

Lista de Exercícios (respostas)

Bancos de Dados
Instituto de Computação
Universidade Estadual de Campinas

Modelagem e Query
2019
André Santanchè

Questão 1

Uma indústria farmacêutica quer desenvolver um banco de dados para registrar os medicamentos que ela produz, bem como os vírus tratados por estes medicamentos. Elabore um modelo conceitual (ER e UML) para este banco de dados conforme o detalhamento a seguir:

- Devem ser armazenados os nomes científicos e populares dos vírus bem como os períodos de incubação.
- Para medicamentos, o devem ser armazenados o nome de venda e o composto ativo.
- Considere que um dado medicamento pode tratar vários vírus e um vírus pode ser tratado por vários medicamentos.
- O banco de dados também precisa armazenar informações sobre o tipo de paciente (e.g. criança, adulto, idoso) infectado por um vírus e se este tipo pode ser tratado pelos respectivos medicamentos. Cada tipo de paciente possui uma dosagem recomendada para a combinação paciente/medicamento.

Vírus podem ser classificados em diversas categorias. A categoria retrovírus é tratada com coquetéis de medicamentos. Um coquetel é composto por vários medicamentos, cada um em uma concentração específica.

Os tratamentos de retrovírus baseados em coquetéis também devem especificar dosagens específicas por tipo de paciente.

Considere dois cenários de restrição:

- somente retrovírus são tratados com coquetéis
- retrovírus só são tratados com coquetéis

Questão 2

Problema geral

Uma organização de saúde está estudando doenças e possui o seguinte conjunto de dados:

- nome e número de habitantes das cidades em que se manifestam doenças;
- nome científico e popular de doenças em estudo;
- CPF, nome e cidade de residência de pessoas que foram contaminadas por doenças (apenas aquelas doenças que estão em estudo por esta organização) e com quais doenças elas foram contaminadas.

Primeira parte

A partir deste problema:

- Desenvolva um modelo ER e um modelo UML a ser usado como base de um banco de dados que armazene estes dados. Busque um modelo mínimo que represente apenas os dados declarados acima e nada além.
- Este modelo poderia ter mais relacionamentos do que aqueles apresentados. Liste que relacionamentos seriam estes e um conjunto de questões relevantes relacionadas a

este contexto de estudo da organização que não podem ser respondidas pela falta de tais relacionamentos.

Segunda parte

Dada uma cidade A que apresenta casos de contaminação por uma doença X. Considere a hipótese de que quanto maior a proximidade de uma cidade B em relação a A, maior a sua probabilidade de ser contaminada por uma doença X que se manifestou em A.

A partir desta hipótese, realize modificações no modelo, de modo que mantenha informações que permitam à organização realizar análises desta natureza. Não é necessário o registro da referida probabilidade, mas do conjunto de dados que permitam inferi-la.

Terceira parte

Proponha um modelo alternativo de proximidade entre cidades e o represente no seu modelo, considerando que você sabe que estradas ligam que cidades.

Quarta parte

Modifique o modelo de modo que seja possível consultar onde e quando (data) as pessoas foram contaminadas por uma doença, de modo que seja possível acompanhar o seu progresso. O modelo deve levar em consideração que a organização sabe em que cidades uma doença X se manifesta, independentemente de registrar casos específicos de pessoas contaminadas pela doença X naquela cidade.

Quinta parte

Se considerarmos que o modelo registra dois tipos de informação: cidades em que houve contaminação de uma doença X (doença X se manifestou na cidade A); pessoas que foram contaminadas por uma doença X e as respectivas cidades da contaminação. Produza dois modelos alternativos para atender os seguintes cenários:

- O modelo deve garantir que só pode haver registros de pessoas contaminadas por uma dada doença X, naquelas cidades em que há registro que a doença se manifestou.
- É possível haver registros de pessoas contaminadas por uma doença X na cidade A, ainda que não haja registro que a doença X se manifestou na cidade A.

Questão 3

Um plano de contas é uma estrutura hierárquica utilizada por sistemas contábeis em empresas, que têm as seguintes características:

- Cada item de um plano de contas (denominado conta) possui um número definido conforme a hierarquia, conforme o recorte abaixo:
 1. Ativo
 - 1.1. Circulante
 - 1.1.1. Disponível
 - 1.1.2. Contas a Receber
 - 1.2. Realizável a Longo Prazo
 2. Passivo
 - 2.1. Circulante
 - ...
- Contas de níveis mais baixos (sem elementos hierarquicamente subordinados) possuem valores atribuídos.
- Contas que possuem elementos subordinados têm seu valor calculado a partir destes elementos.

a) Construa um diagrama ER e um UML para representar este plano de contas. Leve em

consideração critérios como redundância e consistência de valores.

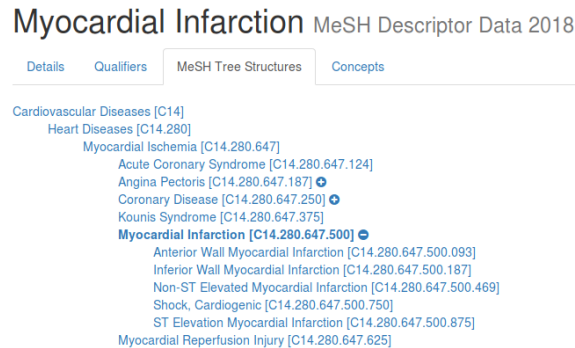
	<p>Tentativa 1 - Se todas as contas fossem iguais. Problema: Não são todas iguais - folhas (contas de níveis mais baixos) não podem ter filhos. A decisão se contas que não são folhas têm saldo depende da opção: se tiverem Saldo haverá redundância para questões de performance, mas podem ter problemas de consistência.</p>
	<p>Tentativa 2 - Usar apenas herança para diferenciar. Problema: Por causa da herança, Folhas podem equivocadamente ter subcontas.</p>
	<p>Tentativa 3: Colocar a Folha como entidade fraca ligada a conta. Problema: Parece correto do ponto de vista do modelo lógico a ser gerado, mas está semanticamente confuso, já que Folha não é uma entidade diferente de Conta.</p>
	<p>Tentativa 4: Duas especializações de Conta. Boa solução. Problema: Neste caso, foram criados dois tipos de relação e o ideal seria apenas uma.</p>
	<p>Tentativa 5: Melhor que a anterior, pois é igualmente expressiva, mas unifica as relações de superior.</p>

b) Transforme o modelo ER em um esquema relacional.

Questão 4

O Medical Subject Headings - MeSH (<https://www.nlm.nih.gov/mesh/meshhome.html>) organiza hierarquicamente sua terminologia para indexação e catálogo de informações biomédicas.

Ao digitar o nome de uma doença no MeSH, ele traz uma hierarquia de sua classificação. Por exemplo, procurar por Myocardial Infarction (Infarto do Miocárdio) ele traz a página:



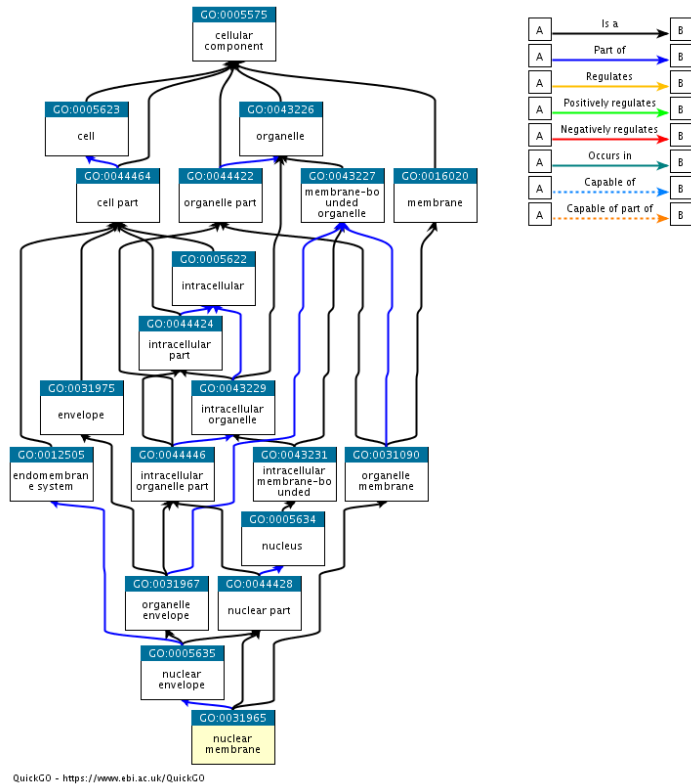
<https://meshb.nlm.nih.gov/record/ui?ui=D009203>

Note que há uma estrutura de codificação que reflete a hierarquia: C14 - Doenças Vasculares; C14.280 - Doenças do Coração; e assim por diante. Construa um modelo conceitual (ER e UML) para representar as informações do mesmo tipo das contidas nesse exemplo, inclusive com a relação hierárquica entre os termos. Considere que a hierarquia pode ter um número diferente de níveis em partes diferentes da representação.

Apesar de o MeSH aceitar mais de uma hierarquia para o mesmo termo, para fins de simplificação, será considerado que cada termo só possui uma hierarquia.

Questão 5

O GeneOntology - GO (<http://geneontology.org/>) define um modelo computacional para a representação de sistemas biológicos. Dentre eles, há uma estrutura de classificação na forma de Grafo Orientado Direcionado que representa componentes de células vivas. Por exemplo, ao procurar por Nuclear Membrane (Membrana Nuclear), o GO apresenta a seguinte classificação:



<https://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0031965>

Note que as relações entre os componentes podem ter tipos (indicados por cores à esquerda).

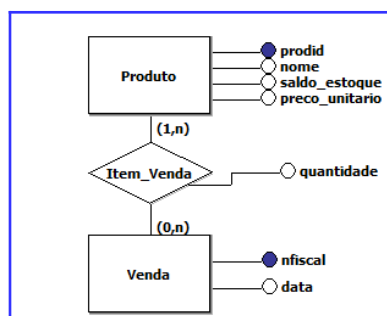
Construa um modelo conceitual (ER e UML) para representar as informações do mesmo tipo das contidas nesse exemplo, inclusive com a relação entre os termos. Considere que cada relação pode ser de um tipo diferente.

Questão 6

Uma empresa controla produtos que estão em seu estoque, bem como o valor de venda unitário de cada um deles. Para fins de simplificação considere que este valor nunca muda.

Estes produtos são vendidos em operações de venda. Cada operação está associada a um número nota fiscal e uma data, incluindo diversos itens de venda. Cada item de venda corresponde à venda de um único produto.

a) Construa um diagrama ER e um UML que represente os produtos e vendas.



b) Transforme o diagrama ER e/ou UML em um esquema relacional.

Produto (ProdId, Nome, SaldoEstoque, PrecoUnitario)
 Venda (Numero, DataVenda)

ItemVenda (NumeroVenda, ProdId, Quantidade)

A tabela Produto contém produtos disponíveis para venda. Cada tupla da tabela Venda registra uma operação de venda de um conjunto de produtos, ela agrupa um conjunto de itens de venda (ItemVenda) que tem chaves estrangeiras para Produto (NomeProduto) e Venda (NumeroVenda).

Depois de ter respondido esta questão, confira o modelo conceitual e esquema na última página e responda os próximos itens.

c) Escreva uma rotina SQL de atualização de preços que aumente um percentual (P%) dos preços unitários, somente para produtos que estiverem entre uma faixa de preços. Para a construção da sentença utilize valores hipotéticos de percentual e faixa de preços.

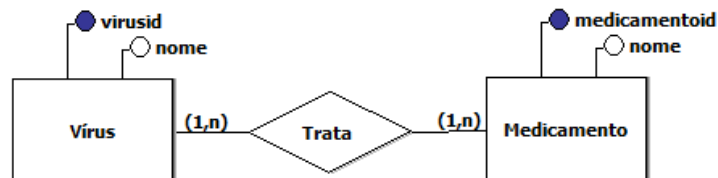
```
UPDATE Produto
SET PrecoUnitario = PrecoUnitario * 1.3
WHERE PrecoUnitario > 50 AND PrecoUnitario < 100
```

d) Escreva uma query SQL que retorne uma relação de vendas agrupada por produto, em que para cada produto deve ser impresso seu nome, a quantidade e valor total de venda do mesmo.

e) Escreva uma query que retorne a última venda de cada produto.

Questão 7

Uma indústria farmacêutica possui um banco de dados que registra os vírus para os quais ela produz medicamentos e os medicamentos que ela produz, cujo modelo ER é definido da seguinte maneira:



O relacionamento Trata indica que um medicamento trata um certo tipo de Vírus.

Duas observações recentes têm mudado a forma como a empresa atua: (i) que a aplicação de vários medicamentos para o mesmo vírus em ordem surtem melhor efeito que um medicamento isolado; (ii) que a ordem de aplicação dos mesmos medicamentos na sequência alteram os resultados, ou seja, uma sequência de medicamentos para um certo tipo de vírus X pode ser mais eficiente que a aplicação da mesma sequência em uma ordem diferente para X.

Para medir a eficiência de um medicamento a indústria definiu um “índice de gravidade” (IG) da doença que varia de 0 (completamente curado) a 10 (pior estado). A aplicação de cada medicamento dentro de uma sequência diminui X pontos (em média) na escala IG (chamado “grau de melhora” - GM).

- Realize modificações e inclusões no modelo ER desta indústria farmacêutica, de modo que se torne possível registrar a eficiência (grau de melhora - GM) da aplicação de sequências de experimentos para tratar um vírus. O controle do GM é feito individualmente por medicamento dentro de uma sequência - note que o GM do mesmo medicamento pode mudar em sequências diferentes. Deve ser possível se registrar sequências diferentes para que se possa compará-las.
- Transforme o modelo ER completo em um esquema relacional.
- Escreva um comando em SQL para criar as tabelas especificadas.
- Escreva uma expressão em álgebra relacional que retorne, para cada sequência de tratamento, o nome vírus tratado, o código da sequência, o nome de cada medicamento

dentro da sequência e o respectivo grau de melhora. Para cada medicamento aplicado em uma sequência deve apresentar o número da ordem de aplicação.

- e) Escreva uma declaração SQL que corresponda a expressão em álgebra de (d).
- f) Defina uma cláusula em (e) que garanta que os medicamentos apareçam pela ordem dentro de uma sequência.
- g) Escreva uma declaração SQL que retorne o total do grau de melhora por sequência de tratamento. Esta query deve mostrar o vírus que foi tratado por sequência e deve nomear os campos apresentados da seguinte maneira: virust, seq, totalmelhora.
- h) Escreva uma query ou sequência de queries (com VIEWS) que apresente no final qual o maior de grau de melhora alcançado para cada vírus, dentre as sequências de tratamento aplicadas (não é necessário apresentar qual a sequência aplicada).
- i) Escreva uma declaração SQL que indique quais as sequências têm dados suspeitos pois o grau de melhora de um de seus medicamentos foi acima de 4. O resultado da query deve também indicar qual o vírus em questão.
- j) Escreva uma declaração SQL que indique quais as sequências têm dados com erro pois a soma dos graus de melhora dos medicamentos é acima de 10. O resultado da query deve também indicar qual o vírus em questão e qual a soma dos graus de melhora.
- k) É possível unificar as queries (i) e (j) em uma única? Se for possível escreva a query unificada, caso contrário justifique porque não é possível.