

Trabalho Prático 1

Introdução a Banco de Dados

Data de Entrega: 27 de maio de 2024

1 Introdução

Este trabalho tem por objetivo a sedimentação dos conhecimentos vistos em sala relativos a bancos de dados relacionais, especialmente álgebra relacional e a linguagem SQL, por meio de 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 geográficas dos países catalogados na ONU (Organização das Nações Unidas), tendo maior foco no Brasil, e casos de dengue e covid de 2024, foi elaborado um esquema especialmente para este trabalho, conforme exibido na Figura 1 ou pelo [modelo online](#). Arquivos contendo o banco de dados (**database.sql**) e uma descrição detalhada das tabelas e campos (**dicionario.pdf**) podem ser baixados no Moodle.

3 Especificação das Consultas

Para submissão, deverão ser elaboradas 10 (dez) consultas na linguagem SQL, sendo 5 (cinco) selecionadas da Seção 3.1 e outras 5 (cinco) da Seção 3.2.

3.1 Especificação 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.

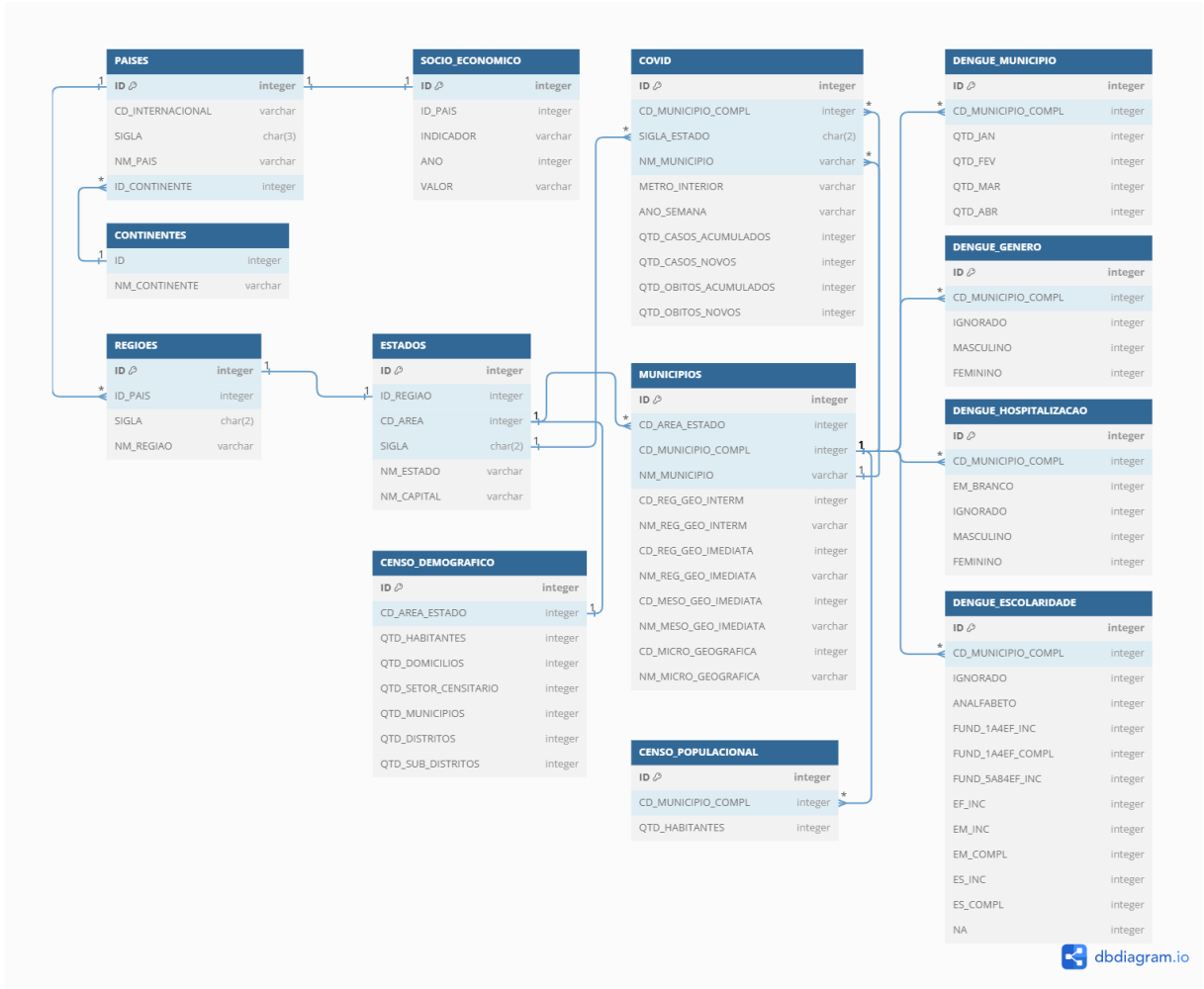


Figure 1: Esquema do banco de dados.

- A0 $\Pi_{NM_PAIS}(\text{CONTINENTES} \bowtie \sigma_{NM_CONTINENTE = 'América do Sul'}(\text{PAISES}))$
- A1 $\Pi_{NM_MUNICIPIO}(\text{MUNICIPIOS}) \cap \Pi_{NM_MUNICIPIO}(\text{DENGUE_MUNICIPIO})$
- A2 $_{NM_ESTADO} \mathcal{F}_{AVG}(QTD_DOMICILIOS)(\text{ESTADOS} \bowtie \text{CENSO_DEMOGRAFICO})$
- A3 $_{NM_MUNICIPIO} \mathcal{F}_{MIN}(QTD_HABITANTES)(\text{MUNICIPIOS} \bowtie \text{CENSO_POPULACIONAL})$
- A4 $\Pi_{NM_MUNICIPIO, NM_ESTADO}(\text{ESTADOS} \bowtie \sigma_{QTD_CASOS_NOVOS \geq 5000 \wedge QTD_OBITOS_NOVOS \geq 50}(\text{COVID}))$
- A5 $\Pi_{NM_REGIAO, NM_ESTADO, NM_CAPITAL}(\text{REGIOES} \bowtie \text{ESTADOS})$
- A6 $\Pi_{NM_MUNICIPIO}(\text{MUNICIPIOS} \bowtie \sigma_{NM_REGIAO = 'Sudeste'}(\text{ESTADOS} \bowtie \text{REGIOES}))$
- A7 $_{NM_ESTADO} \mathcal{F}_{MAX}(QTD_CASOS_ACUMULADO)(\text{ESTADOS} \bowtie \text{COVID})$
- A8 $\Pi_{NM_MUNICIPIO}(\text{MUNICIPIOS} \bowtie \sigma_{QTD_HABITANTES \geq 1000}(\text{CENSO_POPULACIONAL}))$
- A9 $\Pi_{NM_ESTADO}(\text{ESTADOS} \bowtie \sigma_{QTD_MUNICIPIOS \geq 50 \wedge QTD_DISTRITOS \geq 100}(\text{CENSO_DEMOGRAFICO}))$

3.2 Especificação 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 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 Listar (1) quantidade total de municípios brasileiros , (2) quantidade de municípios que tiveram notificações de dengue e (3) quantidade de municípios que tiveram notificações de covid no ano de 2024;
- B1 Listar (1) nomes dos 10 municípios brasileiros com maior quantitativo de notificações de dengue, (2) total de notificações e (3) percentual das notificações de dengue em relação ao total geral (com 2 casas decimais) ordenados do maior para o menor;
- B2 Listar (1) sigla do nome dos países em ordem alfabética, (2) nome dos países e (3) nome do continente;
- B3 Listar (1) nomes dos 7 municípios brasileiros com o maior quantitativo de casos de dengue (mulheres + homens), (2) quantidade de casos de dengue em homens (3) e quantidade de casos de dengue em mulheres no ano de 2024;
- B4 Listar (1) nome dos estados brasileiros em ordem alfabética , (2) percentual de casos de dengue em homens (com duas casas decimais) e (3) percentual de casos de dengue em mulheres (com duas casas decimais) no ano de 2024;
- B5 Listar (1) nomes dos 5 estados brasileiros com maior hospitalização causados pela dengue e (2) gênero mais afetado em cada um dos estados (M ou F) para o ano de 2024;
- B6 Listar (1) nome das regiões brasileiras, (2) total de notificações de dengue por região em ordem decrescente, (3) Analfabeto, (4) 1ª a 4ª Serie Completo, (5) Ensino Médio Completo e (6) Ensino Superior Completo;
- B7 Listar (1) nome dos 10 municípios brasileiros com maior total acumulado de casos de covid, (2) total de notificações, e (3) percentual de casos acumulados de covid em relação ao total geral (com 2 casas decimais) ordenados do maior para o menor;
- B8 Listar (1) nome dos estados (2) e a quantidade de municípios;
- B9 Listar (1) nome dos municípios e (2) nome dos estados brasileiros dos municípios que não tiveram notificação de dengue, ordenado pelo nome dos estados e dos municípios alfabeticamente.

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 ... ;
```

²Considere o número 0 como posterior ao 9.

```
-- 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.

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

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

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.

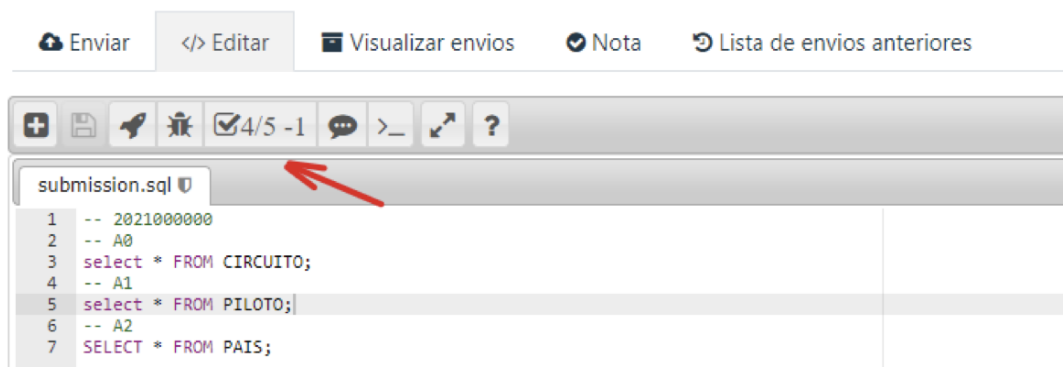


Figure 2: Editor do VPL e botão avaliar.

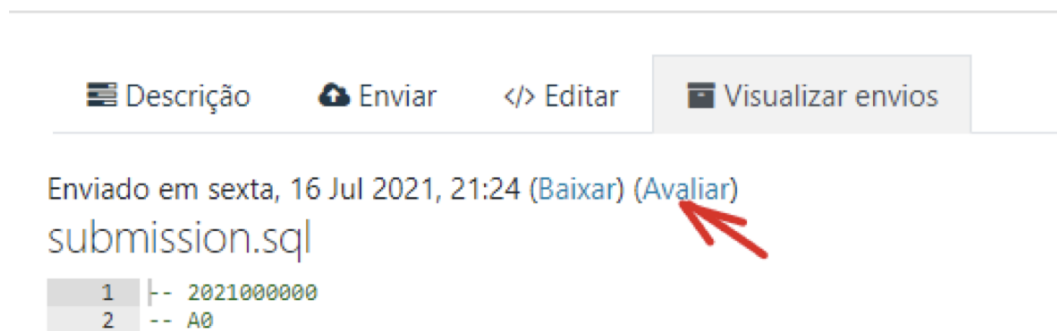


Figure 3: Botão “Avaliar” da tela “Visualizar envios”.

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



Figure 4: Resultado da avaliação na tela do editor VPL.

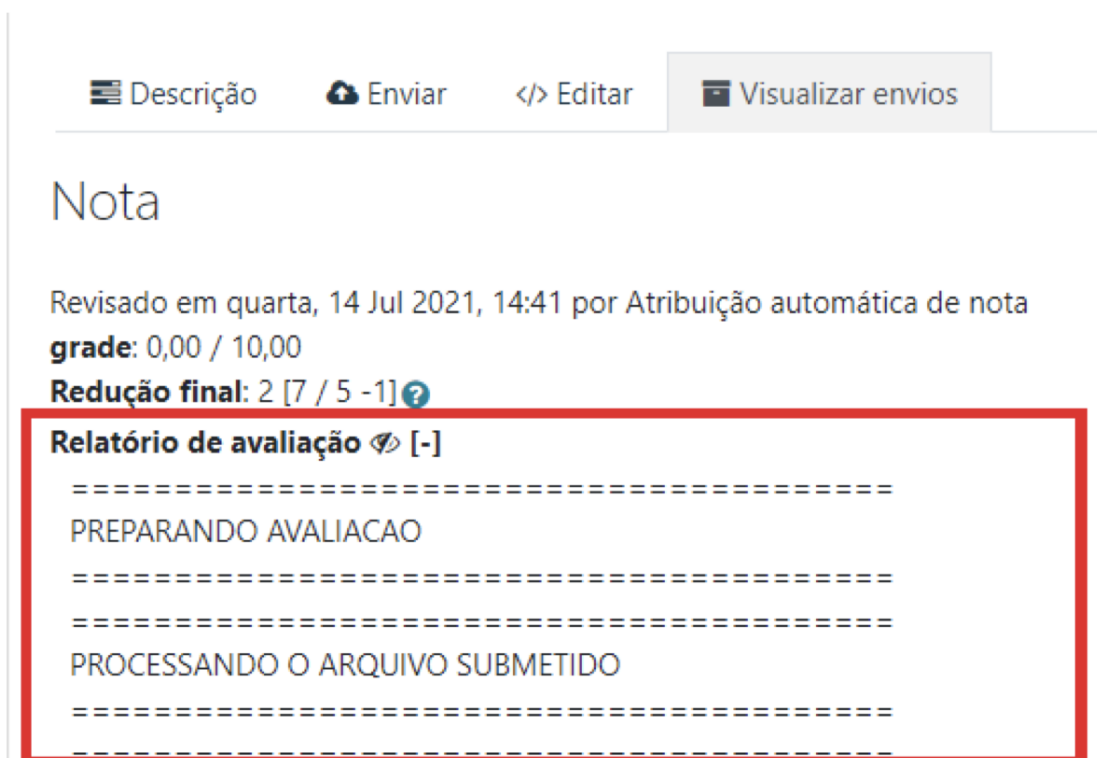


Figure 5: Resultado da avaliação na tela “Visualizar envios”.

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.