS의 표준 언어인 SQL의 이론적인 기반을 제공함

→ **관계대수와 관계해석은 모두 관계 데이터 모델의 중요한 언어이며 실제 동일한 표현 능력을 가지고 있음.**

3.1.2 관계대수 연산자

표 2-5 관계대수 연산자

연산자 종류	대상	연산자 이름			기호	설명
기본	단항	**셀렉션**			σ	릴레이션에서 조건에 만족하는 투플을 선택
기본	단항	**프로젝션**			π	릴레이션의 속성을 선택
추가	단항	개명			ρ	릴레이션이나 속성의 이름을 변경
유도	이항	디비전			$\div$	부모 릴레이션에 포함된 투플의 값을 모두 갖고 있는 투플을 분자 릴레이션에서 추출
기본	이항	**합집합**			$\cup$	두 릴레이션의 합집합
기본	이항	**차집합**			$-$	두 릴레이션의 차집합
유도	이항	교집합			$\cap$	두 릴레이션의 교집합
기본	이항	**카디전 프로덕트**			$\times$	두 릴레이션에 속한 모든 투플의 집합
유도	이항	조인	세타		$\bowtie_{\theta}$	두 릴레이션 간의 비교 조건에 만족하는 집합
			동등		$\bowtie$	두 릴레이션 간의 같은 값을 가진 집합
			자연		$\bowtie_N$	동등 조인에서 중복 속성을 제거
			세미	left	$\ltimes$	자연 조인 후 오른쪽 속성을 제거
				right	$\rtimes$	자연 조인 후 왼쪽 속성을 제거
			외부	left	$⟕$	• 자연 조인 후 각각 왼쪽(left), 오른쪽(right), 양쪽(full)의 모든 값을 결과로 추출 • 조인이 실패(또는 값이 없을 경우)한 쪽의 값을 NULL로 채움
				right	$⟖$	
				full	$⟗$	

3.1.3 관계대수식

■ 관계대수식

관계대수는 릴레이션 간 연산을 통해 결과 릴레이션을 찾는 절차를 기술한 언어로, 이 연산을 수행하기 위한 식을 관계대수식(relational algebra expression)이라고 함. 관계대수식은 대상이 되는 릴레이션과 연산자로 구성되며, 결과는 릴레이션으로 반환됨. 반환된 릴레이션은 릴레이션의 모든 특징을 따름.

- **단항 연산자 : 연산자$_{<조건>}$ 릴레이션**
- **이항 연산자 : 릴레이션1 연산자$_{<조건>}$ 릴레이션2**

R1

A	B	C
a1	b1	c1
a2	b3	c3
a3	b4	c2

R2

A	B	C
a1	b1	c1
a2	b3	c3
a3	b3	c1

그림 2-17 관계대수식을 이해하기 위한 예제 데이터

R1

A	B	C
a1	b1	c1
a2	b3	c3
a3	b4	c2

R2

A	B	C
a1	b1	c1
a2	b3	c3
a3	b3	c1

표 2-6 관계대수 식의 사용 예

주요 연산자	사용 예	결과	설명
셀렉션(σ)	$\sigma_{A=a1 \text{ or } A=a2}$ (R1)	A B C a1 b1 c1 a2 b3 c3	R1에서 조건에 맞는 투플을 추출함.
프로젝션(π)	$\pi_{A,B}$(R2)	A B a1 b1 a2 b3 a3 b3	R2에서 조건에 맞는 속성만을 추출함.
합집합(∪)	R1 ∪ R2	A B C a1 b1 c1 a2 b3 c3 a3 b4 c2 a3 b3 c1	R1과 R2의 합집합을 구함.
차집합(-)	R1 - R2	A B C a3 b4 c2	R1과 R2의 차집합을 구함.
조인	R1 $\bowtie_{R1.C=R2.C}$ R2	R1.A R1.B R1.C R2.A R2.B R2.C a1 b1 c1 a1 b1 c1 a1 b1 c1 a3 b3 c1 a2 b3 c3 a2 b3 c3	R1과 R2의 카티전 프로덕트를 구하여 조건에 맞는 투플을 추출함.

3.2.1 셀렉션(selection)

- **릴레이션의 투플을 추출하기 위한 연산임. 하나의 릴레이션을 대상으로 하는 단항 연산자며, 찾고자 하는 투플의 조건(predicate)을 명시하고 그 조건에 만족하는 투플을 반환함.**
- **형식 :** $\sigma_{<조건>}(R)$ (R은 릴레이션, σ는 그리스 문자이며 대문자는 Σ)

질의 2-1 마당서점에서 판매하는 도서 중 8,000원 이하인 도서를 검색하시오.

$\sigma_{가격<=8000}$ (도서)

도서

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

$\sigma_{가격<=8000}$ (도서)

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
5	피겨 교본	굿스포츠	8000

그림 2-18 셀렉션의 예

3.2.1 셀렉션(selection)의 확장

- **형식 :** $\sigma_{<\text{복합조건}>}$ (R) (R은 릴레이션, σ는 그리스 문자이며 대문자는 Σ)

- 여러 개의 조건을 ∧(and), ∨ (or), ¬ (not) 기호를 이용하여 복합조건을 표시할 수 있다. 예를 들어, "가격이 8,000원 이하이고, 도서번호가 3 이상인 책을 찾아라"는 질의는 다음과 같이 표현한다.

  $\sigma_{(\text{가격}<=8000 \wedge \text{도서번호} >=3)}$ (도서)

3.2.2 프로젝션(projection)

- **릴레이션의 속성을 추출하기 위한 연산으로 단항 연산자임.**
- **형식 :** $\pi_{<속성리스트>}(R)$ (R은 릴레이션, π 는 그리스 문자이며 대문자는 Π)

질의 2-2 신간도서 안내를 위해 고객의 (이름, 주소, 핸드폰)이 적힌 카탈로그 주소록을 만드시오.

$\pi_{이름,\ 주소,\ 핸드폰}(고객)$

고객

고객번호	이름	주민번호	주소	핸드폰
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001
2	김연아	900905-2222222	대한민국 서울	000-6000-0001
3	장미란	831009-2333333	대한민국 강원도	000-7000-0001
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001

$\pi_{이름,\ 주소,\ 핸드폰}(고객)$

이름	주소	핸드폰
박지성	영국 맨체스타	000-5000-0001
김연아	대한민국 서울	000-6000-0001
장미란	대한민국 강원도	000-7000-0001
추신수	미국 클리블랜드	000-8000-0001

그림 2-19 프로젝션의 예

3.3.1 합집합

- 두 개의 릴레이션을 합하여 하나의 릴레이션을 반환함. 이 때 두 개의 릴레이션은 서로 같은 속성 순서와 도메인을 가져야 함.
- **형식 : R ∪ S**

질의 2-3 마당서점은 지점A와 지점B가 있다. 두 지점의 도서는 각 지점에서 관리하며 릴레이션 이름은 각각 도서A, 도서B다. 마당서점의 도서를 하나의 릴레이션으로 보이시오.

도서A ∪ 도서B

도서A

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000

∪

도서B

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

그림 2-20 합집합의 예

3.3.2 교집합

- 합병가능한 두 릴레이션을 대상으로 하며, 두 릴레이션이 공통으로 가지고 있는 투플을 반환함.
- 형식 : R ∩ S

질의 2-4 마당서점의 두 지점에서 동일하게 보유하고 있는 도서 목록을 보이시오.

도서A ∩ 도서B

도서A

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000

∩

도서B

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

→

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000

그림 2-21 교집합의 예

3.3.3 차집합

- 첫 번째 릴레이션에는 속하고 두 번째 릴레이션에는 속하지 않는 투플을 반환함.
- 형식 : R - S

질의 2-5 마당서점 두 지점 중 지점 A에서만 보유하고 있는 도서 목록을 보이시오.

도서A - 도서B

도서A

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000

-

도서B

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

도서번호	도서이름	출판사	가격
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000

그림 2-22 차집합의 예

3.3.4 카티전 프로덕트(cartesian product)

- 두 릴레이션을 연결시켜 하나로 합칠 때 사용함. 결과 릴레이션은 첫 번째 릴레이션의 오른쪽에 두 번째 릴레이션의 모든 투플을 순서대로 배열하여 반환함. 결과 릴레이션의 차수는 두 릴레이션의 차수의 합이며, 카디날리티는 두 릴레이션의 카디날리티의 곱임.

- **형식 : R × S**

질의 2-6 고객 릴레이션과 주문 릴레이션의 카티전 프로덕트를 구하시오 (결과가 많으므로 투플을 일부 삭제한 릴레이션을 사용함).

고객 × 주문

3.3.4 카티전 프로덕트(cartesian product)

고객

고객번호	이름	주소	핸드폰
1	박지성	영국 맨체스타	000-5000-0001
2	김연아	대한민국 서울	000-6000-0001
3	장미란	대한민국 강원도	000-7000-0001

×

주문

주문번호	고객번호	도서번호	판매가격	주문일자
1	2	1	7000	2014-07-01
2	1	2	13000	2014-07-03
3	2	5	8000	2014-07-03
4	1	2	13000	2014-07-04

고객번호	이름	주소	핸드폰	주문번호	고객번호	도서번호	판매가격	주문일자
1	박지성	영국 맨체스타	000-5000-0001	1	2	1	7000	2014-07-01
1	박지성	영국 맨체스타	000-5000-0001	2	1	2	13000	2014-07-03
1	박지성	영국 맨체스타	000-5000-0001	3	2	5	8000	2014-07-03
1	박지성	영국 맨체스타	000-5000-0001	4	1	2	13000	2014-07-04
2	김연아	대한민국 서울	000-6000-0001	1	2	1	7000	2014-07-01
2	김연아	대한민국 서울	000-6000-0001	2	1	2	13000	2014-07-03
2	김연아	대한민국 서울	000-6000-0001	3	2	5	8000	2014-07-03
2	김연아	대한민국 서울	000-6000-0001	4	1	2	13000	2014-07-04
3	장미란	대한민국 강원도	000-7000-0001	1	2	1	7000	2014-07-01
3	장미란	대한민국 강원도	000-7000-0001	2	1	2	13000	2014-07-03
3	장미란	대한민국 강원도	000-7000-0001	3	2	5	8000	2014-07-03
3	장미란	대한민국 강원도	000-7000-0001	4	1	2	13000	2014-07-04

그림 2-23 카티전 프로덕트의 예

3.4 조인(join)

- **두 릴레이션의 공통 속성을 기준으로 속성 값이 같은 투플을 수평으로 결합하는 연산임. 조인을 수행하기 위해서는 두 릴레이션의 조인에 참여하는 속성이 서로 동일한 도메인으로 구성되어야 함. 조인 연산의 결과는 공통 속성의 속성 값이 동일한 투플만을 반환함.**

- **형식 :** $R \bowtie_c S = \sigma_c(R \times S)$ (R과 S는 릴레이션, c 는 조인조건)

- **조인 연산의 구분**
  - **기본연산 : 세타조인($\bowtie_\theta$), 동등조인($\bowtie$), 자연조인($\bowtie_N$)**
  - **확장된 조인 연산 : 세미조인($\ltimes$, $\rtimes$), 외부조인($⟕$, $⟖$, $⟗$)**

3.4.1 세타조인과 동등조인

세타조인(theta join, θ)

- 조인에 참여하는 두 릴레이션의 속성 값을 비교하여 조건을 만족하는 투플만 반환함.
- 세타조인의 조건은 {=, ≠, ≤, ≥, <, >} 중 하나가 됨.
- 형식 : $R \bowtie_{(r\ 조건\ s)} S$ (R과 S는 릴레이션이며 r은 R의 속성, s는 S의 속성)

동등조인(equi join)

- 세타조인에서 = 연산자를 사용한 조인을 말함. 보통 조인 연산이라고 하면 동등조인을 지칭함.
- 형식 : $R \bowtie_{(r = s)} S$

질의 2-7 고객과 고객의 주문 사항을 모두 보이시오.

$고객 \bowtie_{고객.고객번호=주문.고객번호} 주문$

3.4.1 세타조인과 동등조인

고객

고객번호	이름	주소	핸드폰
1	박지성	영국 맨체스타	000-5000-0001
2	김연아	대한민국 서울	000-6000-0001
3	장미란	대한민국 강원도	000-7000-0001
4	추신수	미국 클리블랜드	000-8000-0001

주문

주문번호	고객번호	도서번호	판매가격	주문일자
1	2	1	7000	2014-07-01
2	1	2	13000	2014-07-03
3	2	5	8000	2014-07-03
4	1	2	13000	2014-07-04
5	4	4	35000	2014-07-05
6	5	3	22000	2014-07-07
7	4	3	22000	2014-07-07

고객 $\bowtie_{\text{고객.고객번호=주문.고객번호}}$ **주문**

고객번호	이름	주소	핸드폰	주문번호	고객번호	도서번호	판매가격	주문일자
1	박지성	영국 맨체스타	000-5000-0001	2	1	2	13000	2014-07-03
1	박지성	영국 맨체스타	000-5000-0001	4	1	2	13000	2014-07-04
2	김연아	대한민국 서울	000-6000-0001	1	2	1	7000	2014-07-01
2	김연아	대한민국 서울	000-6000-0001	3	2	5	8000	2014-07-03
4	추신수	미국 클리블랜드	000-8000-0001	5	4	4	35000	2014-07-05
4	추신수	미국 클리블랜드	000-8000-0001	7	4	3	22000	2014-07-07

그림 2-24 동등조인의 예

3.4.2 자연조인(natural join)

- 동등조인에서 조인에 참여한 속성이 두 번 나오지 않도록 두 번째 속성을 제거한 결과를 반환함.
- 형식 : $R \bowtie_{N(r,\ s)} S$

질의 2-8 고객과 고객의 주문 사항을 모두 보여주되 같은 속성은 한 번만 표시하시오.

$고객 \bowtie_{N(고객.고객번호=주문.고객번호)} 주문$

3.4.2 자연조인(natural join)

고객

고객번호	이름	주소	핸드폰
1	박지성	영국 맨체스타	000-5000-0001
2	김연아	대한민국 서울	000-6000-0001
3	장미란	대한민국 강원도	000-7000-0001
4	추신수	미국 클리블랜드	000-8000-0001

주문

주문번호	고객번호	도서번호	판매가격	주문일자
1	2	1	7000	2014-07-01
2	1	2	13000	2014-07-03
3	2	5	8000	2014-07-03
4	1	2	13000	2014-07-04
5	4	4	35000	2014-07-05
6	5	3	22000	2014-07-07
7	4	3	22000	2014-07-07

고객 $\bowtie_{N(\text{고객.고객번호}=\text{주문.고객번호})}$ **주문**

고객번호	이름	주소	핸드폰	주문번호	도서번호	판매가격	주문일자
1	박지성	영국 맨체스타	000-5000-0001	2	2	13000	2014-07-03
1	박지성	영국 맨체스타	000-5000-0001	4	2	13000	2014-07-04
2	김연아	대한민국 서울	000-6000-0001	1	1	7000	2014-07-01
2	김연아	대한민국 서울	000-6000-0001	3	5	8000	2014-07-03
4	추신수	미국 클리블랜드	000-8000-0001	5	4	35000	2014-07-05
4	추신수	미국 클리블랜드	000-8000-0001	7	3	22000	2014-07-07

그림 2-25 자연조인의 예

3.4.3 외부조인과 세미조인

외부조인(outer join)

- 자연조인 시 조인에 실패한 투플을 모두 보여주되 값이 없는 대응 속성에는 NULL 값을 채워서 반환
- 모든 속성을 보여주는 기준 릴레이션 위치에 따라 왼쪽(left) 외부조인, 오른쪽(right) 외부조인, 완전(full) 외부조인으로 나뉨.
- 형식 : 왼쪽(left) 외부조인 – $R ⟕_{(r, s)} S$

  완전(full) 외부조인 – $R ⟗_{(r, s)} S$

  오른쪽(right) 외부조인 - $R ⟖_{(r, s)} S$

R1

A	B
aa	1
cc	2
cc	3

R2

B	C
1	dd
3	ee
4	ff

$R1 ⟕_{(R1.B,\ R2.B)} R2$

A	B	C
aa	1	dd
cc	2	NULL
cc	3	ee

그림 2-26 왼쪽 외부조인의 예

3.4.3 외부조인과 세미조인

질의 2-9　마당서점의 고객과 고객의 주문 내역을 보이시오.

① 고객 기준으로 주문내역이 없는 고객도 모두 보이시오.

② 주문내역이 없는 고객과, 고객 릴레이션에 고객번호가 없는 주문을 모두 보이시오.

③ 주문내역 기준으로 고객 릴레이션에 고객번호가 없는 주문도 모두 보이시오.

① 고객 ⟕ (고객.고객번호, 주문.고객번호) 주문

② 고객 ⟖ (고객.고객번호, 주문.고객번호) 주문

③ 고객 ⟗ (고객.고객번호, 주문.고객번호) 주문

3.4.3 외부조인과 세미조인

고객

고객번호	이름
1	박지성
2	김연아
3	장미란
4	추신수

주문

주문번호	고객번호	판매가격
1	2	7000
2	1	13000
3	2	8000
4	1	13000
5	4	35000
6	5	22000
7	4	22000

고객 ⟕ 고객.고객번호=주문.고객번호 주문

고객번호	이름	주문번호	판매가격
1	박지성	2	13000
1	박지성	4	13000
2	김연아	1	7000
2	김연아	3	8000
3	장미란	**NULL**	**NULL**
4	추신수	**5**	**35000**
4	추신수	7	22000

① 왼쪽 외부조인

고객 ⟗ 고객.고객번호=주문.고객번호 주문

고객번호	이름	주문번호	판매가격
1	박지성	2	13000
1	박지성	4	13000
2	김연아	1	7000
2	김연아	3	8000
3	장미란	**NULL**	**NULL**
4	추신수	5	35000
4	추신수	7	22000
5	**NULL**	6	22000

② 완전 외부조인

고객 ⟖ 고객.고객번호=주문.고객번호 주문

이름	주문번호	고객번호	판매가격
김연아	1	2	7000
박지성	2	1	13000
김연아	3	2	8000
박지성	4	1	13000
추신수	5	4	35000
NULL	6	5	22000
추신수	7	4	22000

③ 오른쪽 외부조인

그림 2-27 외부조인의 예

3.4.3 외부조인과 세미조인

세미조인(semi join)

- 자연조인을 한 후 두 릴레이션 중 한쪽 릴레이션의 결과만 반환하며, 기호에서 닫힌 쪽 릴레이션의 투플만 반환함.
- 형식 : $R \ltimes_{(r,\ s)} S$

질의 2-10 마당서점의 고객 중 주문 내역이 있는 고객의 고객 정보를 보이시오.

$고객 \ltimes_{고객.고객번호,\ 주문.고객번호} 주문$

3.4.3 외부조인과 세미조인

고객

고객번호	이름	주소	핸드폰
1	박지성	영국 맨체스타	000-5000-0001
2	김연아	대한민국 서울	000-6000-0001
3	장미란	대한민국 강원도	000-7000-0001
4	추신수	미국 클리블랜드	000-8000-0001

주문

주문번호	고객번호	도서번호	판매가격	주문일자
1	2	1	7000	2014-07-01
2	1	2	13000	2014-07-03
3	2	5	8000	2014-07-03
4	1	2	13000	2014-07-04
5	4	4	35000	2014-07-05
6	5	3	22000	2014-07-07
7	4	3	22000	2014-07-07

고객 ⋉ (고객.고객번호,주문.고객번호) **주문**

고객번호	이름	주소	핸드폰
1	박지성	영국 맨체스타	000-5000-0001
2	김연아	대한민국 서울	000-6000-0001
4	추신수	미국 클리블랜드	000-8000-0001

그림 2-28 세미조인(왼쪽이 닫힌 경우)의 예

3.5 디비전(division)

- 릴레이션의 속성 값의 집합으로 연산을 수행함.
- 형식 : R ÷ S

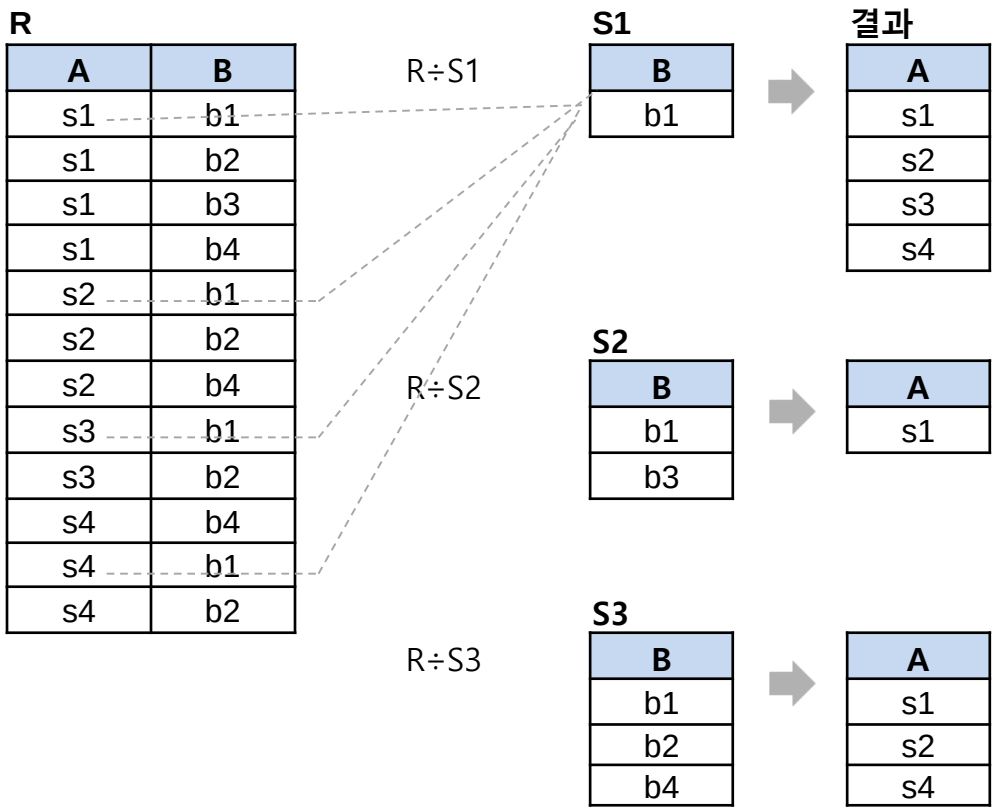

그림 2-29 디비전 연산의 예

3.6.1 셀렉션, 프로젝션, 집합연산의 복합 사용

질의 2-11 마당서점의 도서 중 가격이 8,000원 이하인 도서이름과 출판사를 보이시오.

■ 마당서점의 지점이 하나인 경우

$\pi_{\text{도서이름, 출판사}}(\sigma_{\text{가격}<=8000}\ \text{도서})$

도서

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

$\sigma_{\text{가격}<=8000}$ (도서)

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
5	피겨 교본	굿스포츠	8000

$\pi_{\text{도서이름, 출판사}}$

도서이름	출판사
축구의 역사	굿스포츠
피겨 교본	굿스포츠

그림 2-30 단일 릴레이션에서 셀렉션, 프로젝션 연산의 복합 사용

3.6.1 셀렉션, 프로젝션, 집합연산의 복합 사용

■ 마당서점의 지점이 둘 이상인 경우

$\pi_{\text{도서이름, 출판사}}((\sigma_{\text{가격}<=8000}\ \text{도서A}) \cup (\sigma_{\text{가격}<=8000}\ \text{도서B}))$

도서A

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
2	축구아는 여자	나무수	13000
3	축구의 이해	대한미디어	22000

도서B

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
4	골프 바이블	대한미디어	35000
5	피겨 교본	굿스포츠	8000

$\sigma_{\text{가격}<=8000}$ **(도서A)**

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000

$\cup$

$\sigma_{\text{가격}<=8000}$ **(도서B)**

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
5	피겨 교본	굿스포츠	8000

도서번호	도서이름	출판사	가격
1	축구의 역사	굿스포츠	7000
5	피겨 교본	굿스포츠	8000

$\pi_{\text{도서이름, 출판사}}$

도서이름	출판사
축구의 역사	굿스포츠
피겨 교본	굿스포츠

그림 2-31 두 개 이상의 릴레이션에서 셀렉션, 프로젝션, 합집합 연산의 복합 사용

3.6.2 카티전 프로덕트를 사용한 연산과 조인을 사용한 연산

질의 2-12 마당서점의 박지성 고객의 거래 내역 중 주문번호, 이름, 가격을 보이시오.

■ 카티전 프로덕트를 사용한 연산

$$\pi_{\text{주문.주문번호, 고객.이름, 주문.판매가격}} (\sigma_{\text{고객.고객번호=주문.고객번호 AND 고객.이름='박지성'}} (\text{고객} \times \text{주문}))$$

고객

고객번호	이름	주민번호	주소	핸드폰
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001
2	김연아	900905-2222222	대한민국 서울	000-6000-0001
3	장미란	831009-2333333	대한민국 강원도	000-7000-0001
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001

주문

주문번호	고객번호	도서번호	판매가격	주문일자
1	1	1	7000	2014-07-01
2	1	2	13000	2014-07-03
3	2	5	8000	2014-07-03
4	3	2	13000	2014-07-04
5	4	4	35000	2014-07-05
6	1	3	22000	2014-07-07
7	4	3	22000	2014-07-07

고객 × 주문

고객번호	이름	주민번호	주소	핸드폰	주문번호	고객번호	도서번호	판매가격	주문일자
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	1	1	1	7000	2014-07-01
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	2	1	2	13000	2014-07-03
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	3	2	5	8000	2014-07-03
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	4	3	2	13000	2014-07-04
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	5	4	4	35000	
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	6	1	3	22000	
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	7	4	3	22000	
2	김연아	900905-2222222	대한민국 서울	000-6000-0001	1	1	1	7000	201 -01
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001	5	4	4	35000	2014-07-05
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001	6	1	3	22000	2014-07-07
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001	7	4	3	22000	2014-07-07

결과생략

$\sigma_{\text{고객.고객번호=주문.고객번호 AND 고객.이름='박지성'}}$

고객번호	이름	주민번호	주소	핸드폰	주문번호	고객번호	도서번호	판매가격	주문일자
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	1	1	1	7000	2014-07-01
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	2	1	2	13000	2014-07-03
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	6	1	3	22000	2014-07-07

$\pi_{\text{주문.주문번호,고객.이름,주문.판매가격}}$

주문번호	이름	판매가격
1	박지성	7000
2	박지성	13000
6	박지성	22000

그림 2-32 카티전 프로덕트를 사용한 연산

3.6.2 카티전 프로덕트를 사용한 연산과 조인을 사용한 연산

■ 조인을 사용한 연산

$\pi_{\text{주문번호, 이름, 판매가격}}$ ($\sigma_{\text{이름='박지성'}}$ (고객 $\bowtie_{\text{고객.고객번호=주문.고객번호}}$ 주문))

■ 카티전 프로덕트를 사용한 연산 (위 연산식과 동일함)

$\pi_{\text{주문.주문번호, 고객.이름, 주문.판매가격}}$ ($\sigma_{\text{고객.고객번호=주문.고객번호 AND 고객.이름='박지성'}}$ (고객×주문))

고객

고객번호	이름	주민번호	주소	핸드폰
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001
2	김연아	900905-2222222	대한민국 서울	000-6000-0001
3	장미란	831009-2333333	대한민국 강원도	000-7000-0001
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001

주문

주문번호	고객번호	도서번호	판매가격	주문일자
1	1	1	7000	2014-07-01
2	1	2	13000	2014-07-03
3	2	5	8000	2014-07-03
4	3	2	13000	2014-07-04
5	4	4	35000	2014-07-05
6	1	3	22000	2014-07-07
7	4	3	22000	2014-07-07

$$\text{고객} \bowtie_{\text{고객.고객번호=주문.고객번호}} \text{주문}$$

고객번호	이름	주민번호	주소	핸드폰	주문번호	고객번호	도서번호	판매가격	주문일자
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	1	1	1	7000	2014-07-01
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	2	1	2	13000	2014-07-03
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	6	1	3	22000	2014-07-07
2	김연아	900905-2222222	대한민국 서울	000-6000-0001	3	2	5	8000	2014-07-03
3	장미란	831009-2333333	대한민국 강원도	000-7000-0001	4	3	2	13000	2014-07-04
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001	5	4	4	35000	2014-07-05
4	추신수	820713-1444444	미국 클리블랜드	000-8000-0001	7	4	3	22000	2014-07-07

$$\sigma_{\text{이름='박지성'}}$$

고객번호	이름	주민번호	주소	핸드폰	주문번호	고객번호	도서번호	판매가격	주문일자
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	1	1	1	7000	2014-07-01
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	2	1	2	13000	2014-07-03
1	박지성	810225-1111111	영국 맨체스타	000-5000-0001	6	1	3	22000	2014-07-07

$$\pi_{\text{주문번호,이름,판매가격}}$$

주문번호	이름	판매가격
1	박지성	7000
2	박지성	13000
6	박지성	22000

그림 2-33 조인을 사용한 연산