

# Trabalho Prático 1

## Introdução a Banco de Dados

Data de entrega: 31 de Janeiro 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 reais de estatísticas de jogos de Beisebol entre 2010-2019, 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**.

### 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 posterior<sup>1</sup> 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)

A0.  $\pi_{\text{NOME\_TIME}}(\text{TIME\_NOMES} \bowtie \text{TIMES} \bowtie \sigma_{\text{ANO}=2000}(\text{TIMES}))$

---

<sup>1</sup>Considere o número 0 como posterior ao 9

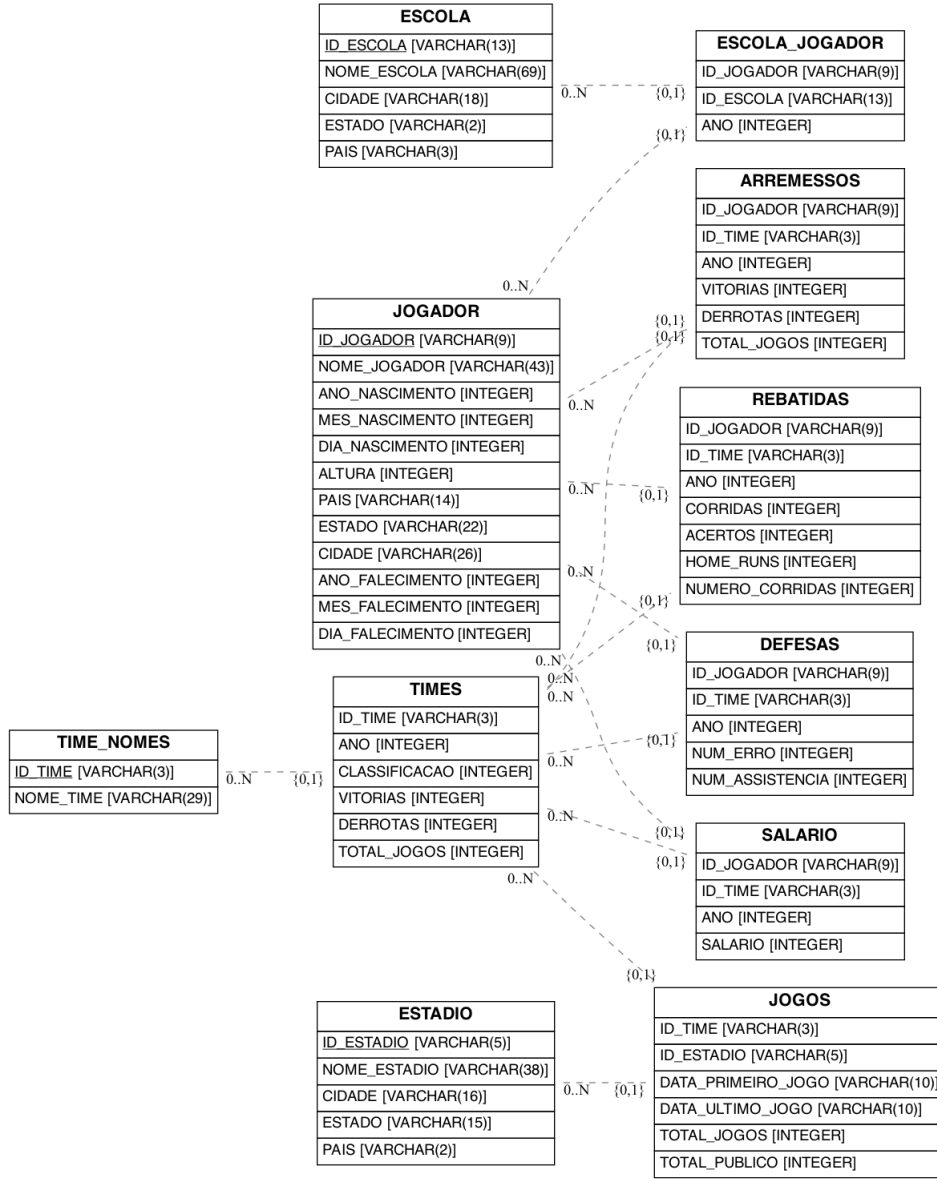


Figura 1: Esquema relacional do banco de dados

- A1.  $\pi_{\text{NOME\_ESTADIO}, \text{NOME\_TIME}} (\text{ESTADIO} \bowtie \text{JOGOS} \bowtie \text{TIMES} \bowtie \text{TIME\_NOMES} \bowtie \sigma_{\text{ESTADO}='CA'}(\text{ESTADIO}))$
- A2.  $\pi_{\text{NOME\_JOGADOR}, \text{NOME\_ESCOLA}} (\text{JOGADOR} \bowtie \text{ESCOLA\_JOGADOR} \bowtie \text{ESCOLA} \bowtie \sigma_{\text{PAIS}='Mexico'}(\text{JOGADOR}))$
- A3.  $\pi_{\text{NOME\_TIME}} (\text{TIMES} \bowtie \text{TIME\_NOMES} \bowtie \text{SALARIO} \bowtie \sigma_{\text{SALARIO} \geq 100000}(\text{SALARIO}) \wedge \sigma_{\text{SALARIO} \leq 200000}(\text{SALARIO}))$
- A4.  $\pi_{\text{ID\_JOGADOR}, \text{ACERTOS}} (\text{REBATIDAS} \bowtie \sigma_{\text{ACERTOS} > 200}(\text{REBATIDAS}))$

- A5.  $\pi_{ID\_TIME, VITORIAS} (TIMES \bowtie \sigma_{VITORIAS > 5}(TIMES) \wedge \sigma_{ANO=2010} (TIMES))$
- A6.  $\pi_{NOME\_JOGADOR, SALARIO} (JOGADOR \bowtie SALARIO \bowtie \sigma_{ANO\_NASCIMENTO=1991}(JOGADOR))$
- A7.  $\pi_{NOME\_TIME, NOME\_ESTADIO} (TIME\_NOMES \bowtie TIMES \bowtie JOGOS \bowtie ESTADIO \bowtie \sigma_{ESTADO='NY'}(ESTADIO) \vee \sigma_{PAIS='MX'}(ESTADIO))$
- A8.  $\pi_{ID\_TIME, ANO} (TIMES \bowtie \sigma_{CLASSIFICACAO=1}(TIMES))$
- A9.  $\pi_{NOME\_JOGADOR, SALARIO} (SALARIO \bowtie TIMES \bowtie JOGADOR \bowtie \sigma_{CLASSIFICACAO=1}(TIMES) \wedge \sigma_{ANO=2016} (SALARIO))$

### 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 posterior<sup>2</sup> 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 jogadores, (2) o Estado e (3) a Cidade da escola dos jogadores que frequentaram uma escola em uma cidade e Estado diferente da cidade e estado natal
- B1. Liste (1) o jogador, (2) qual foi o ano e (3) o total de acertos do jogador que teve a maior soma de acertos de rebatidas
- B2. Liste (1) o nome dos times e (2) o ano em que o time obteve o somatório de vitórias maior que derrotas em arremessos, ordenado pelo maior total de vitórias
- B3. Liste (1) o nome das escolas que já receberam mais de um jogador no mesmo ano
- B4. Liste (1) o nome dos times e (2) seu somatório de erros de defesa de todos os times que em 2015 tiveram mais derrotas que vitórias
- B5. Liste (1) o nome dos times que tiveram o primeiro jogo em Março de 2019
- B6. Liste (1) o nome dos jogadores e (2) a quantidade de salários que eles já tiveram, de todos jogadores que já faleceram
- B7. Liste (1) o id dos jogadores e (2) o nome dos jogadores que tiveram acertos de rebatidas em pelo menos 2 times diferentes
- B8. Liste (1) o nome do estádio que teve o maior número de jogos e (2) qual foi esse total de jogos
- B9. Liste (1) o nome dos jogadores brasileiros que não tiveram nenhum erro nas defesas do ano de 2019

---

<sup>2</sup>Considere o número 0 como posterior ao 9

## 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 (;)<sup>3</sup>. 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 Colab<sup>4</sup> 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:

---

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

<sup>4</sup><https://colab.research.google.com>

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 Jaccard <sup>5</sup>:

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

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 10 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.

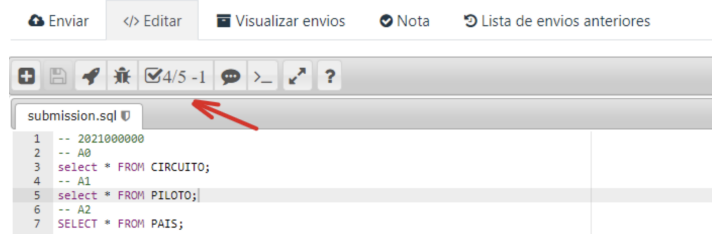


Figura 2: Editor do VPL e botão avaliar

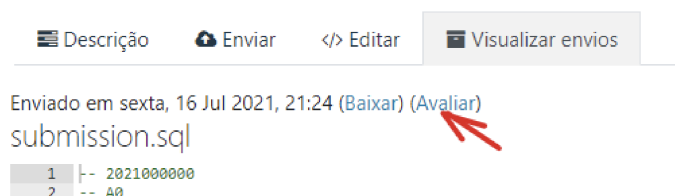


Figura 3: Botão “Avaliar” da tela “Visualizar envios”

<sup>5</sup>[https://en.wikipedia.org/wiki/Jaccard\\_index](https://en.wikipedia.org/wiki/Jaccard_index)

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.**

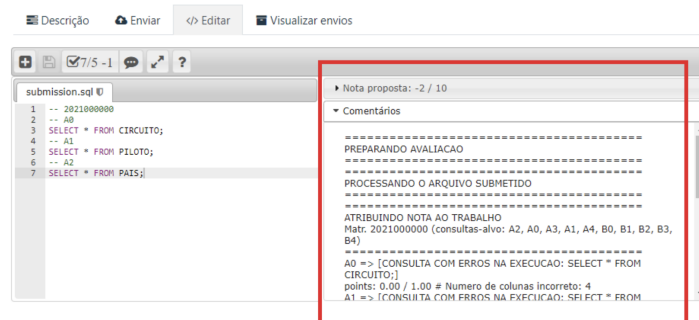


Figura 4: Resultado da avaliação na tela do editor VPL

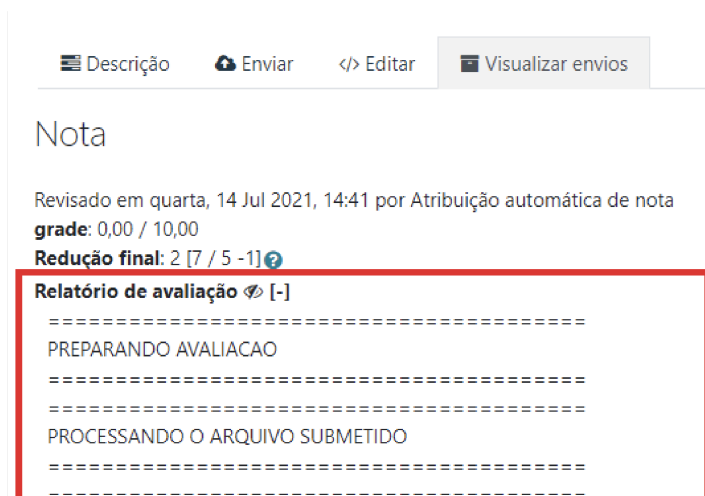


Figura 5: Resultado da avaliação na tela “Visualizar envios”.

## 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.

```

=====
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 ou mensagem direta no Moodle para a monitora Cora Silberschneider.