DI-FCT-NOVA

6 de julho de 2017

Bases de Dados

Exame de Recurso, 2016/17

Consulta limitada

Parte I

(correspondente à matéria do 1º teste)

Grupo I.1

Nota: O enunciado deste grupo é *intencionalmente* vago. Ao responder deve, sempre que necessário, dizer o que assumiu e que acha que não está suficientemente detalhado no enunciado.

A polícia judiciária resolveu organizar a informação sobre os “casos” de futebol de Verão. A sua tarefa neste grupo é propor uma base de dados relacional que permita armazenar toda essa informação.

Mais concretamente, a base de dados deve permitir armazenar informação sobre as várias pessoas envolvidas, bastando de cada uma guardar uma identificação e um nome (ambos obrigatórios). As pessoas podem simpatizar com clubes, sendo que de cada clube a polícia só pretende guardar um código e o nome. Cada pessoa só pode simpatizar, no máximo, com um clube.

Além disso, a polícia quer distinguir entre vários tipos de pessoas: os árbitros, os comentadores desportivos e os dirigentes de instituições de futebol. Dos comentadores a polícia quer saber em que televisão fazem comentários. Só o podem fazer numa televisão, e a televisão tem que ser uma de: RTP, SIC, TVI e CM.

Um dirigente pode pertencer a mais do que uma instituição (por exemplo ser da assembleia da federação, e presidente de uma associação distrital de futebol). Claro que uma instituição pode ter mais do que um dirigente. De cada instituição, a polícia pretende apenas guardar um código e um nome.

Para cada dirigente que pertence a uma instituição, pretende-se saber qual, ou quais, os cargo que o dirigente ocupa nessa instituição, desde quando e, caso tenham terminado, até quando. Se um dirigente ocupou mais do que uma vez um mesmo cargo numa mesma instituição, a polícia apenas quer guardar a última vez.

Por fim, a polícia quer guardar informação sobre os vários processos em curso e as várias evidências associadas aos processos. Cada processo tem um número, um nome e uma descrição. Deve também ser armazenada informação sobre qual a pessoa que fez a denúncia que deu origem ao processo, sempre que haja uma, a data em que o fez, e sobre as várias pessoas que estão envolvidas no processo. Relativamente às evidências, cada uma tem um número, uma descrição, e a data em que foi recolhida. Uma evidência pode estar associada a mais do que um processo.

1. [Exame]

- a) Esboce um diagrama de entidades e relações para a base de dados da polícia.
[Cotação em exame: 2,5 valores; Cotação em substituição de teste: 4,5 valores]
- b) Proponha uma base de dados relacional para armazenar a informação sobre os casos de futebol de Verão. Para cada uma das relações que propuser deve indicar a chave primária, e todas as chaves estrangeiras. Deve ainda indicar qual o domínio de cada um dos atributos. Para definição da base de dados pode usar SQL, ou outra qualquer notação, desde que seja claro quais as relações, chaves, atributos e respetivos domínios. **[2,5; 4,5]**

2. [EXTRA] Nestes casos é comum as evidências referirem-se a pessoas. Claro está que se uma evidência está associada a um processo, só pode referir pessoas que estão envolvidas no processo (se não estavam antes, passam a estar!).

- a) Altere o diagrama de entidades e relações que apresentou para que seja possível armazenar informação sobre pessoas referidas em evidências, conforme descrito acima. Nesta pergunta basta apresentar aquilo que muda e/ou é acrescentado **[0; 1,5]**
- b) Proponha um conjunto de relações para armazenar esta informação (incluindo chaves primárias, candidatas e estrangeiras. Pode usar SQL, ou outra qualquer notação, desde que seja claro quais as relações, chaves, atributos e respetivos domínios. **[0; 1,5]**

Grupo I.2

A polícia pensou melhor sobre a forma como estava a armazenar a informação sobre os cargos que as várias pessoas tinham nas várias instituições, e resolveu criar a seguinte relação:

$$R = (\text{Pessoa}, \text{Instituição}, \text{Cargo}, \text{DataInício}, \text{Nome}, \text{Descrição})$$

e impor as seguintes dependências funcionais F:

Pessoa $\rightarrow$ Pessoa, Nome

Instituição $\rightarrow$ Descrição

Cargo, Pessoa, Instituição $\rightarrow$ DataInício

Pessoa, Instituição $\rightarrow$ Cargo

- a) **[Exame]** Qual (ou quais) a(s) chave(s) candidata(s) de R? **[0,5; 2]**
- b) **[Exame]** Apresente uma cobertura canónica do conjunto de dependências F. **[0,75; 2]**
- c) **[Exame]** Usando o algoritmo dado nas aulas, decomponha sem perdas o esquema R, por forma a obter um conjunto de esquemas na forma normal de Boyce-Codd. **[0,75; 2]**
- d) **[EXTRA]** Qual o significado intuitivo de cada uma das dependências funcionais em F? **[0; 2]**

Parte II

(correspondente à matéria do 2º teste)

Como nos processos dos casos deste Verão há uma série de evidências relacionadas com mensagens de email, a polícia resolveu fazer uma base de dados para armazenar todas essas mensagens. Essa base de dados tem as seguintes relações, onde os atributos sublinhados constituem a chave:

<i>personas</i> (<u>idPessoa</u> , nome, clube)	<i>contas</i> (<u>email</u> , idPessoa)
<i>mensagens</i> (idMens, <u>email</u> , <u>subject</u> , <u>texto</u> , <u>data</u>)	<i>destinatários</i> (idMens, <u>para</u>)
<i>replies</i> (idMens, mensagemOrig)	<i>forwards</i> (idMens, mensagemOrig)
<i>tags</i> (<u>tag</u> , <u>descrição</u>)	<i>classificações</i> (idMens, <u>tag</u>)

A tabela *personas* guarda informação sobre o nome e clube das várias pessoas, sendo que cada pessoa tem um identificador único. A tabela de *contas*, guarda as várias contas de email, tendo para cada uma (para além do endereço), o identificador da pessoa que é dona dessa conta.

As várias mensagens são armazenadas na tabela *mensagens*. Cada mensagem tem um identificador único, o email de quem enviou a mensagem, o subject, o texto e a data em que foi enviada. Os destinatários das várias mensagens são guardados na tabela *destinatários* que, em cada linha, tem um identificador de mensagem e o email de um dos seus destinatários (em *para*).

Sempre que uma mensagem é um reply de uma outra mensagem, é guardado uma linha na tabela *replies*, com o identificador da mensagem (que é o reply) e o identificador da mensagem original a que esta está a responder (em *mensagemOrig*). O mesmo se passa para mensagens que são forward de outras mensagens, neste caso armazenando a informação em *forwards*.

Por fim, as mensagens podem ser classificadas pela polícia com tags (e.g. bruxaria, corrupção, parvoíce, etc). A tabela *tags* guarda todas as tags, com um código (em *tag*) e uma breve descrição. Na tabela *classificações* é armazenada a informação sobre que tags são associadas a que mensagens.

Nesta base de dados, para além das chaves primárias indicadas, estão ainda definidas as seguintes chaves estrangeiras: *idPessoa* em *contas* referencia *idPessoa* em *personas*, *tag* em *classificações* referencia *tag* em *tags* e, em qualquer uma das tabelas, *idMens* referencia *idMens* em *mensagens*.

Grupo II.1

1. Apresente uma expressão em **álgebra relacional e uma consulta SQL** para cada uma das perguntas:
 - a) [Exame] Quais as descrições das tags com que foram classificadas mensagens enviadas por Francisco J Marques? [1; 3]
 - b) [Exame] Quais os nomes das pessoas do FCP que enviaram pelo menos uma vez uma mensagem a uma pessoa do SLB? [1; 3]
 - c) [EXTRA] Quais as mensagens que nunca foram reencaminhadas (forward) para ninguém, e de que endereço foi enviada cada uma dessas mensagens? [0; 3]
2. Apresente **uma consulta SQL** para cada uma das perguntas:
 - a) [Exame] Quantas mensagens mencionam, no texto, o General Nhaga? [0,75; 1,5]
 - b) [Exame] Que pessoas já enviaram mensagens com mais do que 50 destinatários? [0,75; 1,5]
 - c) [EXTRA] Por cada clube, quantas mensagens enviadas por pessoas desse clube foram classificadas pela polícia como sendo parvoíce (i.e. com tag parvoíce)? [0; 1,5]

Grupo II.2

1. A base de dados descrita acima não consegue garantir uma série de condições que é normal serem garantidas em mensagens de email. Neste grupo vamos resolver alguns desses problemas. Imponha restrições de integridade sobre a base de dados acima, para garantir que:
 - a) [Exame] Só se podem enviar mensagens para endereços de contas que existem na base de dados. [0,75; 1,5]

- b) **[Exame]** Só quem foi destinatário de uma mensagem é que pode fazer reply dessa mensagem. **[0,75; 1,5]**
 - c) **[EXTRA]** Os replies a mensagens têm que incluir como destinatário pelo menos quem enviou a mensagem original. **[0; 1,5]**
2. A polícia quer ser avisada de algumas situações que possam ocorrer em mensagens de email. Para isso criou uma tabela de mensagens a inspecionar, com o seguinte esquema:

ainspecionar(idMens)

Crie os mecanismos necessários na base de dados para que esta tabela seja preenchida automaticamente com o identificador da mensagem sempre que:

- a) **[Exame]** Apareça uma nova mensagem que no seu texto menciona o General Nhaga. **[0,75; 1]**
- b) **[Exame]** Apareça uma nova mensagem enviada de pintodacosta@fcp.pt para o lfvieira@slb.pt, ou vice-versa. **[0,75; 1]**

Parte III

(correspondente à matéria do 3º teste)

Considere novamente a base de dados de mensagens de email descrita na Parte II

(Se não respondeu à parte II, leia a introdução dessa parte até ao início das perguntas!).

Grupo III.1

1. Apresente um conjunto de regras em Datalog que permita obter resposta às seguintes perguntas:
 - a) [Exame] Quais as descrições das tags com que foram classificadas mensagens enviadas por Francisco J Marques? [0,75; 1]
 - b) [Exame] Acontece frequentemente uma pessoa enviar uma mensagem, que depois um dos destinatários faz forward para mais umas quantas pessoas, que por sua vez fazem forward para mais pessoas, etc. Todas essas pessoas são “atingidas” pela mensagem original. Quais as pessoas “atingidas” por mensagens enviadas por pedroguerra@tvi.pt? [0,75; 1,25]
 - c) [EXTRA] Quais os clubes das pessoas que tenham endereços de email (na base de dados) para quem pedroguerra@tvi.pt nunca enviou uma mensagem ? [0; 1,25]

Grupo III.2

1. [Exame] Na base de dados descrita na parte II não é garantido que cada mensagem tenha pelo menos um destinatário. Descreva sucintamente porque é que é impossível garantir que não há mensagens sem destinatários, sem se ter alguma noção de transações. [0,75; 1,5]
2. [Exame] Considere as seguintes transações concorrentes (não assuma, para já, nenhuma ordem entre as operações de uma e da outra transação – i.e. não se sabe se o **select** da transação 2 é executado antes do 1º **insert** da transação 1, ou entre o 1º e o 2º **insert**, ou depois do 2º **insert**):

Transação 1	Transação 2
<pre>insert into destinatários values ('m1','a@pt'); insert into destinatários values ('m1','b@pt');</pre>	<pre>select count(*) from destinatários where idMens = 'm1';</pre>

- a) Sabendo que antes da execução destas transações não existe na tabela *destinatários* nenhum tuplo em que *idMens* seja 'm1', que não há mais transações a ser executadas na base de dados para além destas duas, e que ambas são executadas com isolation mode serializable, diga quais os possíveis resultados devolvidos pela operação de **select** da transação 2. [1; 1,5]
- b) Suponha agora que as transações são ambas executadas em isolation mode read uncommitted (i.e. sem nenhum tipo de isolamento). Apresente um escalonamento das 3 operações envolvidas nas duas transações, em que o resultado do **select** seja diferente daqueles que se obtinham no modo serializable, e diga qual esse resultado. [1; 1,5]

Grupo III.3

Para transferir dados de mensagens de email, decidiu-se usar ficheiros XML de acordo com o seguinte DTD:

```
<!DOCTYPE mensagens[
  <!Element mensagens(conta*, mensagem*)>
  <!Element conta(nomePessoa)>
    <!ATTLIST conta email ID #REQUIRED>
  <!Element mensagem(destinatário*,subject,texto,data)>
    <!ATTLIST mensagem de IDREF #REQUIRED>
  <!Element destinatario(#PCDATA)>
    <!ATTLIST destinatario para IDREF #REQUIRED>
  <!Element nomePessoa(#PCDATA)>
  <!Element subject(#PCDATA)>
  <!Element texto(#PCDATA)>
  <!Element data(#PCDATA)>
]>
```

1. **[Exame]** Apresente o texto dum documento XML que use este DTD, e que contenha a seguinte informação: **[1; 2]**
 - Conta do João, com endereço joao@pt; conta da Maria com endereço maria@pt; conta do Pedro, com endereço pedro@pt.
 - Mensagem da Maria para o Pedro e para o João, enviada a 05.07.17, com subject "Olá" e texto "Cá estou".
2. Apresente expressões XPATH sobre um ficheiro de acordo com o DTD acima que devolvam os resultados das seguintes perguntas:
 - a) **[Exame]** Quais os endereços para os quais já foram enviadas mensagens? **[0,5; 1]**
 - b) **[Exame]** Quais os nomes das pessoas que enviaram mensagens no dia 06.07.17? **[0,75; 1]**
 - c) **[EXTRA]** Quais as datas em que foram enviadas mensagem para joao@pt? **[0; 1]**
3. **[EXTRA]** Como terá reparado, o DTD acima não tem nenhuma informação sobre as classificações das mensagens por parte da polícia. A sua tarefa nesta pergunta é alterar o DTD, por forma a que o ficheiro XML tenha informação: **[0; 3]**
 - Sobre as várias tags para classificar as mensagens;
 - Sobre a classificação de cada uma das mensagens (onde uma mensagem pode ter associadas várias tags);
 - Na classificação de mensagens só podem ser usadas tags que existam no próprio ficheiro.

Grupo III.4

Resolveu-se alterar a base de dados apresentada na Parte II usando as *features* objeto-relacional do SGBD. Nomeadamente, pensou-se que seria uma boa ideia ter uma tabela de mensagens semelhante à indicada na Parte II, mas armazenando logo aí o conjunto de todos as tags de classificação de cada mensagem pela polícia.

Ter-se-ia então, em vez das tabelas *mensagens* e *classificações*, uma só tabela, *mensagensClass*, com, para além dos atributos que agora já estão em *mensagens*, um atributo *classifs* com um conjunto de tags com que a mensagem está classificada.

1. **[EXTRA]** Apresente o código SQL completo de criação da nova tabela *mensagensClass*. Não se preocupe, nesta questão, com a imposição de restrições de chaves. **[0; 2]**
2. **[EXTRA]** Considerando a base de dados com esta tabela em vez da originais *mensagens* e *classificações*, apresente o código SQL para a pergunta: **[0; 2]**

"Quais as pessoas que enviaram mensagens classificadas com a tag 'parvoíce'?"