



Universidade Nova de Lisboa

OMNIS CIVITAS CONTRA SE DIVISA NON STABIT

Faculdade de Ciências e Tecnologia



Dependências Funcionais; Formas Normais – Resumo e exercícios

(13 Jan 2017. Aula prática 2)

O modelo de dados que desenhámos até agora não foi nem deverá ser determinado pelas aplicações particulares possíveis relacionadas com o vídeo-club: pesquisas particulares, relatórios particulares, etc. Deve apenas ser determinado pela Normalização com base no conjunto de dependências funcionais explícitas e implícitas (deduzíveis). O que fizemos até aqui foi a aplicação indirecta dessa Normalização recorrendo a algumas regras práticas mas nem por isso menos robustas. Vamos agora verificar se o fizemos correctamente, recorrendo para isso à Normalização.

A chamada Normalização serve para evitar as redundâncias do Modelo de Dados, causadoras de inconsistências, e para facilitar o acesso aos dados de forma rigorosa e completa. Eis as Formas Normais mais importantes que a constituem.

1ª Forma Normal

Um esquema R (ou se quisermos, uma tabela R) está na primeira forma normal se os domínios dos seus atributos forem atómicos, isto é, se os seus elementos forem unidades indivizíveis. Por exemplo, imagine-se que a tabela sócio tinha um atributo chamado *filhos_sócio* onde se escrevia, para cada sócio, os nomes dos filhos: *João, Maria, Luis* para o sócio cujo *num_sócio* é igual a 5, e *Pedro Manuel, Joana* para o sócio com *num_sócio* igual a 6. Claramente, estes conteúdos não são atómicos porque é necessário retirar de cada cadeia de caracteres os nomes individuais dos filhos. Esta tabela não estaria na 1ª Forma Normal. Outro exemplo de falha da 1ª Forma Normal está relacionado com o Código Postal. Assim, imaginemos que a tabela sócio tinha um campo separado chamado *cod_postal_sócio* que pretendia conter o Código Postal da morada do sócio e que, era sistematicamente preenchido com a estrutura *Numero – Localidade*. Também aqui a conteúdo não é atómico. Esta falha não é grave mas gera redundância, já que se houver pelo menos dois sócios com o mesmo Código Postal, a Localidade aparecerá repetida.

2ª Forma Normal

Um esquema R está na 2ª Forma Normal se e só se estiver na 1ª e se cada atributo de R não pertencente à chave depender funcionalmente de toda a chave. Na prática esta exigência é satisfeita se R estiver na 3ª Forma Normal, como vamos compreender.

Regras de Armstrong

Quando nos é dado um conjunto de dependências funcionais, normalmente há outras dependências que se podem concluir a partir desse conjunto e de regras coerentes. As regras de Armstrong são coerentes e a sua aplicação exaustiva a um conjunto de dependências funcionais, formam aquilo a que se chama o fecho (conjunto completo) de dependências funcionais que se podem deduzir de forma coerente. Eis as regras:

Se $\beta \subseteq \alpha$, então $\alpha \rightarrow \beta$ (reflexividade)

Se $\alpha \rightarrow \beta$, então $\gamma\alpha \rightarrow \gamma\beta$ (aumento)

Se $\alpha \rightarrow \beta$, e $\beta \rightarrow \gamma$, então $\alpha \rightarrow \gamma$ (transitividade)

3ª Forma Normal

Um esquema R está na 3ª Forma Normal se e só se para toda a dependência funcional $\alpha \rightarrow \beta \in F^+$, pelo menos uma das seguintes condições é verdadeira (F^+ é o fecho (conjunto completo) de todas as dependências funcionais directas e deduzíveis pelas regras de Armstrong.):

. $\alpha \rightarrow \beta$ é trivial (i.e., $\beta \subseteq \alpha$)

. α é superchavede R (i.e., $\alpha \rightarrow R \in F^+$)

. Todo o atributo $A \subseteq (\beta - \alpha)$ está contido numa chave candidata de R.

(NOTA: cada um desses atributos pode pertencer a uma chave candidata distinta)

Por outras palavras R está na 3ª Forma Normal se nenhum dos seus atributos depende de outros atributos que não façam parte de uma das chaves candidatas de R.

Para ilustrar a 3ª Forma Normal, vamos supor que a dependência funcional $ano_filme \rightarrow preço_dia_filme \in F^+$: nesse caso, esta dependência colocaria a tabela filme fora da 3ª forma normal, uma vez que o atributo $preço_dia_filme$ dependeria de um atributo (ano_filme) que não faz parte de nenhuma chave candidata de *filme*.

Podemos considerar que a 3ª Forma Normal é a meta mínima a atingir no processo de Normalização dum Modelo de Dados. A 3ª Forma Normal não evita redundâncias.

Forma Normal de Boyce-Codd (BCNF)

Um esquema R diz-se na BCNF, relativamente a um conjunto de dependências F, se e só se para toda a dependência em F^+ da forma $\alpha \rightarrow \beta$, onde $\alpha \subseteq R$ e $\beta \subseteq R$, pelo menos uma das condições é verdadeira:

- . $\alpha \rightarrow \beta$ é trivial (i.e., $\beta \subseteq \alpha$)
- . α é superchave de R (i.e., $\alpha \rightarrow R \in F^+$)

Por outras palavras, só podem existir dependências funcionais não triviais aplicáveis a um esquema R, se a parte esquerda da dependência for superchave. Esta exigência é mais forte do que a determinada pela 3ª Forma Normal. Se um esquema R estiver na BCNF também está na 3ª Forma Normal. A BCNF evita redundâncias e constitui uma meta a tentar atingir na concepção dos Modelos de Dados.

Imaginemos que tínhamos incluído o atributo nome_ator na tabela filme_ator. Existiriam agora mais novas dependências funcionais a ter em conta, por exemplo: $cod_filme, nome_ator \rightarrow cod_filme, nome_ator, cod_actor$, sendo por isso outra chave candidata, e $cod_actor \rightarrow nome_ator$, o que implica que a tabela filme_ator estaria na 3ª Forma Normal (nome_ator em $cod_actor \rightarrow nome_ator$ faria parte duma chave candidata) mas não estaria na BCNF precisamente porque cod_actor em $cod_actor \rightarrow nome_ator$ não é superchave em filme_ator.

Exercício 1

Com base na análise das dependências funcionais, Crie um Modelo de Dados na 4FN, para um consultório médico. Considere consultas, doentes, médicos e especialidades (estomatologia, pneumologia, etc.). Cada médico pode ser especialista em mais do que uma especialidade. Nas consultas os médicos receitam medicamentos (um ou mais). Existem também marcações de consultas. Considere ainda como válidas as seguintes dependências funcionais do conjunto F:

$cod_medico \rightarrow nome_médico$

$(cod_médico, nome_médico) \rightarrow (cod_especialidade, nome_especialidade)$

$nome_especialidade \rightarrow cod_especialidade$

$cod_especialidade \rightarrow nome_especialidade$

$cod_fármaco \rightarrow nome_fármaco$

$cod_doente \rightarrow Nome_doente$

$(data_consulta, hora_consulta, cod_médico) \rightarrow (cod_doente, nome_doente, data_marcação, cod_especialidade)$

$(data_consulta, hora_consulta, cod_doente) \rightarrow (cod_médico, nome_médico, data_marcação, cod_especialidade)$

$(data_consulta, hora_consulta, cod_médico, cod_fármaco) \rightarrow qt_receita$

Antes de mais, é conveniente encontrar uma cobertura canónica G de F.

Assim, vemos que $nome_médico$ é redundante à esquerda em $(cod_médico, nome_médico) \rightarrow (cod_especialidade, nome_especialidade)$, uma vez que F implica logicamente uma cobertura em que a DF que contem a redundância é substituída por outra sem essa redundância (ver slide 7.23), ou seja: por pseudo-transitividade (PT) as DFs $cod_medico \rightarrow nome_médico$ e $(cod_medico, nome_médico) \rightarrow (cod_especialidade, nome_especialidade)$ derivam $cod_medico \rightarrow (cod_especialidade, nome_especialidade)$. Esquemáticamente,

$$\begin{array}{l} cod_medico \rightarrow nome_médico \\ (cod_medico, nome_médico) \rightarrow (cod_especialidade, nome_especialidade) \\ \hline \end{array} \quad (PT)$$

$cod_medico \rightarrow (cod_especialidade, nome_especialidade)$

Por outro lado, nesta nova DF, ainda há uma redundância à esquerda: $nome_especialidade$.

Com efeito, se substituirmos esta DF por outra sem esta redundância, obteremos uma nova cobertura que implica logicamente a anterior. Esquemáticamente:

$$\begin{array}{l} cod_medico \rightarrow cod_especialidade \qquad cod_especialidade \rightarrow nome_especialidade \\ \hline \end{array} \quad (T)$$

$cod_medico \rightarrow nome_especialidade$

$$\begin{array}{l} \qquad \qquad \qquad cod_medico \rightarrow cod_especialidade \\ \hline cod_medico \rightarrow (cod_especialidade, nome_especialidade) \\ \end{array} \quad (U)$$

Aplicando os mesmos mecanismos de eliminação de redundância à esquerda e à direita, juntamente com a união de DFs com a mesma parte esquerda, pode obter-se a cobertura canónica G:

$cod_medico \rightarrow (nome_médico, cod_especialidade)$
 $cod_especialidade \rightarrow nome_especialidade$
 $nome_especialidade \rightarrow cod_especialidade$
 $cod_fármaco \rightarrow nome_fármaco$
 $cod_doente \rightarrow Nome_doente$
 $(data_consulta, hora_consulta, cod_médico) \rightarrow (cod_doente, data_marcação)$
 $(data_consulta, hora_consulta, cod_doente) \rightarrow (cod_médico, data_marcação)$

$(data_consulta, hora_consulta, cod_médico, cod_fármaco) \rightarrow qt_receita$

O esquema inicial é $R=(cod_medico, nome_médico, cod_especialidade, nome_especialidade, cod_fármaco, nome_fármaco, cod_doente, nome_doente, data_consulta, hora_consulta, data_marcação, qt_receita)$. Uma das chaves de R é $(data_consulta, hora_consulta, cod_médico, cod_fármaco)$ dado que $\{data_consulta, hora_consulta, cod_médico, cod_fármaco\}^+ = R$. Outra chave (conjunto de atributos cujo fecho inclui todos os atributos de R) é $(data_consulta, hora_consulta, cod_doente, cod_fármaco)$. Acontece que R não está na BCNF já que, por exemplo, a dependência funcional (DF) $cod_doente \rightarrow nome_doente$ viola esta forma normal porque cod_doente não é chave em R nem esta DF é trivial. Assim, é necessário decompor o esquema R .

Decomposição (ver slide 7.36):

$Result:=R$

$R = R1$

$\alpha = cod_doente \quad \beta = nome_doente$

$Result := (Result - R1) \cup (R1 - \beta) \cup (\alpha, \beta)$

Assim, $R2=(cod_doente, nome_doente)$, $R3=(cod_medico, nome_médico, cod_especialidade, nome_especialidade, cod_fármaco, nome_fármaco, cod_doente, data_consulta, hora_consulta, data_marcação, qt_receita)$. $R2$ está na BCNF porque nenhuma outra DF se aplica a $R2$.

Mas, $R3$ não está na BCNF por violação, por exemplo, de $cod_especialidade \rightarrow nome_especialidade$. Isso faz evoluir a decomposição para: $R2 \cup R4 \cup R5$, sendo $R4 = (cod_especialidade, nome_especialidade)$ e $R5=(cod_medico, nome_médico, cod_especialidade, cod_fármaco, nome_fármaco, cod_doente, data_consulta, hora_consulta, data_marcação, qt_receita)$. $R4$ está na BCNF mas $R5$ não está porque $cod_médico \rightarrow (nome_médico, cod_especialidade)$ viola a BCNF uma vez que $cod_médico$ não é chave em $R5$ nem a DF em questão é trivial. Assim, $R5$ decompor-se-á e na próxima iteração do algoritmo, $Result = R2 \cup R4 \cup R6 \cup R7$, com $R6 = (cod_médico, nome_médico, cod_especialidade)$ que está na BCNF, e $R7 = (cod_medico, cod_fármaco, nome_fármaco, cod_doente, data_consulta, hora_consulta, data_marcação, qt_receita)$. Porém, $cod_fármaco \rightarrow nome_fármaco$ impede $R7$ de estar na BCNF. Na próxima iteração, $Result = R2 \cup R4 \cup R6 \cup R8 \cup R9$, com $R8 = (cod_fármaco, nome_fármaco)$, que está na BCNF, e $R9 = (cod_medico, cod_fármaco, cod_doente, data_consulta, hora_consulta, data_marcação, qt_receita)$. Mas $R9$ não está na BCNF porque em $(data_consulta, hora_consulta, cod_médico) \rightarrow (cod_doente, data_marcação, data_consulta, hora_consulta, cod_médico)$ a parte esquerda da DF não é chave em $R9$ nem a DF é trivial. Assim $Result = R2 \cup R4 \cup R6 \cup R8 \cup R10 \cup R11$ com $R10 = (data_consulta, hora_consulta, cod_médico, cod_doente, data_marcação)$ que está na BCNF e $R11=(cod_medico, data_consulta, hora_consulta, cod_fármaco, qt_receita)$, que está na BCNF.

Assim, a base de dados será composta pelos seguintes esquemas que, dada a sua composição, pode ter os seguintes nomes de tabelas e conteúdos:

Doente = (cod_doente, nome_doente)

Especialidade = (cod_especialidade, nome_especialidade)

Médico = (cod_médico, nome_médico, cod_especialidade)

Fármaco = (cod_fármaco, nome_fármaco)

Consulta = (data_consulta, hora_consulta, cod_médico, cod_doente, data_marcação)

Receita = (cod_medico, data_consulta, hora_consulta, cod_fármaco, qt_receita)

Quanto à preservação das dependências, é fácil ver que o fecho da união das DFs que se aplicam a cada esquema coincide com o fecho da cobertura G, isto é:

F_Doente = { cod_medico → nome_médico, cod_especialidade };

F_especialidade = { cod_especialidade → nome_especialidade, nome_especialidade → cod_especialidade }

F_Médico = { cod_medico → nome_médico, cod_especialidade }

F_Fármaco = { cod_fármaco → nome_fármaco }

F_Marcação = { data_consulta, hora_consulta, cod_médico → cod_doente, data_marcação, data_consulta, hora_consulta, cod_doente → cod_médico, data_marcação }

F_Receita = { data_consulta, hora_consulta, cod_médico, cod_fármaco → qt_receita }

Por outras palavras, {F_Doente ∪ F_especialidade ∪ F_Médico ∪ F_Fármaco ∪ F_Marcação ∪ F_Receita}⁺ = G⁺.

Exercício 2:

a) Verifique se a base de dados sobre o clude de vídeo, desenhada por meio do DER, está na BCNF.

Considerem-se como válidas as seguintes DFs de F:

cod_género → nome_género

cod_editora → nome_editora

cod_actor → nome_actor

cod_realizador → nome_realizador

cod_filme → (nome_filme, ano_filme, preço_dia_filme, dias_sem_multa_filme, multa_dia_filme, cod_género, cod_editora)

(num_copia, cod_filme, data_aluguer) → (num_sócio, data_devolução, estado_devolução)

num_sócio → (nome_sócio, bi_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio)

bi_sócio → (nome_sócio, num_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio)

Eis as tabelas:

Sócio=(num_sócio, nome_sócio, bi_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio)

Género=(cod_género, nome_género)

Editora=(cod_editora, nome_editora)

Actor=(cod_actor, nome_actor)

Realizador=(cod_realizador, nome_realizador)

Filme=(cod_filme, nome_filme, ano_filme, preço_dia_filme, dias_sem_multa_filme, multa_dia_filme, cod_género, cod_editora)

Filme_actor=(cod_filme, cod_actor) *Ffilme_realizador*=(cod_filme, cod_realizador)

Cópia=(cod_filme, num_cópia)

Aluquer=(cod_filme, num_cópia, data_aluquer, num_sócio)

Devolução=(cod_filme, num_cópia, data_aluquer, data_devolução, estado_devolução)

Teremos agora que verificar se cada esquema (correspondente a cada tabela) está na BCNF. Assim, seguindo o slide 7.38, vemos que não é preciso computar F+ (o fecho completo de F) mas é preciso computar o fecho de cada subconjunto dos atributos da cada esquema *R* e ver se esse fecho se limita a ele mesmo ou ao cunjunto *R*. Se isso não acontecer, *R* não está na BCNF.

Assim, no esquema *Género*, temos o atriburo *cod_género*, cujo fecho $\{cod_género\}^+ = \{cod_género, nome_genero\}$ dado que contem *cod_género* por reflexividade e, por aplicação da DF $cod_género \rightarrow nome_genero$, contém *nome_género*, ou seja, o fecho contém o esquema inteiro. Por outro lado, o fecho de *nome_género* é apenas $\{nome_género\}$. Nenhuma outra DF se aplica a esque esquema, logo *Género* está na BCNF.

Uma análise idêntica prova que os esquemas *Editora*, *Actor* e *Realizador* estão também na BCNF. Quanto ao esquema *Sócio*, apenas os fechos de *num_sócio*, de *bi_sócio* e os dos subconjuntos contendo um ou dois deste atributos é que não contêm apenas os próprios atributos (por exemplo, mesmo $\{morada_sócio, tlf_sócio\}^+ = \{morada_sócio, tlf_sócio\}$). Quanto a $\{num_sócio\}^+ = \{num_sócio, nome_sócio, bi_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio\}$, ou seja todos os atributos do esquema, devido à reflexividade de *num_sócio* e à DF $num_sócio \rightarrow nome_sócio, bi_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio$ (ver slide 7.19). Devido a $bi_sócio \rightarrow (nome_sócio, num_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio)$, e à reflexividade de *bi_sócio*, também $\{bi_sócio\}^+ = \{bi_sócio, nome_sócio, num_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio\}$. Assim, também o esquema *Sócio* está na BCNF.

Quanto ao esquema *Filme*, também se percebe que, apenas os subconjuntos de atributos deste esquema e que incluem o atributo *cod_filme* é que têm um fecho que inclui outros atributos, em particular, inclui todos os atributos do esquema (repare-se na DF $cod_filme \rightarrow (nome_filme, ano_filme, preço_dia_filme, dias_sem_multa_filme, multa_dia_filme, cod_gênero, cod_editora)$). Todos os outros subconjuntos têm um fecho limitado ao próprio conjunto do qual se quer calcular o fecho, tendo em conta o conteúdo do esquema. Por isso, *Filme* está também na BCNF. De notar que, embora $\{cod_gênero\}^+ = \{cod_gênero, nome_gênero\}$ (repare-se na DF $cod_gênero \rightarrow nome_gênero$), o atributo *nome_gênero* não faz parte do esquema em causa e portanto verifica-se a condição definida no slide 7.38: (the attribute closure of α either includes no attributes of $R_i - \alpha$ or includes all attributes of R_i . Sendo $\alpha = cod_gênero$ e $R_i = (cod_filme, nome_filme, ano_filme, preço_dia_filme, dias_sem_multa_filme, multa_dia_filme, cod_gênero, cod_editora)$).

Quanto ao esquema *Filme_actor*, a única Df que se lhe aplica é a que se obtém por reflexividade $cod_filme, cod_actor \rightarrow cod_filme, cod_actor$, que não viola a condição em 7.38. O Com efeito, as Dfs que envolvem cada um dos atributos ($cod_filme \rightarrow nome_filme$ e $cod_actor \rightarrow nome_actor$) não violam a referida condição. *Filme_actor* está pois na BCNF. Uma análise semelhante se aplica ao esquema ao esquema *Filme_realizador* e concluir-se-á facilmente que está na BCNF.

Relativamente ao esquema *Cópia* não há qualquer DF que se lhe aplique e que viole a condição em 7.38, como facilmente se conclui. De facto, $\{num_cópia\}^+$ apenas contém *num_cópia* por reflexividade e, relativamente ao fecho de *cod_filme*, todos os atributos nele contidos e diferentes de *cod_filme* não estão neste esquema. Este esquema está portanto na BCNF.

Quanto ao esquema *Aluguer*, uma vez que tem 4 atributos $\{cod_filme, num_cópia, data_aluguer, e num_sócio\}$, existem potencialmente $2^4 - 1 = 15$ subconjuntos cujos fechos há que considerar. No entanto, apenas o fecho $\{cod_filme, num_cópia, data_aluguer\}^+ = \{cod_filme, num_cópia, data_aluguer\}$, obtido com base na DF $(num_cópia, cod_filme, data_aluguer) \rightarrow num_sócio$ (por sua vez obtida por decomposição da DF $(num_cópia, cod_filme, data_aluguer) \rightarrow (num_sócio, data_devolução, estado_devolução)$) requer análise, já que os restantes subconjuntos de atributos têm fechos limitados aos próprios elementos de cada conjunto. Assim, dado que o fecho $\{cod_filme, num_cópia, data_aluguer\}^+$ contém o conjunto completo de atributos do esquema, também aqui não há violação da condição no slide 7.38. O esquema *Aluguer* está também na BCNF.

Relativamente ao esquema *Devolução*, também só a $(num_cópia, cod_filme, data_aluguer) \rightarrow (num_sócio, data_devolução, estado_devolução)$ requer a nossa análise, já que todos os outros

subconjuntos derivados do esquema e diferentes de $\{num_copia, cod_filme, data_aluguer\}$ geram fechos iguais ao conjunto dos seus atributos. Assim, como se vê facilmente, o fecho $\{num_copia, cod_filme, data_aluguer\}^+$ contem todos os atributos do esquema devido à referida DF e à reflexividade dos atributos à esquerda dessa mesma DF. Logo, também este esquema está na BCNF. E portanto toda a base de dados está na BCNF.

b) Verifique se a base de dados Preserva as DFs.

As DFs serão preservadas se $(F_Sócio \cup F_Género \cup F_Editora \cup F_Actor \cup F_Realizador \cup F_Filme \cup F_Filme_actor \cup F_Filme_realizador \cup F_Cópia \cup F_Aluguer \cup F_devolução)^+ = F^+$.

Em que F_X é o conjunto de DFs de F^+ que se aplicam ao esquema X . Assim, começando por $F_Sócio$, as DFs que são aplicáveis a este esquema são:

$num_sócio \rightarrow (nome_sócio, bi_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio)$

$bi_sócio \rightarrow (nome_sócio, num_sócio, data_nsc_sócio, morada_sócio, tlf_sócio, sexo_sócio)$

e todas as que derivam por reflexividade dos $2^9 - 1 = 511$ subconjuntos não vazios formados a partir destes 9 atributos.

Quanto a $F_Género$, as DFs que se lhe aplicam são a $cod_género \rightarrow nome_género$ e as que derivam por reflexividade e decomposição: $(cod_género, nome_género) \rightarrow nome_género$; $(cod_género, nome_género) \rightarrow cod_género$; $cod_género \rightarrow cod_género$ e $nome_género \rightarrow nome_género$.

Relativamente aos restantes conjuntos de DFs, seria necessário fazer o mesmo trabalho. No entanto, não é difícil perceber que não o fecho da união destes conjuntos de DFs é equivalente ao fecho F^+ , já que, não existem DFs “escondidas e não refletidas” nos esquemas.

Exercício 3:

a) Verifique se a base de dados sobre o tema que o seu grupo escolheu, desenhada por meio do DER, está na BCNF. Caso não esteja, verifique se está na 3FN.

