ÁLGEBRA RELACIONAL

Feliz Gouveia

UFP

fribeiro@ufp.edu.pt

Linguagens relacionais

- Álgebra relacional, mais operacional
- Cálculo relacional, mais declarativo

- Ambas têm o mesmo poder expressivo

Uma álgebra para consultas

- A álgebra relacional permite utilizar relações para consultas
- Simplifica a escrita e a manipulação de consultas
  - A álgebra relacional é uma álgebra de conjuntos
  - SQL utiliza multiconjuntos (coleções onde os elementos podem aparecer repetidos)
- Uma expressão representando uma consulta chama-se um **plano lógico**. É normalmente representado sob forma arborescente, onde os nós são os operadores
- No entanto, o utilizador escreve apenas SQL
  - Internamente, o SGBD transforma as consultas em álgebra relacional

Álgebra Relacional

- Os resultados de um operador são relações
- Por isso diz-se que é a álgebra relacional é **fechada**, e podem-se aplicar operadores sucessivamente
  - Op3(Op2(Op1(R)))

Operadores relacionais: resumo

- União
- Diferença
- Seleção
- Projeção
- Produto

- A estes 5 operadores básicos acrescentaram-se outros

Operadores derivados

- Interseção
- Junção (teta, equi, natural)
- Divisão
- Renomear

- Estes operadores simplificam a escrita das consultas, não acrescentam nada à álgebra

Extensões

- Eliminação de duplicados
- Agrupamento e agregação
- Ordenação (na realidade a ordenação serve apenas para ordenar resultados, não devolve uma relação diferente)

- Junção externa
- Semi-junção
- Anti-junção

Uma álgebra de multiconjuntos

- R ∪ S : um tuplo aparece no resultado o número de vezes que aparece em R mais o número de vezes que aparece em S
- R ∩ S : um tuplo aparece no resultado o mínimo do número de vezes que aparece em R e em S
- R – S : um tuplo aparece no resultado o número de vezes em R menos o número de vezes em S, mas nunca menos do que 0
- R e S devem ter o mesmo esquema

Se R e S fossem conjuntos

- As operações $R \cap S$ e $R - S$ teriam o mesmo resultado que na álgebra de conjuntos
- $R \cup S$ já não teria o mesmo resultado: um elemento que aparece em R e em S iria aparecer duas vezes no resultado

Em SQL

- UNION, INTERSECT e EXCEPT eliminam duplicados do resultado
  - Trabalham sobre multiconjuntos e no fim eliminam tuplos duplicados
- Pode-se especificar ALL para manter duplicados
  - select …. UNION ALL select ….

Operador Seleção

- σ_c (R) : aplica a condição *c* à relação *R*
- Resultado é o multiconjunto que satisfaz *c*
- A condição *c* pode envolver:
  - Operadores artiméticos, sobre texto,…., comparações, e operadores booleanos (AND, …)
- SQL: podem-se ter sub-consultas e usar operadores como EXISTS
- Como as sub-consultas não podem ser expressas na condição *c*, devem ser reescritas como junções

Operador Projeção

- $\pi_L(R)$: projeta a relação R na lista L
- A lista L pode renomear atributos, e usar operações numéricas e sobre texto (por exemplo concatenar texto)
  - diz-se “projeção generalizada”
- O esquema da relação resultante é a lista L
- O resultado pode ter duplicados, mesmo que R não os tenha

Operador Produto

- $R \times S$: o esquema da relação resultante tem todos os atributos de R mais todos de S (qualificados se forem iguais)
- Um tuplo do resultado é formado por cada tuplo de R junto com cada tuplo de S

Operador Junção

- R ⋈ S : equivalente a $\pi_L(\sigma_c (R \times S))$
- Junção Natural: compara as colunas comuns de ambas as tabelas (a condição *c* é "=")
    - Se existirem colunas comuns
    - Projeta apenas uma das colunas comuns
- Junção cartesiana: é o produto
- θ-junção: ou equi-junção se θ for o operador =
    - A equi-junção é a junção mais comum (chamada junção interna)
    - Colunas comparadas aparecem no resultado, ao contrário da junção natural

Renomear

- $\rho(R)$
- Operador usado para renomear colunas ou relações
- $\rho_{\text{ects} \leftarrow \text{créditos}}(\text{disciplina})$

Eliminação de duplicados

- δ (R) : o resultado deste operador não contém duplicados, como em SQL DISTINCT
- Na prática converte um multiconjunto (com duplicados) num conjunto (sem duplicados)

Ordenação

- τ_L (R)
- Converte um multiconjunto numa lista
- Os tuplos da relação R são ordenados pelos atributos especificados na lista L (ordem ascendente por defeito)
- Em SQL, usa-se ORDER BY com os atributos de L
- Para ordenar por ordem descendente usa-se a palavra DESC
- A ordem é ignorada se outro operador for aplicado a seguir
  - Por isso só deve ser usado por último

Agrupamento

- γ_L (R)
- Agrupa tuplos em sub-conjuntos
- A agregação pode usar uma função *f* (min, max, count, sum, avg)
- Os elementos de L podem ser:
  - Atributos a serem agrupados
  - Uma função de agregação

Junções externas

- Permitem tuplos sozinhos (sem um par correspondente) aparecerem no resultado, usando o valor ⊥ (nulo) para preencher o par em falta
- Junções externas à esquerda mantêm todos os tuplos da relação esquerda: R ⟕ S
- Junções externas à direita mantêm todos os tuplos da relação direita: R ⟖ S
- Junções externas totais mantêm tuplos sem correspondência de R e de S: R ⟗ S

Semi-junções

- $R \ltimes S$ ($R \rtimes S$)
- Faz a junção e mantém apenas o esquema à esquerda (direita, respetivamente) sem duplicados
  - Permite retirar do resultado colunas que se sabe não vão ser necessárias
- Não existe um operador correspondente em SQL, mas existem operações equivalentes
- Pode ser escrito:
  - $R \ltimes S = \pi_R(R \bowtie S)$

Anti-junções

- $R \rhd S$, ou $R \lhd S$
- Tuplos de R (ou S) para os quais não existe correspondência em S (ou R)
- Padrão típico de diferença de conjuntos
- Também conhecido por anti-semi-junção, pode ser escrito:
  - $R \rhd S = R - R \ltimes S$, ou
  - $R \rhd S = R \bowtie (R - S)$

Divisão

- $R \div S$
- Tuplos de R para os quais existem todos os valores de todos os tuplos de S nos atributos comuns
- Não existe um operador em SQL, mas existem várias alternativas
- Útil para responder a questões do tipo "todos os que satisfazem uma dada condição"
- Pode ser traduzida como:
- $R \div S \equiv \pi_{R—S}(R) — \pi_{R—S}((\pi_{R—S}(R) \times S) — R)$
- $R(A, B) \div S(B)$ devolve os A que respeitam todos os $S(B)$
- É uma "junção de conjuntos"

Divisão

- Inscrito ÷ disciplina

ALUNO_ID	DISCIPLINA_ID
100	BD
100	PG
150	BD

ID	DISCIPLINA
BD	Bases de Dados
PG	Programação

- resulta em:

ALUNO_ID
100

Tipos de operadores

- Monotónicos
  - A inserção de mais tuplos nunca invalida respostas anteriores, isto é qualquer tuplo presente na resposta anterior também está na resposta após inserção. Por exemplo “qual o nome do aluno 25?”
- Não-monotónicos
  - A inserção de mais tuplos pode invalidar respostas anteriores. Por exemplo “qual o produto mais caro?”

Operadores monotónicos e não-monotónicos

- Monotónicos
  - Os operadores σ, π, $\times$, $\bowtie$, $\cup$ são monotónicos, qualquer consulta que use apenas estes operadores é monotónica
- Não-monotónicos
  - A diferença de conjuntos R – S é não-monotónica em S; juntando um tuplo a S podem deixar de aparecer tuplos no resultado
  - A divisão $R \div S$ é não-monotónica em S; juntando um novo tuplo a S alguns tuplos de R podem deixar de conter todo o S e desaparecem do resultado

Propriedades dos operadores

- As junções, multiplicações, uniões e intersecções são associativas e comutativas
  - $R \bowtie S \equiv R \bowtie S$, $(R \bowtie S) \bowtie T \equiv R \bowtie (S \bowtie T)$
  - $R \times S \equiv R \times S$, $(R \times S) \times T \equiv R \times (S \times T)$
  - $R \cup S \equiv R \cup S$, $(R \cup S) \cup T \equiv R \cup (S \cup T)$
  - $R \cap S \equiv R \cap S$, $(R \cap S) \cap T \equiv R \cap (S \cap T)$
- A seleção é comutativa:
  - $\Sigma_{c1} (\sigma_{c2} (R)) \equiv \sigma_{c2} (\sigma_{c1} (R))$
- A diferença não é associativa nem comutativa
- A junção *teta*, ao contrário da junção natural, pode não ser associativa

Equivalências

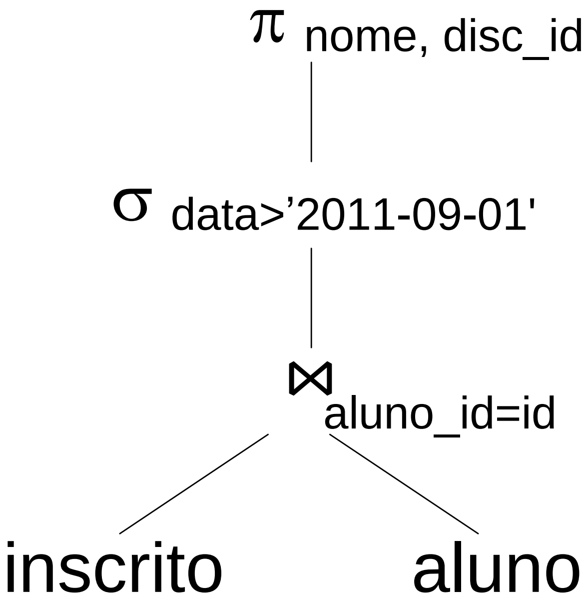
- $\sigma_{A \wedge B}(R) \equiv \sigma_A(\sigma_B(R))$
- $\sigma_{A \vee B}(R) \equiv \sigma_A(R) \cup \sigma_B(R))$
- $\sigma_c(R \cup S) \equiv \sigma_c(R) \cup \sigma_c(S)$
- $\sigma_c(R \cap S) \equiv \sigma_c(R) \cap \sigma_c(S)$
- $\sigma_c(R - S) \equiv \sigma_c(R) - \sigma_c(S)$

Árvores de expressões

- Podem-se aplicar operadores a resultados de operadores, construindo uma árvore
- As folhas da árvore são nomes de relações
- Os nós interiores são operadores
- A execução é ascendente (das folhas para a raiz)
- Geralmente existem várias alternativas equivalentes à mesma árvore, dependendo de como se agrupam as operações

Exemplo de árvore

- É um exemplo típico de SPJ (seleção-projeção-junção)
- Escreva a expressão em álgebra relacional e o SQL de:



Transformações

- As expressões (planos) podem ser transformadas pelo SGBD
- As melhorias conseguidas dependem de informação constante no catálogo da BD (tamanho das colunas e número de tuplos)
- O SGBD tem um componente só para otimizar as consultas: o optimizador de consultas!