Өгөгдлийн сангийн үндэс (CSII202 - 3 кр)
Database Systems

Lecture2: Relations and Relational Algebra

МУИС, ХШУИС, МКУТ-ийн багш

Маг. Довдонгийн Энхзол

Хэдийгээр дунд зэргийн чадал авьяастай байлаа ч гэсэн өөртөө том зорилго тавиад түүнийхээ төлөө бүх хүчээрээ тэмцдэг хүн л миний бодлоор аз жаргалтай баймаар санагддаг.

М.И.Калинин

In this Lecture

- The Relational Model
  - Relational data structure
  - Relational data manipulation
- For more information
  - Connolly and Begg – Chapters 3.1-3.2.2 and 4
  - Ullman and Widom – Chapter 3.1, 5.1

Relations

- We will use tables to represent relations:

Anne	aaa@cs.nott.ac.uk
Bob	bbb@cs.nott.ac.uk
Chris	ccc@cs.nott.ac.uk

Relations

- This is a relation between people and email addresses

Anne	aaa@cs.nott.ac.uk
Bob	bbb@cs.nott.ac.uk
Chris	ccc@cs.nott.ac.uk

Relations

- A mathematician would say that it is a set of pairs: <Anne, aaa @ cs.nott.ac.uk>, <Bob, bbb @ cs.nott.ac.uk>, and <Chris, ccc @ cs.nott.ac.uk>.

Anne	aaa@cs.nott.ac.uk
Bob	bbb@cs.nott.ac.uk
Chris	ccc@cs.nott.ac.uk

Relations

- Each value in the first column is a name, each value in the second column is an email address.
- In general, each column has a ***domain*** – a set from which all possible values can come.

Anne	aaa@cs.nott.ac.uk
Bob	bbb@cs.nott.ac.uk
Chris	ccc@cs.nott.ac.uk

Relations

- A mathematical relation is a set of tuples: sequences of values. Each tuple corresponds to a row in a table:

Anne	aaa@cs.nott.ac.uk	0115111111
Bob	bbb@cs.nott.ac.uk	0115222222
Chris	ccc@cs.nott.ac.uk	0115333333

Relations: terminology

- ***Degree of a relation***: how long the tuples are, or how many columns the table has
  - In the first example (name,email) degree of the relation is 2
  - In the second example (name,email,telephone) degree of the relation is 3
  - Often relations of degree 2 are called binary, relations of degree 3 are called ternary etc.
- ***Cardinality of the relation***: how many different tuples are there, or how many different rows the table has.

Relations: mathematical definition

- Mathematical definition of a relation R of degree n, where values come from domains $A_1, \ldots, A_n$:

$R \subseteq A_1 \times A_2 \times \ldots \times A_n$

(relation is a subset of the Cartesian product of domains)

Cartesian product:

$A_1 \times A_2 \times \ldots \times A_n =$

$\{<a_1, a_2, \ldots, a_n>: a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n\}$

Relational model: data manipulation

- Data is represented as relations.
- Manipulation of data (query and update operations) corresponds to operations on relations
- Relational algebra describes those operations. They take relations as arguments and produce new relations.
- Think of numbers and corresponding operators +,-,\, * or booleans and corresponding operators &,|,! (and, or, not).
- Relational algebra contains two kinds of operators: common set-theoretic ones and operators specific to relations (for example projecting on one of the columns).

Union

- Standard set-theoretic definition of union:

$A \cup B = \{x: x \in A \text{ or } x \in B\}$

- For example, $\{a,b,c\} \cup \{a,d,e\} = \{a,b,c,d,e\}$

- For relations, we want the result to be a relation again: a set $R \subseteq A_1 \times \ldots \times A_n$ for some n and domains $A_1,\ldots,A_n$. (or, in other words, a proper table, with each column associated with a single domain of values).

- So we require in order to take a union of relations R and S that R and S have the same number of columns and that corresponding columns have the same domains.

Union-compatible relations

- Two relations R and S are ***union-compatible*** if they have the same number of columns and corresponding columns have the same domains.

Example: not union-compatible

(different number of columns)

Anne	aaa	111111
Bob	bbb	222222
Chris	ccc	333333

Tom	1980
Sam	1985
Steve	1986

Example: not union-compatible

(different domains for the second column)

Anne	aaa
Bob	bbb
Chris	ccc

Tom	1980
Sam	1985
Steve	1986

Example: union-compatible

Anne	1970
Bob	1971
Chris	1972

Tom	1980
Sam	1985
Steve	1986

Union of two relations

- Let R and S be two union-compatible relations. Then their union $R \cup S$ is a relation which contains tuples from both relations:

$R \cup S = \{x: x \in R \text{ or } x \in S\}$.

- Note that union is a partial operation on relations: it is only defined for some (compatible) relations, not for all of them.

- Similar to division for numbers (result of division by 0 is not defined).

Example: shopping lists

R

Cheese	1.34
Milk	0.80
Bread	0.60
Eggs	1.20
Soap	1.00

S

Cream	5.00
Soap	1.00

$R \cup S$

Cheese	1.34
Milk	0.80
Bread	0.60
Eggs	1.20
Soap	1.00
Cream	5.00

Difference of two relations

Let R and S be two union-compatible relations. Then their ***difference*** R - S is a relation which contains tuples which are in R but not in S:

$R - S = \{x: x \in R \text{ and } x \notin S\}$.

- Note that difference is also a partial operation on relations.

Example

R

Cheese	1.34
Milk	0.80
Bread	0.60
Eggs	1.20
Soap	1.00

S

Cream	5.00
Soap	1.00

R - S

Cheese	1.34
Milk	0.80
Bread	0.60
Eggs	1.20

Intersection of two relations

Let R and S be two union-compatible relations. Then their ***intersection*** is a relation $R \cap S$ which contains tuples which are both in R and S:

$R \cap S = \{x: x \in R \text{ and } x \in S\}$

- Note that intersection is also a partial operation on relations.

Example

R

Cheese	1.34
Milk	0.80
Bread	0.60
Eggs	1.20
Soap	1.00

S

Cream	5.00
Soap	1.00

$R \cap S$

Soap	1.00

Cartesian product

- Cartesian product is a total operation on relations.
- Usual set theoretic definition of product:

$R \times S = \{<x,y>: x \in R, y \in S\}$

- Under the standard definition, if <Cheese, 1.34> $\in$ R and <Soap,1.00> $\in$ S, then

< <Cheese, 1.34>, <Soap,1.00>> $\in R \times S$

(the result is a pair of tuples).

Extended Cartesian product

- ***Extended*** Cartesian product flattens the result in a 4-element tuple: <Cheese, 1.34, Soap, 1.00>
- For the rest of the course, “product” means extended product.

Extended Cartesian product of relations

Let R be a relation with column domains $\{A_1,\ldots,A_n\}$ and S a relation with column domains $\{B_1,\ldots,B_m\}$. Then their ***extended Cartesian product*** $R \times S$ is a relation

$$R \times S = \{<c_1,\ldots,c_n,c_{n+1},\ldots,c_{n+m}>:$$

$$<c_1, \ldots,c_n>\in R,\ \ <c_{n+1},\ldots,c_{n+m}>\in S\}$$

Example

R

Cheese	1.34
Milk	0.80
Bread	0.60
Eggs	1.20
Soap	1.00

S

Cream	5.00
Soap	1.00

$R \times S$

Cheese	1.34	Cream	5.00
Milk	0.80	Cream	5.00
Bread	0.60	Cream	5.00
Eggs	1.20	Cream	5.00
Soap	1.00	Cream	5.00
Cheese	1.34	Soap	1.00
Milk	0.80	Soap	1.00
Bread	0.60	Soap	1.00
Eggs	1.20	Soap	1.00
Soap	1.00	Soap	1.00

Projection

- Let R be a relation with n columns, and X is a set of column identifiers (at the moment, we will use numbers, but later we will give then names, like “Email”, or “Telephone”). Then projection of R on X is a new relation $\pi_X(R)$ which only has columns from X.
- For example, $\pi_{1,2}(R)$ is a table with only the 1^{st} and 2^{nd} columns from R.

Example: $\pi_{1,3}(R)$

R:

1	2	3
Anne	aaa@cs.nott.ac.uk	0115111111
Bob	bbb@cs.nott.ac.uk	0115222222
Chris	ccc@cs.nott.ac.uk	0115333333

Example: $\pi_{13}(R)$

$\pi_{13}(R)$:

Anne	0115111111
Bob	0115222222
Chris	0115333333

Selection

- Let R be a relation with n columns and α is a property of tuples.
- ***Selection from R subject to condition*** $\boldsymbol{\alpha}$ is defined as follows:

$$\sigma_{\alpha}(R) = \{<a_1, \ldots, a_n> \in R : \alpha(a_1, \ldots, a_n)\}$$

What is a reasonable property?

- We assume that properties are written using {and, or, not} and expressions of the form col(i) Θ col(j) (where i,j are column numbers) or col(i) Θ v, where v is a value from the domain A_i.

$\forall$ Θ is a comparator which makes sense when applied to values from i and j columns ($=$, $\neq$, maybe also $\leq$, $\geq$, $<$, $>$ if there is a natural order on values).

What is a meaningful comparison

- We can always at least tell if two values in the same domain are equal or not (database values are finitely represented).

- In some cases, it makes sense to compare values from different column domains: for example, if both are domains contain strings, or both contain dates.

- For example, 1975 > 1987 is a meaningful comparison, “Anne” = 1981 is not.

- We can only use a comparison in selection property if its result is true or false, never undefined.

Example: selection

$\forall\ \sigma_{\text{col(3) < 2002 and col(2) = Nolan}}\ (R)$

R

Insomnia	Nolan	2002
Magnolia	Anderson	1999
Insomnia	Skjoldbjaerg	1997
Memento	Nolan	2000
Gattaca	Niccol	1997

Example: selection

$\forall\sigma$ $_{\text{col(3) < 2002 and col(2) = Nolan (R)}}$

Insomnia	Nolan	2002
Magnolia	Anderson	1999
Insomnia	Skjoldbjaerg	1997
Memento	Nolan	2000
Gattaca	Niccol	1997

Example: selection

$\forall\sigma_{\text{col(3) < 2002 and col(2) = Nolan}}\ (R)$

Insomnia	Nolan	2002
Magnolia	Anderson	1999
Insomnia	Skjoldbjaerg	1997
Memento	Nolan	2000
Gattaca	Niccol	1997

Example: selection

$\forall\sigma_{\text{col(3) < 2002 and col(2) = Nolan}}\text{(R)}$

Insomnia	Nolan	2002
Magnolia	Anderson	1999
Insomnia	Skjoldbjaerg	1997
Memento	Nolan	2000
Gattaca	Niccol	1997

Example: selection

$\forall \sigma_{\text{col(3) < 2002 and col(2) = Nolan}} \text{(R)}$

Insomnia	Nolan	2002
Magnolia	Anderson	1999
Insomnia	Skjoldbjaerg	1997
Memento	Nolan	2000
Gattaca	Niccol	1997

Example: selection

$\forall\sigma_{\text{col(3) < 2002 and col(2) = Nolan}}\text{(R)}$

Insomnia	Nolan	2002
Magnolia	Anderson	1999
Insomnia	Skjoldbjaerg	1997
Memento	Nolan	2000
Gattaca	Niccol	1997

Example: selection

∀ $\sigma_{\text{col(3) < 2002 and col(2) = Nolan}}$ (R)

Memento	Nolan	2000

Summary

- Data is represented as tables
- Operations on tables:
  - union of two union-compatible tables (tables with the same number of columns and the same domains for corresponding columns)
  - set difference of two union-compatible tables
  - intersection of two union-compatible tables
  - extended Cartesian product of two tables
  - project a table on some of its columns
  - select rows in a table satisfying a property
- Result of an operation is again a table, so operations can be chained

Example exam question

- What is the result of

$\pi_{1,3}(\sigma_{\text{col(2) = col(4)}}(R \times S))$, where R and S are:

R

Anne	111111
Bob	222222

S

Chris	333333
Dan	111111

Other operations

- Not all SQL queries can be translated into relational algebra operations defined in the lecture.
- Extended relational algebra includes counting and other additional operations.