Álgebra Relacional

- Tópicos:
  - Definição da Álgebra Relacional
  - Operadores básicos da Álgebra Relacional
  - Operações estendidas em Álgebra Relacional
  - Modificações de Bases de Dados
  - Vistas
- Bibliografia:
  - Secções 2.5, 2.6 e 6.1 do livro recomendado
  - Capítulos 2 e 3 do livro *The theory of relational databases*

Linguagem de Consulta/Interrogação

- E agora que temos uma base de dados relacional, já com dados, o que é que fazemos com ela?
  - Programas que manipulam os dados na base de dados
  - Programas que consultam os dados que estão na base de dados (e depois fazem "coisas" com eles)
- São necessárias linguagens que acedam e modificam os dados
  - Linguagens de interrogação e manipulação
- Categorias de linguagens:
  - Procedimentais (para dizer **como** obter os dados)
  - Declarativas (para dizer **que** dados queremos)
- Linguagens Declarativas
  - Como estamos a trabalhar sobre um modelo que não reflete necessariamente a forma como os dados estão armazenados, as procedimentais não são muito apropriadas
  - As declarativas tornam mais fácil lidar com a complexidade e relações entre os dados

Interrogação de bases de dados

- Como deve ser uma linguagem de interrogação?
- O que são perguntas (ou interrogações)?
- E como são as respostas?
- Uma resposta a uma pergunta a uma base de dados relacional é uma relação (ou tabela).
  - Eg. a resposta à pergunta quais os clientes com contas com saldos inferiores a 10, e de quanto são esses saldos, é uma relação com atributos nome (ou nº) de cliente e saldo, e cujos tuplos são aqueles que se pretendem
- Logo uma pergunta é uma "função" (ou, melhor, uma operação) que, dado um conjunto de relações, devolve uma relação
- Para "formalizar" uma pergunta precisamos de um conjunto de operadores que operem sobre relações
- **Álgebra relacional** como linguagem de interrogação

Álgebra Relacional

- Seis operadores básicos
  - seleção
  - projeção
  - união
  - diferença de conjuntos
  - produto cartesiano
  - renomeação
- Os operadores têm como argumentos relações de entrada e devolvem uma relação como resultado (i.e. são fechados sobre relações)

Operação de Seleção

- Notação: $\sigma_p(r)$
- *p* é designado por predicado de seleção
- Definida como:

$$\sigma_p(r) = \{t \mid t \in r \textbf{ e } p(t)\}$$

onde *p* é uma fórmula do cálculo proposicional constituída por termos ligados por: $\wedge$ (**e**), $\vee$ (**ou**), $\neg$ (**não**)

  - Cada termo é da forma:

<atributo> *op* <atributo> ou <constante>

onde *op* pode ser: $=, \neq, >, \geq. <. \leq$

- Exemplo de seleção:

$$\sigma_{\textit{branch-name}=\textit{'Perryridge'} \wedge \text{balance}>100}(\textit{account})$$

- Propriedades da seleção:
  - Comutatividade: $\sigma_{P1}(\sigma_{P2}(r)) = \sigma_{P2}(\sigma_{P1}(r)) = \sigma_{P1 \wedge P2}(r))$

Operação de Seleção – Exemplo

- Relação **r**

r

A	B	C	D
α	α	1	**7**
α	β	5	**7**
β	β	12	3
β	β	23	**10**

- $\sigma_{A=B \wedge D > 5}$ (r)

$\sigma_{A = B \wedge D > 5}$ **(r)**

A	B	C	D
α	α	1	**7**
β	β	23	**10**

Operação de Projeção

- Notação:

$$\prod_{A1,\, A2,\, \ldots,\, Ak} (r)$$

  onde $A_1, \ldots, A_k$ são nomes de atributos e r é uma relação.
- O resultado é a relação com os k atributos (*k* colunas) obtidos eliminando os atributos (colunas) que não estão listadas
- Relembrar: o resultado não tem tuplos duplicados, pois as relações são conjuntos
- E.g. eliminar o atributo *branch-name* de *account*

$$\prod_{account\text{-}number,\ balance} (account)$$

- Propriedades da Projeção:
  - Absorção: $\prod_{Cn}(\ldots \prod_{C1}(r)\ ) = \prod_{Cn}(r)$
  - Comutatividade com seleção: $\prod_{Cn}(\sigma_P(r)) = \sigma_P(\prod_{Cn}(r))$

Operação de Projeção – Exemplo

- Relação *r*:

r

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

- $\prod_{A,C}(r)$

A	C
α	1
α	1
β	1
β	2

$\Rightarrow$

$\prod_{A,C}(r)$

A	C
α	1
β	1
β	2

(Integridade de referência com álgebra relacional)

- Agora que definimos o operador de projeção temos outra forma de definir chaves estrangeiras!

- Os atributos *A* na relação *r* são chave estrangeira referenciando o atributos *B* na relação *s* sse

$$\prod_{A}(r) \subseteq \prod_{B}(s)$$

- Exemplo:
  - *id* em *customer* ser chave estrangeira referenciado *id* em *person*, significa/impõe que:

$$\prod_{id}(customer) \subseteq \prod_{id}(person)$$

Operação de União

- Notação: $r \cup s$
- Definida como:

$$r \cup s = \{t \mid t \in r \textbf{ ou } t \in s\}$$

- Para $r \cup s$ ser válida:

  1. *r, s* devem ter a mesma *aridade* (igual número de atributos)

  2. Os domínios dos atributos devem ser *compatíveis* (e.g., os valores da 2ª coluna de *r* são do mesmo tipo dos valores da 2ª coluna de *s*)

- E.g. determinar quais os clientes que têm uma conta ou um empréstimo

$$\prod_{customer\text{-}name}(depositor) \cup \prod_{customer\text{-}name}(borrower)$$

Operação de União – Exemplo

- Relações *r, s:*

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

- r ∪ s:

r ∪ s

A	B
α	1
α	2
β	1
β	3

Propriedade da União

- Propriedades
  - Comutatividade: $r \cup s = s \cup r$
  - Associatividade: $r \cup (s \cup t) = (r \cup s) \cup t$
  - Distributividade sobre seleção: $\sigma_P(r \cup s) = \sigma_P(r) \cup \sigma_P(s)$
  - Distributividade sobre projeção: $\prod_L(r \cup s) = \prod_L(r) \cup \prod_L(s)$

Operação de Diferença de Conjuntos

- Notação: $r - s$
- Definida como:

$$r - s = \{t \mid t \in r \textbf{ e } t \notin s\}$$

- As diferenças de conjuntos só podem ser efectuadas entre relações compatíveis.
  - *r* e *s* devem ter a mesma aridade
  - os domínios dos atributos de *r* e *s* devem ser compatíveis
- Propriedades
  - Distributividade sobre seleção: $\sigma_P(r - s) = \sigma_P(r) - \sigma_P(s)$

Operação de Diferença de Conjuntos-Ex.

- Relações *r, s:*

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

- *r* – *s*:

r - s

A	B
α	1
β	1

Operação de Produto Cartesiano

- Notação: $r \times s$
- Definida como:

$$r \times s = \{tq \mid t \in r \textbf{ e } q \in s\}$$

- Assume que os atributos de r(R) e s(S) são disjuntos. (Ou seja, $R \cap S = \emptyset$).
- Se os atributos de *r(R)* e *s(S*) não são disjuntos, então têm que se utilizar renomeações:
  - Se o nome das relações for diferente, distinguem-se os atributos com o mesmo nome prefixando-os com o nome da relação.
  - Se não for possível usar o nome da relação para desambiguar os atributos com o mesmo nome (e.g. as relações têm o mesmo nome, ou o nome é ambíguo por resultar de operações compostas), veremos mais à frente...

Operação de Produto Cartesiano-Ex.

- Relações **r** e **s**:

r

A	B
α	1
β	2

s

C	D	E
α	10	a
α	13	a
β	20	b
γ	10	b

- **r** x **s**:

r x s

A	B	C	D	E
α	1	α	10	a
α	1	α	13	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	α	13	a
β	2	β	20	b
β	2	γ	10	b

Propriedade do Produto

- Propriedades
  - Comutatividade: r x s = s x r
  - Associatividade: r x (s x t) = (r x s) x t

Composição de Operações

- Pode-se construir expressões combinando várias operações
- Exemplo: $\sigma_{A=C}(r \times s)$

r x s

A	B	C	D	E
α	1	α	10	a
α	1	α	13	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	α	13	a
β	2	β	20	b
β	2	γ	10	b

$\sigma_{A=C}(r \times s)$

A	B	C	D	E
α	1	α	10	a
α	1	α	13	a
β	2	β	20	b

Operação de Renomeação

- Permite dar um nome aos resultados de expressões de álgebra relacional.
- Permite que uma relação seja referida por mais de um nome.
- A expressão:

$$\rho_X(E)$$

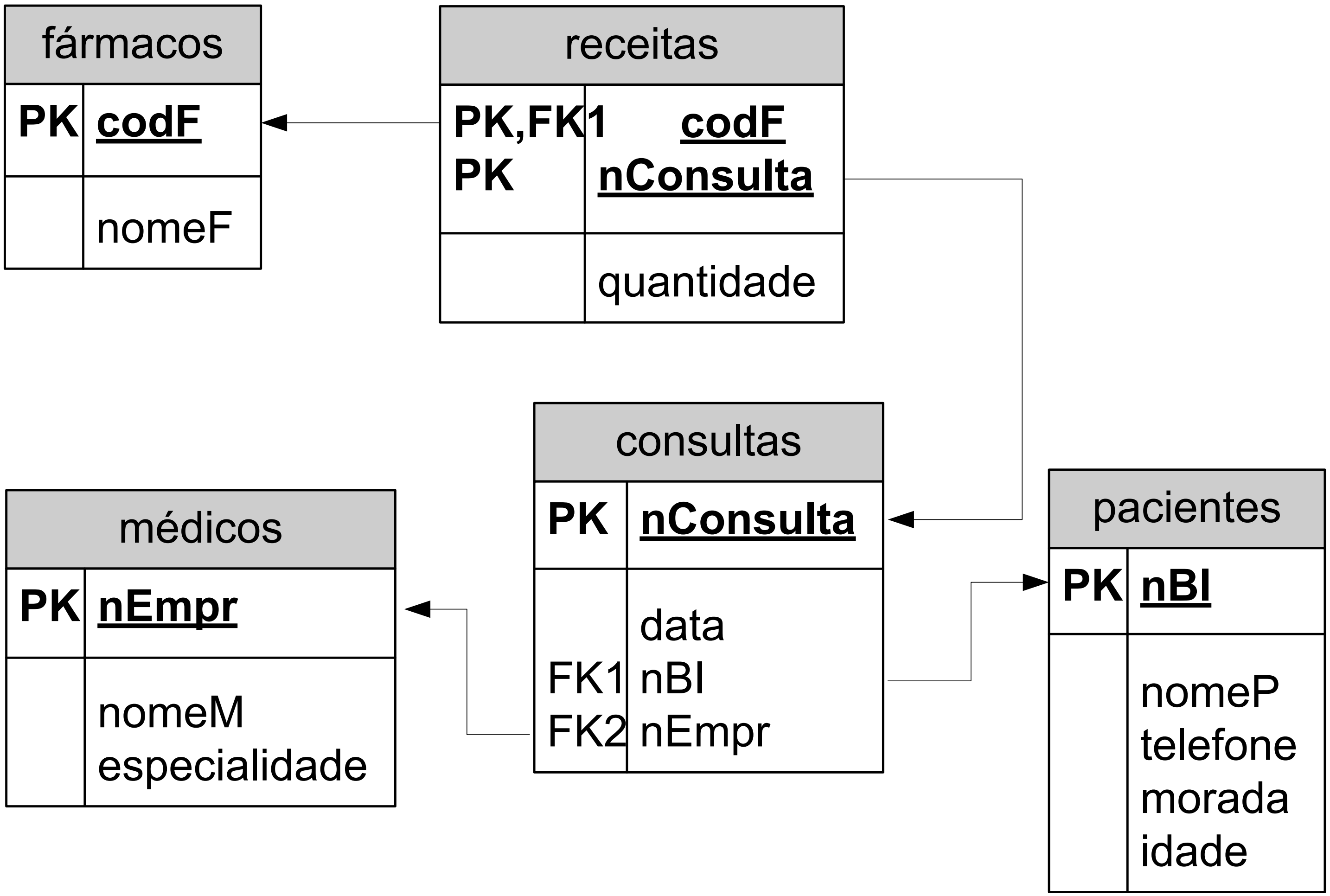
devolve a expressão *E* com o nome *X*

- Se uma expressão de álgebra relacional *E* tem aridade *n*, então

$$\rho_{X\,(A1,\, A2,\, \ldots,\, An)}\,(E)$$

devolve a expressão *E* com o nome *X*, e com os atributos renomeados para *A1, A2, …., An*.

Exemplo Clínica



Exemplo Clínica

- *médicos(<u>nEmpr</u>,nomeM,especialidade)*
- *pacientes(<u>nBI</u>,nomeP,telefone,morada,idade)*
- *fármacos(<u>codF</u>,nomeF)*
- *consultas(<u>nConsulta</u>,data,nBI,nEmpr)*
- *receitas(<u>codF,nConsulta</u>,quantidade)*

Exemplo Clínica

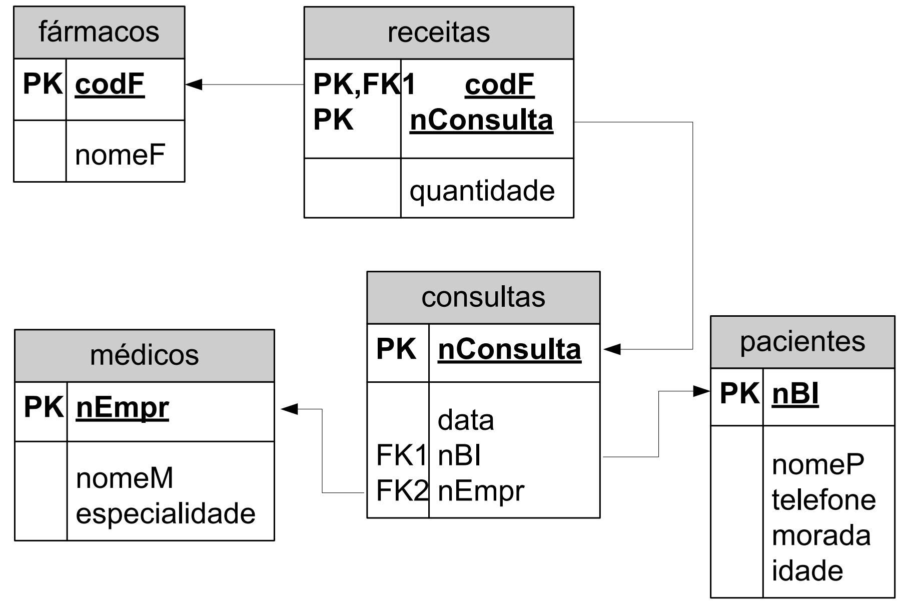
médicos(nEmpr,nomeM,especialidade)

fármacos(codF,nomeF)

pacientes(nBI,nomeP,telefone,morada,idade)

consultas(nConsulta,data,nBI,nEmpr)

receitas(codF,nConsulta,quantidade)



- Quais os pacientes com mais de 50 anos de idade?

$$\sigma_{idade > 50}(pacientes)$$

- Quais os nomes dos pacientes com mais de 50 anos de idade?

$$\Pi_{nomeP}(\sigma_{idade > 50}(pacientes))$$

Exemplo Clínica

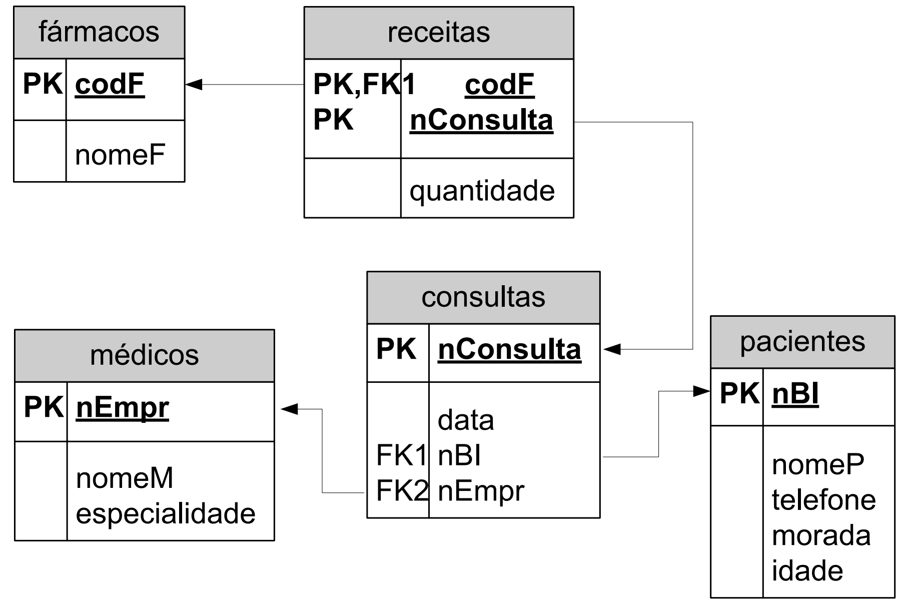
médicos(<u>nEmpr</u>,nomeM,especialidade)

fármacos(<u>codF</u>,nomeF)

pacientes(<u>nBI</u>,nomeP,telefone,morada,idade)

consultas(<u>nConsulta</u>,data,nBI,nEmpr)

receitas(<u>codF,nConsulta</u>,quantidade)



- Quais os nomes dos fármacos que já foram receitados em consultas da clínica?

$$\Pi_{nomeF}((\sigma_{receitas.codF\ =\ fármacos.codF}(receitas\ x\ fármacos)))$$

Exemplo Clínica

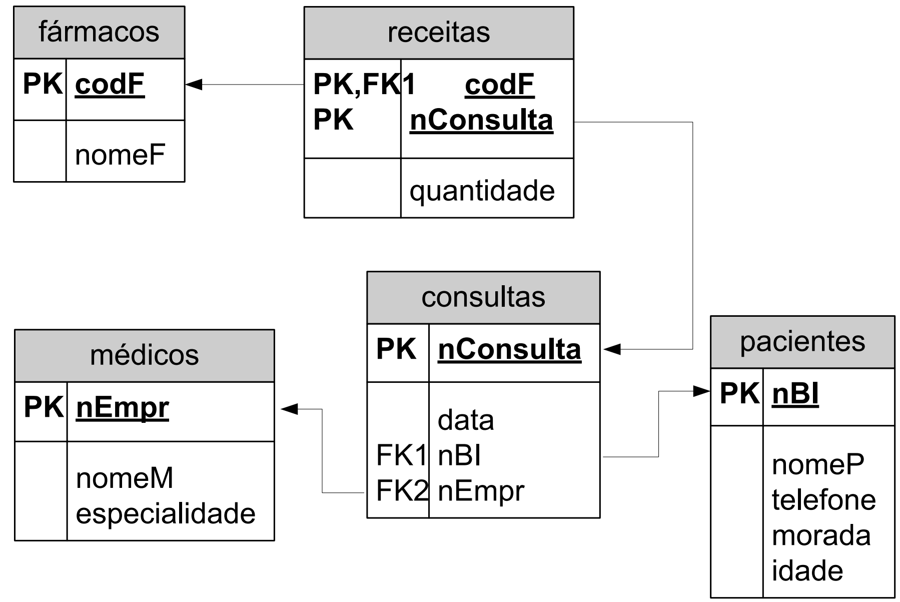
médicos(nEmpr,nomeM,especialidade)

fármacos(codF,nomeF)

pacientes(nBI,nomeP,telefone,morada,idade)

consultas(nConsulta,data,nBI,nEmpr)

receitas(codF,nConsulta,quantidade)



- Quais os nomes e códigos dos fármacos que nunca foram receitados?

fármacos –

$$\Pi_{receitas.codF,nomeF}(\sigma_{receitas.codF\,=\,fármacos.codF}(receitas \text{ x } fármacos)))$$

Exemplo Clínica

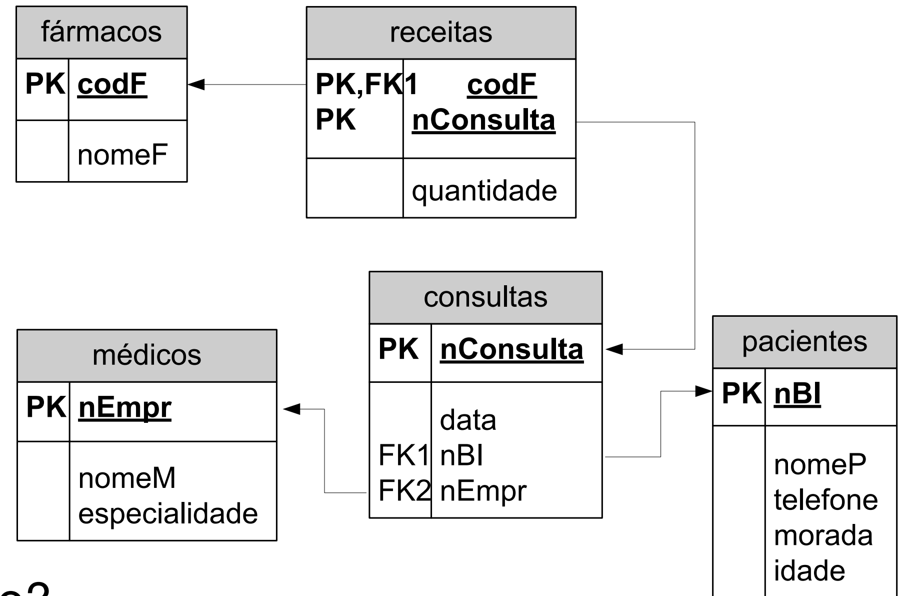
médicos(nEmpr,nomeM,especialidade)

fármacos(codF,nomeF)

pacientes(nBI,nomeP,telefone,morada,idade)

consultas(nConsulta,data,nBI,nEmpr)

receitas(codF,nConsulta,quantidade)



- Qual a idade do paciente mais velho?
  - Renomear a relação *pacientes* como *d*
  - A consulta é:

$$\Pi_{idade}(pacientes) - \Pi_{pacientes.idade}(\sigma_{pacientes.idade < d.idade}(pacientes \times \rho_d(pacientes)))$$

Exemplo Clínica

pacientes

Nome	Idade
Ana	30
Rui	20
Carla	25

$\Rightarrow$

ρ_d(pacientes)

d.Nome	d.Idade
Ana	30
Rui	20
Carla	25

$\Rightarrow$

pacientes x ρ_d(pacientes)

Nome	Idade	d.Nome	d.Idade
Ana	30	Ana	30
Ana	30	Rui	20
Ana	30	Carla	25
Rui	20	Ana	30
Rui	20	Rui	20
Rui	20	Carla	25
Carla	25	Ana	30
Carla	25	Rui	20
Carla	25	Carla	25

$\Leftarrow$

$\sigma_{\text{pacientes.Idade < d.Idade}}$ (pacientes x ρ_d(pacientes))

Nome	Idade	d.Nome	d.Idade
Rui	20	Ana	30
Rui	20	Carla	25
Carla	25	Ana	30

$\Downarrow$

$\Pi_{\text{pacientes.Idade}}$ ($\sigma_{\text{pacientes.Idade < d.Idade}}$ (pacientes x ρ_d(pacientes)))

Idade
20
25

Π_{Idade} (pacientes)

Idade
30
20
25

Π_{Idade} (pacientes) –
$\Pi_{\text{pacientes.Idade}}$ ($\sigma_{\text{pacientes.Idade < d.Idade}}$ (pacientes x ρ_d(pacientes)))

$\Rightarrow$

Idade
30

Exemplo Clínica

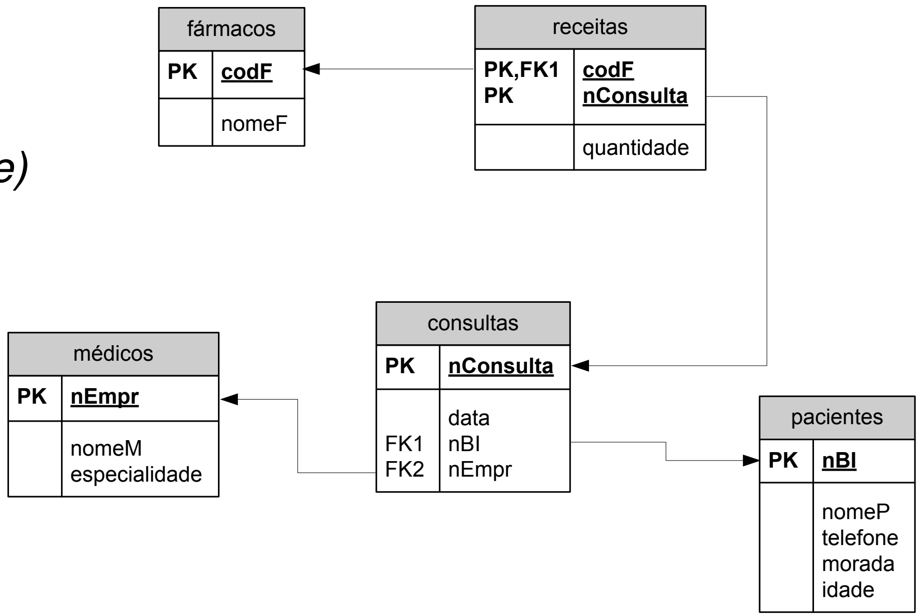
médicos(<u>nEmpr</u>,nomeM,especialidade)

pacientes(<u>nBI</u>,nomeP,telefone,morada,idade)

fármacos(<u>codF</u>,nomeF)

consultas(<u>nConsulta</u>,data,nBI,nEmpr)

receitas(<u>codF</u>,<u>nConsulta</u>,quantidade)



- E quais os (nomes dos) pacientes com essa idade?
  - Seja r a relação da pergunta anterior:

$$\Pi_{nomeP}(\sigma_{\ pacientes.idade\ =\ r.idade}\ (pacientes\ x\ r))$$

Definição Formal

- Uma expressão básica na álgebra relacional é:
  - Uma relação na base de dados
  - Uma relação constante
- Sejam E_1 e E_2 expressões de álgebra relacional; então todas as expressões abaixo são expressões de álgebra relacional:
  - $E_1 \cup E_2$
  - $E_1 - E_2$
  - $E_1 \times E_2$
  - $\sigma_p (E_1)$, P é um predicado nos atributos de E_1
  - $\prod_S(E_1)$, S é uma lista com alguns dos atributos de E_1
  - $\rho_x (E_1)$, x é um novo nome para o resultado de E_1