

Dependências Funcionais

Fernando Lobo

Base de Dados, Universidade do Algarve

Motivação: Normalização

- Por vezes o esquema relacional tem falhas.
- Geralmente são derivadas de falhas no modelo conceptual.
- Vamos aprender a teoria de “Normalização” de relações, a qual permite eliminar certo tipo de falhas.

Exemplo

Filmes1(nome, ano, duração, nomeEstúdio, moradaEstúdio)

nome	ano	duração	nomeEstúdio	moradaEstúdio
Star Wars	1977	124	Fox	10 Elm St., Los Angeles
Empire Strikes Back	1980	143	Fox	10 Elm St., Los Angeles
Gone With the Wind	1939	181	Paramount	44 Pine St., Los Angeles
Lion King	1994	124	Disney	56 Oak St., Los Angeles
Return of the Jedi	1983	165	Fox	10 Elm St., Los Angeles
Pocahontas	1995	115	Disney	56 Oak St., Los Angeles

- Qual o problema desta relação?

Problemas

① Redundância

- ▶ a morada de um estúdio aparece repetida várias vezes.

② Pode dar origem a anomalias em “updates”.

- ▶ ex: actualizar a morada no 1º tuplo e não o fazer no 2º e 5º tuplo.

③ Pode dar origem a anomalias em “deletes”.

- ▶ ex: se apagarmos o filme “Gone With the Wind”, o estúdio “Paramount” desaparece da BD!

Normalização

- Estudo de Normalização de relações permite-nos decompor uma relação em várias relações de modo a que não haja este tipo de anomalias.
- Antes de aprender a fazê-lo, precisamos de aprender o conceito de *Dependência Funcional*.

Dependências Funcionais

- Generalização do conceito de chave.
- Notação: $X \rightarrow Y$
- Significado: se fixarmos valores para os atributos X , os valores dos atributos Y são únicos (i.e., não há repetidos).
 - ▶ Nota: X e Y são conjuntos de atributos.
 - ▶ Usaremos \dots, X, Y, Z para representar conjuntos de atributos; A, B, C, \dots para atributos individuais.
 - ▶ Por vezes vamos omitir o uso de chavetas para representar conjuntos, i.e. usaremos ABC em vez de $\{A, B, C\}$.

No nosso exemplo ...

Filmes1(nome, ano, duração, nomeEstúdio, moradaEstúdio)

nome	ano	duração	nomeEstúdio	moradaEstúdio
Star Wars	1977	124	Fox	10 Elm St., Los Angeles
Empire Strikes Back	1980	143	Fox	10 Elm St., Los Angeles
Gone With the Wind	1939	181	Paramount	44 Pine St., Los Angeles
Lion King	1994	124	Disney	56 Oak St., Los Angeles
Return of the Jedi	1983	165	Fox	10 Elm St., Los Angeles
Pocahontas	1995	115	Disney	56 Oak St., Los Angeles

- nome ano → duração nomeEstúdio moradaEstúdio
- nomeEstúdio → moradaEstúdio

Definição é independente da instância

- A definição de DF é independente da instância da relação.
- Uma DF é uma asserção que é obedecida por todas as possíveis instâncias de uma relação.
- O mesmo acontece com o conceito de chave.

Chaves e Superchaves

- X é uma **superchave** da relação R se:
 - ▶ $X \rightarrow$ todos os atributos de R .
- X é uma **chave** da relação R , se nenhum subconjunto próprio de X for superchave de R (i.e., X é uma “superchave mínima”).
 - ▶ NOTA: Y é subconjunto próprio de X , se $Y \subseteq X \wedge Y \neq X$
- X é chave de $R \Rightarrow X$ é superchave de R .

Exemplo

Filmes1(nome, ano, duração, nomeEstúdio, moradaEstúdio)

Com DF's:

nome ano → duração nomeEstúdio moradaEstúdio

nomeEstúdio → moradaEstúdio

- {nome,ano} é chave de Filmes1. Porquê?
- Quantas superchaves há na relação Filmes1?

DF's triviais

- $X \rightarrow Y$ é **trivial** se $Y \subseteq X$
- Exemplos
 - ▶ nome ano \rightarrow nome
 - ▶ nome \rightarrow nome

Regras sobre DF's

- $X \rightarrow A_1 A_2 \dots A_n$
- É equivalente a:
 - ▶ $X \rightarrow A_1$
 - ▶ $X \rightarrow A_2$
 - ▶ ...
 - ▶ $X \rightarrow A_n$
- Em inglês, “Splitting Rule”.

Não se pode fazer o mesmo do lado esquerdo

Exemplo:

- nome ano \rightarrow duração
- não é equivalente a:
 - ▶ nome \rightarrow duração
 - ▶ ano \rightarrow duração
- Porquê?

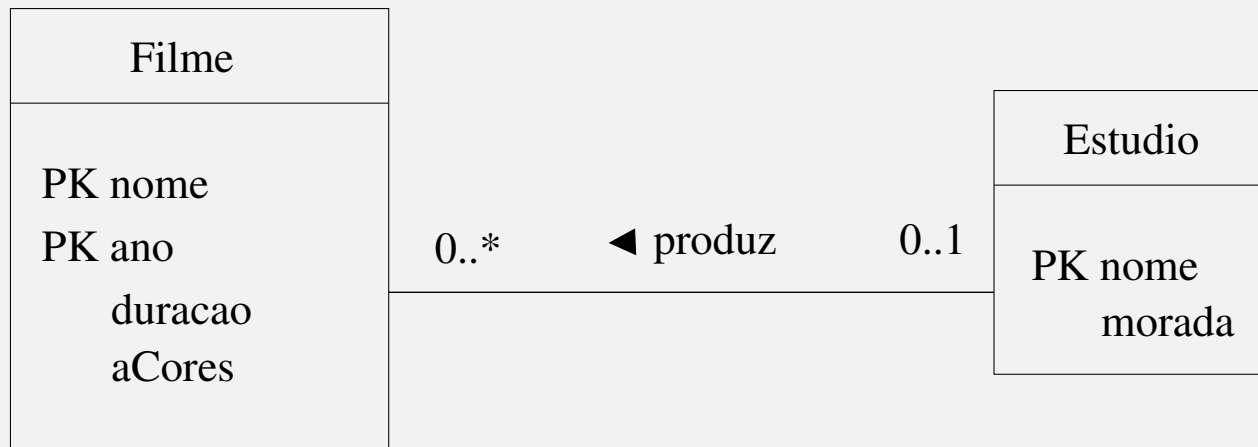
Regras sobre DF's (cont.)

- **Transitividade**: Se $X \rightarrow Y$ e $Y \rightarrow Z$, então $X \rightarrow Z$.
- **“Augmentation”**: Se $X \rightarrow Y$, então $XZ \rightarrow Y$.

De onde vêm as DF's

- 1 Chaves obtidas no modelo conceptual.
- 2 Relações muitos-um.
- 3 Conhecimento do problema.

Exemplo: casos 1 e 2



Filmes(nome, ano, duração, aCores, nomeEstúdio)

- nome ano → duração aCores
- nome ano → nomeEstúdio

Exemplo: conhecimento sobre o problema (caso 3)

- Não pode haver 2 disciplinas que funcionam no mesmo dia da semana, à mesma hora, e na mesma sala.
 - ▶ sala dia hora → disciplina

Fecho de um conjunto de atributos

- Notação: o fecho de X designa-se por X^+ .
- Significado: X^+ é o conjunto de todos os atributos que são funcionalmente determinados por X .
- Algoritmo para calcular X^+ :
 - 1 $X^+ = X$
 - 2 Se $A_1, A_2, \dots, A_m \rightarrow B$, e todos os A 's estão contidos em X^+ , adicionar B a X^+ .
 - 3 Termina quando não conseguirmos adicionar mais nada a X^+ .

Exemplo

- Dada a relação $R(ABCDEF)$ com as seguintes DF's:

- ▶ $AB \rightarrow C$

- ▶ $BC \rightarrow A$

- ▶ $BC \rightarrow D$

- ▶ $D \rightarrow E$

- ▶ $CF \rightarrow B$

$(AB)^+$ = ?
= AB
= ABC // porque $AB \rightarrow C$
= $ABCD$ // porque $BC \rightarrow D$
= $ABCDE$ // porque $D \rightarrow E$

Projeção de DF's

- Nota: Projeção = eliminar algumas colunas.
- Motivação: Normalização, o processo de decompor uma relação em 2 ou mais relações.
- Exemplo:
 - ▶ $R(ABCD)$ com DF's: $AB \rightarrow C$, $C \rightarrow D$, e $D \rightarrow A$.
 - ▶ Decompor em $R_1(ABC)$ e $R_2(AD)$.
 - ▶ Quais as DF's em R_1 ?
 - ▶ $AB \rightarrow C$, mas também $C \rightarrow A$.

Projectão de DF's (cont.)

Dada uma relação R com um conjunto de DF's, projectar R em R_1 . Quais as DF's de R_1 ?

- 1 Começar com as DF's de R e encontrar todas as DF's não triviais que se podem deduzir a partir delas.
- 2 Ficar apenas com as DF's cujos atributos façam parte do esquema de R_1 .

Algoritmo para projectar as DF's de uma relação R numa relação R_1

- 1 Seja T o eventual output. Inicialmente, $T = \emptyset$.
- 2 Para cada subconjunto X de atributos de R_1 , calcular X^+ (usando as DF's de R).
- 3 Inserir $X \rightarrow A$ em T , para todo o $A \in (X^+ - X)$, desde que A seja atributo de R_1 .

Simplificação

- Calcular o fecho de todos os subconjuntos de atributos é trabalhoso.
- Simplificação nº1:
 - ▶ Não é necessário calcular o fecho do conjunto vazio, nem do conjunto de todos os atributos.
- Simplificação nº2:
 - ▶ Se $X^+ =$ todos os atributos, não é necessário calcular o fecho de subconjuntos que contenham X .

Exemplo

- Seja $R(ABCD)$ com DF's $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$.
- Projectar R em $R_1(ACD)$ (i.e., eliminar o atributo B).

Fechos	T
$A^+ = ABCD$	$A \rightarrow C, A \rightarrow D$
$C^+ = CD$	$C \rightarrow D$
$D^+ = D$	
$(CD)^+ = CD$	

- T é o conjunto de DF's de R_1 . Mas podemos eliminar $A \rightarrow D$ porque pode ser deduzida a partir das outras duas.
- DF's de R_1 : $A \rightarrow C$, $C \rightarrow D$.

Equivalência de conjuntos de DF's

- Dois conjuntos de DF's S e T são **equivalentes**, se o conjunto de instâncias que satisfaz S for o mesmo que o conjunto de instâncias que satisfaz T .
- No exemplo que vimos há pouco:
 - ▶ $\{A \rightarrow C, A \rightarrow D, C \rightarrow D\}$ é equivalente a $\{A \rightarrow C, C \rightarrow D\}$

Base mínima de DF's

- Seja S um conjunto de DF's para uma relação R .
- Qualquer conjunto de DF's equivalente a S é uma *base* para S .
- Uma *base mínima* para uma relação R , é uma base B que satisfaça:
 - 1 Todas as DF's em B , têm no lado direito um só atributo.
 - 2 Se removermos uma das DF's de B , deixa-se de ter uma base.
 - 3 Se removermos um atributo do lado esquerdo de uma das DF's de B , deixa-se de ter uma base.