

Trabalho Prático 1

Introdução a Banco de Dados

Data de entrega: 24 de Outubro de 2022

1 Introdução

Este trabalho tem por objetivo a consolidação dos conhecimentos vistos em sala relativos a banco de dados relacionais, principalmente álgebra relacional e linguagem SQL. O trabalho será um contato prático com SGBDs relacionais. Conforme detalhado adiante, o trabalho consiste na elaboração de consultas SQL a partir de expressões em álgebra relacional e linguagem natural das informações que se pretende recuperar. Este trabalho utilizará o módulo VPL (*Virtual Programming Lab*) do Moodle e o SGBD SQLite.

2 Esquema Relacional

O banco de dados a ser utilizado contempla dados históricos das Copas do Mundo entre 1930 e 2014, com um esquema elaborado especialmente para este trabalho, conforme a Figura 1. Arquivos contendo o banco de dados (`database.sql`) e uma descrição detalhada das tabelas e campos (`dicionario.pdf`) podem ser baixados no Moodle. Atenção: as tabelas possuem muitos relacionamentos 1:n e n:n, **lembre-se de considerar os valores distintos**.

ATENÇÃO: Os resultados de algumas consultas podem não coincidir com os dados reais, uma vez que este banco de dados foi simplificado para este trabalho.

3 Especificações das consultas

3.1 Especificações em Álgebra Relacional

Neste bloco, você deverá elaborar consultas SQL equivalentes a expressões em álgebra relacional. Selecione as 5 (cinco) consultas da lista a seguir correspondentes aos 5 primeiros dígitos de seu número de matrícula, prefixados com A. Caso haja algum dígito repetido dentre esses, considere a consulta posterior¹ mais próxima ainda não selecionada. Por exemplo, para o número de matrícula 2015021992, as consultas seriam **A2** (correspondente ao 1º dígito), **A0** (2º), **A1** (3º), **A5** (4º) e **A3** (5º, considerando que as consultas **A0**, **A1**, e **A2** já teriam sido selecionadas)

¹Considere o número 0 como posterior ao 9

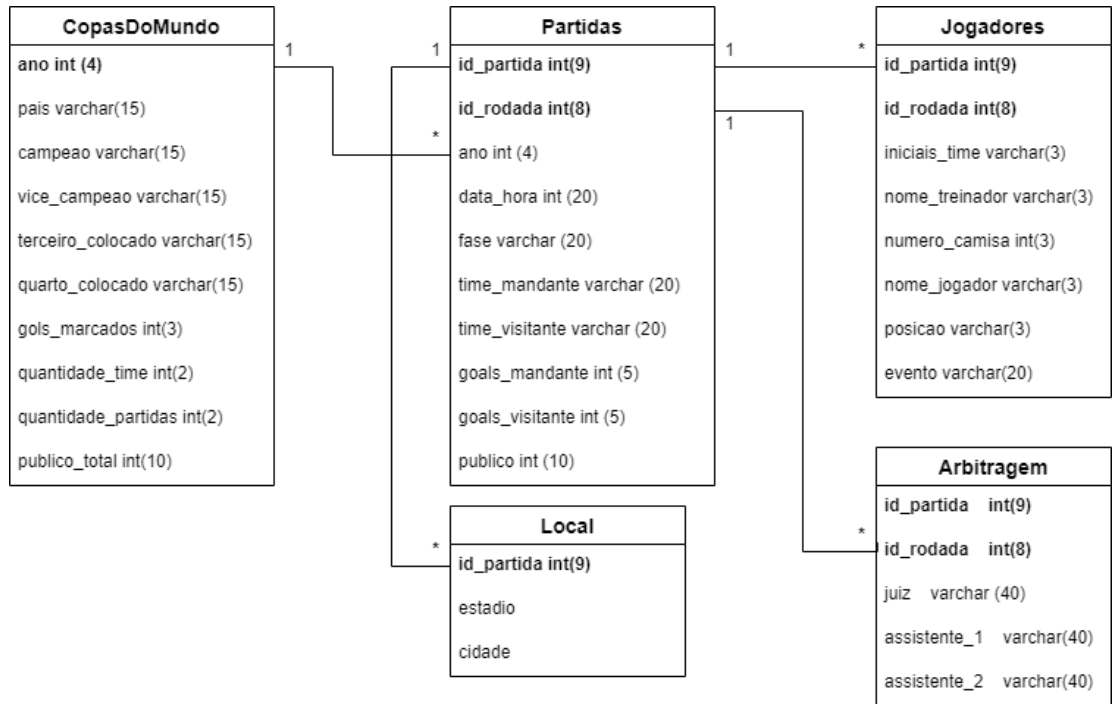


Figura 1: Esquema relacional do banco de dados.

- A0. $\pi_{\text{nome_jogador}} \sigma_{\text{numero_camisa}=14}$ (jogador)
- A1. $\pi_{\text{nome_jogador}} \sigma_{\text{iniciais_time}='ITA' \wedge \text{ano}=1994}$ (jogador \bowtie partidas)
- A2. $\pi_{\text{estadio}} \sigma_{\text{ano}='1958'}$ (local \bowtie partidas)
- A3. $\pi_{\text{nome_treinador}} \sigma_{\text{gols_mandante} \geq 7}$ (jogadores \bowtie partidas)
- A4. $\pi_{\text{time_mandante}} \sigma_{\text{gols_marcados} \geq 150}$ (copasdomundo \bowtie partidas)
- A5. $\pi_{\text{data_hora}, \text{ano}} \sigma_{\text{publico} \geq 80000}$ (partidas)
- A6. $\pi_{\text{data_hora}, \text{ano}} \sigma_{\text{gols_mandante} \geq 5}$ (partidas)
- A7. $\pi_{\text{cidade}} \sigma_{\text{ano}='2002'}$ (local \bowtie partidas)
- A8. $\pi_{\text{estadio}} \sigma_{\text{publico} \geq 90000}$ (local \bowtie partidas)
- A9. $\pi_{\text{nome_completo}} \sigma_{\text{posicao}='C' \wedge \text{ano}=1986}$ (jogadores \bowtie partidas)

3.2 Especificações em Linguagem Natural

Neste bloco, você deverá elaborar consultas SQL equivalentes a expressões em linguagem natural. Selecione as 5 (cinco) consultas da lista a seguir correspondentes aos 5 últimos dígitos de seu número de matrícula, prefixados com B. Caso haja algum dígito repetido dentre esses, considere a consulta posterior²

²Considere o número 0 como posterior ao 9

mais próxima ainda não selecionada. Por exemplo, para o número de matrícula 2015021992, as consultas seriam **B2** (correspondente ao 6º dígito), **B1** (7º), **B9** (8º), **B0** (9º, considerando que a consulta **B9** já teria sido selecionada) e **B3** (10º, considerando que a consulta **B2** já teria sido selecionada).

- B0. Liste (1) o nome dos juízes, (2) assistente um e (3) assistente dois que apitaram a final e semifinal de 2014.
- B1. Liste (1) iniciais do time, (2) o nome do treinador que foram para as quartas de finais em 1970.
- B2. Liste (1) o id_partida, (2) a data e hora e (3) o ano dos jogos finais em todas as copas
- B3. Liste (1) o nome das cidades que sediaram mais de uma edição de Copas do Mundo e (2) seus respectivos países.
- B4. Liste (1) quantas fases diferentes teve a copa que teve um público total maior que 300 mil (2) e qual foi o país campeão.
- B5. Liste (1) o nome das cidades, (2) estádios e (3) o público dos jogos da copa do ano de 2006 em qualquer um dos times fizeram 3 ou mais gols.
- B6. Liste (1) o time mandante e (2) o ano dos países que participaram de apenas uma copa do mundo. Ordene os resultados da edição mais antiga para a mais nova
- B7. Liste (1) o nome dos treinadores, (2) o nome dos juizes dos jogos finais e (3) o ano em que o Brasil foi campeão. Ordene os resultados da edição mais antiga para a mais recente.
- B8. Liste (1) a quantidade de gols totais que o Brasil fez na copa de 2002. Considere os gols como mandantes e como visitantes.
- B9. Liste (1) o nome do juiz (2) a data e hora dos 5 maiores públicos da copa de 2010 e ordenados de maneira decrescente de públicos

4 Instruções para Submissão

Para submissão, as consultas elaboradas deverão ser incluídas em um único arquivo **submission.sql**, conforme template fornecido na tarefa criada para este trabalho prático no Moodle. A seguir, apresentamos um exemplo desse arquivo, considerando o número de matrícula 2015021992 utilizado nos exemplos da Seção 3.

```
-- 2015021992
-- A2
SELECT ... ;
-- A0
SELECT ... ;
-- A1
SELECT ... ;
-- A5
```

```

SELECT ... ;
-- A3
SELECT ... ;
-- B2
SELECT ... ;
-- B1
SELECT ... ;
-- B9
SELECT ... ;
-- B0
SELECT ... ;
-- B3
SELECT ... ;

```

Note que a primeira linha contém um comentário (denotado por `--` no início da linha) incluindo o número de matrícula do aluno. Em seguida, são listadas as 10 consultas elaboradas conforme a Seção 3, cada uma precedida por seu código correspondente (A2, A0, etc.) e encerrada com ponto-e-vírgula (;)³. O arquivo **submission.sql** completo deverá ser submetido via Moodle VPL para avaliação automática.

IMPORTANTE: Antes de submeter as consultas para avaliação automática no VPL, é aconselhável que você as teste localmente. Para tanto, você deverá criar e popular localmente o banco de dados a partir do script **database.sql** fornecido. Como preparação para as demais atividades do curso, em particular o próximo trabalho prático, recomendamos que utilize o Google Colab⁴ para desenvolvimento, conforme visto nas aulas.

5 Avaliação

Este trabalho utilizará um modelo de avaliação automática via Moodle VPL. O avaliador segue as seguintes premissas básicas:

1. Serão avaliadas somente as 10 consultas selecionadas para cada aluno conforme seu número de matrícula (vide Seção 3), devidamente identificadas conforme o formato de submissão descrito (vide Seção 4)
2. As consultas elaboradas deverão executar sem erro no SQLite (sqlite3).
3. Cada consulta deverá retornar exatamente as colunas definidas e na ordem definida na especificação, conforme determinadas pela operação de projeção na Seção 3.1 ou explicitamente listadas na Seção 3.2.

Respeitadas as premissas, cada consulta será avaliada automaticamente quanto à correspondência do conjunto de resultados retornados R em relação ao conjunto de resultados esperados G . Para medida de correspondência, o avaliador utiliza o índice de Jaccard⁵:

³Isso é fundamental, pois o arquivo será executado como uma lista única de comandos SQL.

⁴<https://colab.research.google.com>

⁵https://en.wikipedia.org/wiki/Jaccard_index

$$J(R, G) = \frac{|R \cap G|}{|R \cup G|}$$

Note que $J(R, G) \in [0, 1]$, o que permite atribuir pontuação parcial a cada uma das 10 consultas submetidas. A nota final do trabalho prático será a soma das notas atribuídas às 10 consultas, totalizando o máximo de 15 pontos.

5.1 Solicitação de Avaliação

Não há limite quanto ao número de submissões permitidas ao longo da duração do trabalho. Entretanto, somente a versão mais recente do arquivo **submission.sql** será avaliada. Para solicitar uma avaliação, o aluno deverá clicar no botão de avaliar disponível no editor do VPL, conforme Figura 2, ou na tela de “Visualizar envios”, conforme Figura 3.

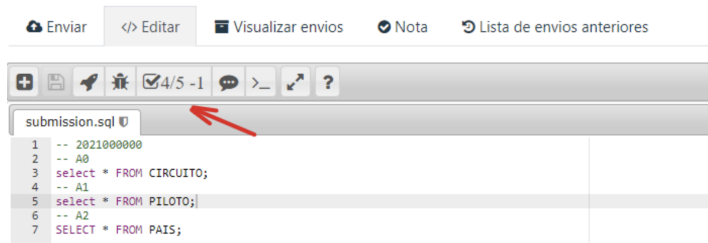


Figura 2: Editor do VPL e botão avaliar

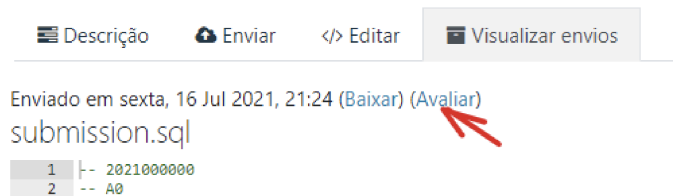


Figura 3: Botão “Avaliar” da tela “Visualizar envios”

Após a execução da avaliação, a nota atribuída, bem como as mensagens de erro/sucesso poderão ser conferidas no painel da direita, no caso do editor do VPL (Figura 4), ou diretamente na tela de “Visualizar envios” (Figura 5). Caso deseje, o aluno poderá utilizar o feedback provido por essa avaliação para melhorar seu trabalho, realizar nova submissão e nova avaliação. **O aluno poderá solicitar até 5 (cinco) avaliações sem desconto na nota.** A partir da 6ª avaliação será descontado 1 (um) ponto da nota final a cada nova solicitação. **A nota final será a nota da última avaliação realizada.**

5.2 Exemplo de Avaliação

A título de exemplo somente, o trecho a seguir mostra a saída produzida pelo avaliador para um arquivo de teste em que somente 4 consultas foram submetidas. A saída foi editada para omitir o texto das consultas, substituídas pelo texto “[OMITIDO]”. As saídas produzidas são explicadas na sequência.

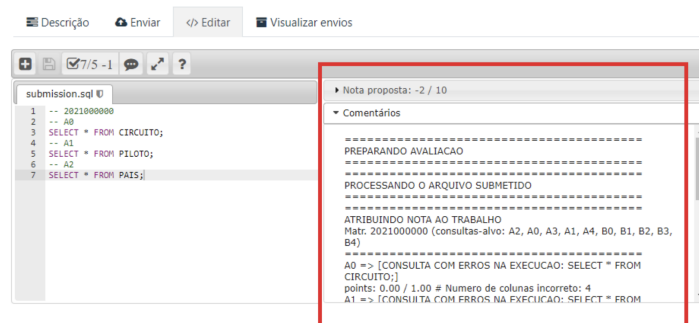


Figura 4: Resultado da avaliação na tela do editor VPL

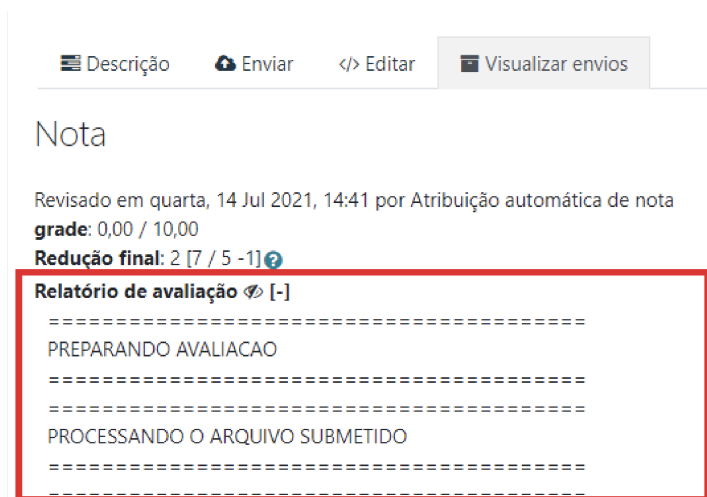


Figura 5: Resultado da avaliação na tela “Visualizar envios”.

```

=====
PREPARANDO AVALIACAO
=====
=====
PROCESSANDO O ARQUIVO SUBMETIDO
=====
=====
ATRIBUINDO NOTA AO TRABALHO
Matr. 2015021992 (consultas-alvo: A2, A0, A1, A5, A3, B2, B1, B9, B0, B3)
=====
A0 => [CONSULTA EXECUTADA COM SUCESSO: [OMITIDO]]
pontos: 0.00 / 1.00
A1 => [CONSULTA EXECUTADA COM SUCESSO: [OMITIDO]]
pontos: 1.00 / 1.00
A2 => [CONSULTA EXECUTADA COM SUCESSO: [OMITIDO]]
pontos: 0.40 / 1.00
A3 => [CONSULTA COM ERROS NA EXECUCAO: [OMITIDO]]

```

```

pontos: 0.00 / 1.00 # Numero de colunas incorreto: 1
A5 => [CONSULTA NAO ENCONTRADA]
pontos: 0.00 / 1.00
B0 => [CONSULTA NAO ENCONTRADA]
pontos: 0.00 / 1.00
B1 => [CONSULTA NAO ENCONTRADA]
pontos: 0.00 / 1.00
B2 => [CONSULTA NAO ENCONTRADA]
pontos: 0.00 / 1.00
B3 => [CONSULTA NAO ENCONTRADA]
pontos: 0.00 / 1.00
B9 => [CONSULTA NAO ENCONTRADA]
pontos: 0.00 / 1.00
TOTAL (antes dos descontos): 1.40 / 10.00

```

A consulta “A0” não levantou erros de execução e tinha as colunas corretas, por isso o avaliador informa sua execução com sucesso. No entanto, como nenhuma linha produzida coincidiu com o esperado, a nota atribuída foi 0 (zero) de 1 ponto possível.

A consulta “A1” não continha erros e retornou o resultado esperado, por isso recebeu nota total de 1 ponto.

A consulta “A2” não continha erros e especificava as colunas corretamente, no entanto o coeficiente de similaridade de Jaccard entre o conjunto das linhas retornadas e esperadas foi 0,4 e a nota atribuída foi proporcional.

A consulta “A3” especificava incorretamente a quantidade de colunas frente ao solicitado e por isso recebeu avaliação 0 (zero).

As consultas “A5” a “B9” não estavam presentes no arquivo.

A nota exibida na última linha é a soma das notas individuais, sem contar eventuais descontos por solicitações adicionais de avaliação. No exemplo, a submissão obteve a nota não descontada de 1,40 de um total de 10 pontos possíveis.

6 Dúvidas

Para dúvidas relacionadas ao TP referente, favor enviar mensagem no fórum de discussão do Moodle, mensagem direta no Moodle ou e-mail para os monitores em alvinolessa@dcc.ufmg.br e pamelacarvalho@dcc.ufmg.br.