

Manipulação de Dados com .CSV e MySQL

Cartilha de boas práticas para tratamento de dados referente à disciplina de 'Tópicos em Gestão'



Instituto Federal do Maranhão - Campus Monte Castelo

Autor: José Carlos Chaves Brasil Junior
Graduando em Sistemas de Informação

Desenvolvido para a disciplina de 'Tópicos em Gestão'
Ministrada pelo professor Abraão Neiver de Miranda Azevedo

São Luís, 2023
Maranhão

SUMÁRIO

1. Introdução.....	3
2. Modelo de banco de dados.....	4
3. Formulários e tratamento de planilhas de dados.....	5
3.1 Tratando planilhas.....	5
3.2 Exportando os arquivos '.csv'.....	8
4. Importando dados.....	10
5. Manipulando dados.....	13
6. Contribua com esse guia.....	14

1. Introdução

Esta cartilha tem como objetivo padronizar alguns procedimentos referentes à confecção de pesquisas que serão realizadas em sala de aula, tal como pontuar as condições mínimas que devem oferecer para que sua utilidade seja extraída ao máximo, além de propor um guia aos alunos, principalmente, sobre manipulação de arquivos **‘.csv’** em conjunto com a plataforma **MySQL Workbench**.

É válido ressaltar que este guia — ou cartilha, como preferir — foi estruturado com base em experimentos realizados em sala de aula, principalmente. Sendo assim, está aberta a mudanças e espera-se que mude, visto que alguns procedimentos apresentados até então são realizados sem automação alguma(*); procedimentos esses que podem, supostamente, ser atribuídos a algum script de preferência.

Como essas questões, em conjunto com a utilização de recursos avançados como *‘procedures’*, *‘triggers’* e funções, além de estruturas lógicas através da utilização de linguagens de programação, não estão em foco, **vamos nos ater apenas à migração de dados entre planilhas e banco de dados de maneira equânime.**

Espera-se que aluno possua conhecimentos em e esteja habituado com:

- **Google Forms**
- **Google Sheets**
- **MySQL Workbench e banco de dados relacional.**

E que já possua:

- **Pesquisa realizada através do Google Forms**
- **Modelo de banco de dados** (opcional, disponível em [GitHub](https://github.com/jjuniorbrasil/topicosemgestaobd): github.com/jjuniorbrasil/topicosemgestaobd)

2. Modelo de banco de dados

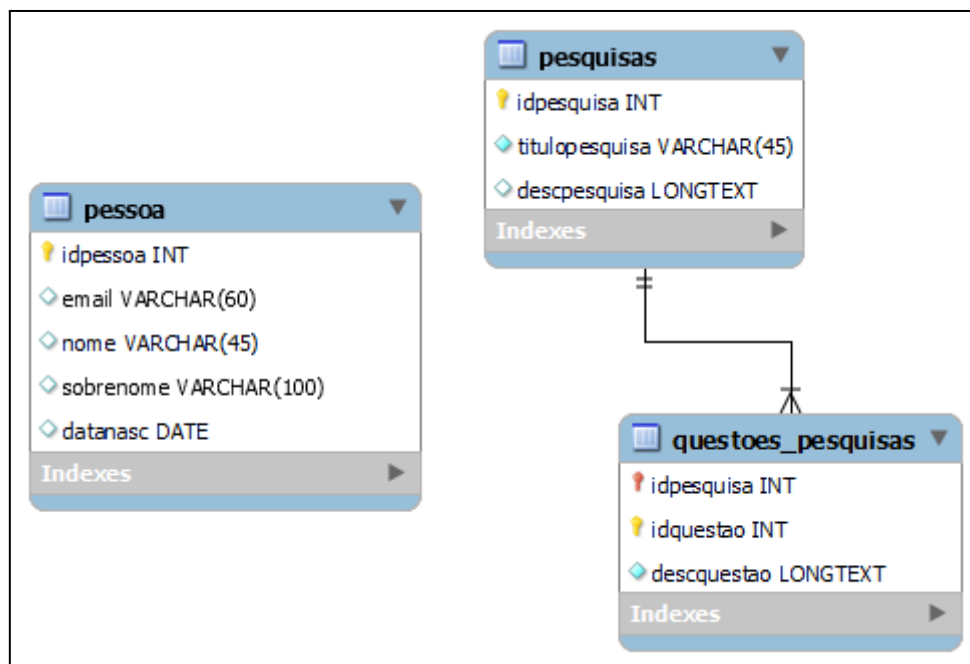


Figura 2.1 - Modelo simplificado, feito no MySQL Workbench

É válido (muito válido, inclusive) ter em mente que, além dos dados obtidos pelas pesquisas, é importante que haja, no próprio banco de dados, dados sobre as próprias pesquisas. Chamaremos, em vários momentos e sem rigor, esses mesmos dados de **‘metadados da pesquisa’**.

Isso é necessário pela dificuldade que há em manipular pesquisas, importadas como grandes planilhas, sem um tratamento prévio que **retire a dependência das respostas vinculadas diretamente às perguntas**. Imagine, por exemplo (que será reiterado adiante), a necessidade de digitar uma *string* inteira para se referir a uma pergunta em toda consulta SQL que pensar em fazer. Além de ser inviável, muitas vezes o próprio MySQL Workbench não conseguirá interpretar colunas não tratadas (ver o próximo capítulo).

De maneira simples: na tabela *pessoa*, irão se armazenar dados sobre indivíduos pesquisados. É uma tabela que considero importante, visto que correlacionar dados a longo prazo, em um mesmo banco de dados, viria a se tornar mais fácil. Na tabela *‘pesquisa’* se registrarão as próprias pesquisas, cada qual com seu *‘id’* único, nome da tabela de acordo com *‘titulopesquisa’* e breve descrição armazenada em *‘descpesquisa’*. A tabela *‘questoes_pesquisas’* armazenará apenas questões.

Importante: Até aqui, ainda não criamos nenhuma estrutura para armazenar dados coletados. Isso será feito mais a frente.

3. Formulários e tratamento de planilhas de dados

Talvez este seja o tópico mais importante deste guia, visto que há abordagens diversas quando o escopo é maior do que a simples manipulação em um banco de dados. É de se esperar que, caso tenha pulado direto para este capítulo, já possua uma pesquisa realizada através da plataforma **Forms**, do Google, com campos que podem ser utilizados como '**chave primária**' dentro do MySQL (por exemplo, **exigir e-mail ou matrícula** dentro do questionário). Optamos por essa abordagem devido à facilidade que é trabalhar com o ambiente Google e suas aplicações, tais quais o **Google Drive** e o **Google Sheets** (esse último em especial) e pela necessidade de utilizarmos uma chave primária para correlacionar diferentes pesquisas.

3.1 Tratando planilhas

Arquivos '**.csv**', quando importados como tabelas para o **MySQL Workbench**, funcionam como entidades. Todo arquivo '**.csv**' terá um **título**, **colunas** com respectivos nomes e **valores**, que são todas as tuplas com exceção da primeira. Sendo assim, precisam ser previamente tratados antes de serem exportados pelo Google Sheets.

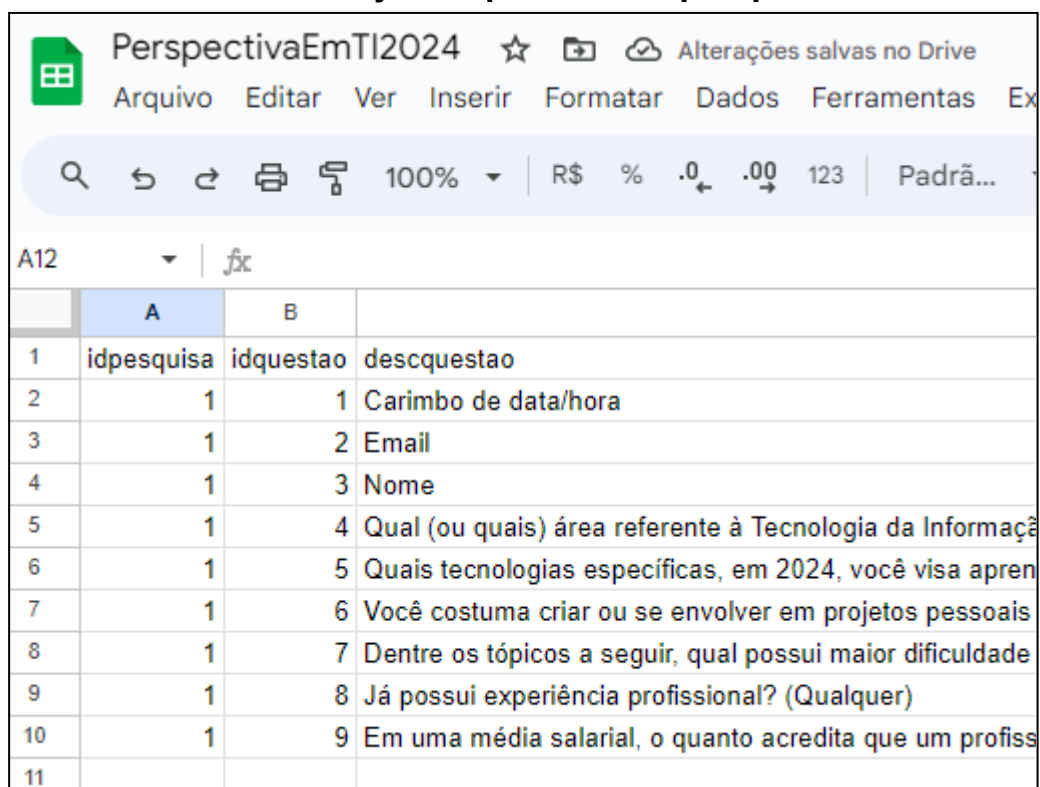
Tendo realizado a pesquisa pelo Google Forms, garanta que:

1. Os dados da pesquisa sejam exportadas para uma planilha

	A	B	C	D	E	F
1	Carimbo de data	Email	Nome	Qual (ou quais)	Quais tecnologias	Você costuma cr
2	07/03/2024 18:0	[Redacted]	[Redacted]	Desenvolvimento	HTML, CSS, Jav	Sim, possuo/ten
3	07/03/2024 18:0	[Redacted]	[Redacted]	Desenvolvimento	JavaScript	Não, programo v
4	07/03/2024 18:0	[Redacted]	[Redacted]	Desenvolvimento	React, TailwindC	Sim, possuo/ten
5	07/03/2024 18:0	[Redacted]	[Redacted]	Desenvolvimento	SQL, Python par	Sim, possuo/ten
6	07/03/2024 18:0	[Redacted]	[Redacted]	Desenvolvimento	Python, JavaScri	Sim, possuo/ten
7	07/03/2024 18:2	[Redacted]	[Redacted]	Desenvolvimento	back-end, Segu	Não, programo v
8	07/03/2024 18:2	[Redacted]	[Redacted]	Desenvolvimento	Desenvolvimento	Sim, possuo/ten
9	07/03/2024 18:2	[Redacted]	[Redacted]	Desenvolvimento	front-end (web	Sim, possuo/ten
10	12/03/2024 17:1	[Redacted]	[Redacted]	Segurança da Informação		Sim, possuo/ten
11	12/03/2024 18:0	[Redacted]	[Redacted]	Desenvolvimento	Todas	Sim, possuo/ten
12	12/03/2024 18:0	[Redacted]	[Redacted]	Segurança da In	Inteligência artifi	Sim, possuo/ten
13	12/03/2024 18:1	[Redacted]	[Redacted]	Desenvolvimento	react	Sim, possuo/ten

Figura 3.1.1 - Planilha originada do Google Forms

2. Os metadados estejam separados da pesquisa



	A	B	
1	idpesquisa	idquestao	descquestao
2	1	1	Carimbo de data/hora
3	1	2	Email
4	1	3	Nome
5	1	4	Qual (ou quais) área referente à Tecnologia da Informaçã
6	1	5	Quais tecnologias específicas, em 2024, você visa apren
7	1	6	Você costuma criar ou se envolver em projetos pessoais
8	1	7	Dentre os tópicos a seguir, qual possui maior dificuldade
9	1	8	Já possui experiência profissional? (Qualquer)
10	1	9	Em uma média salarial, o quanto acredita que um profiss
11			

Figura 3.1.2 - Modelo simplificado, feito no MySQL Workbench

Aqui, basta que uma nova planilha seja criada de acordo com a tabela 'questoes_pesquisas' (ver Figura 2.1). Perceba que a planilha (Figura 3.1.2) possui as colunas 'idpesquisa', 'idquestao' e 'descquestao'. São os mesmos atributos apresentados no modelo de banco de dados que estamos utilizando. Não é necessário que o nome das colunas e dos atributos sejam os mesmos, mas ambos tornam-se mais fáceis de se identificar.

Posteriormente, utilizaremos o arquivo '.csv' dessa planilha para inserir os *metadados da pesquisa* ao banco de dados.

3. As colunas da primeira planilha sejam renomeadas

	A	B	C	D	E	F	G
1	Carimbo de data	Email	Nome	Qual (ou quais) : Quais tecnologia	Você costuma cri	D	
2	07/03/2024 18:0	fe	ac Mo	Desenvolviment	HTML, CSS, Jav	Sim, possuo/ten	E
3	07/03/2024 18:0	he	@ He	Desenvolviment	JavaScript	Não, programo v	E
4	07/03/2024 18:0	ju	ci Jo	Desenvolviment	React, TailwindC	Sim, possuo/ten	A
5	07/03/2024 18:0	gu	ca Zi	Desenvolviment	SQL, Python par	Sim, possuo/ten	E
6	07/03/2024 18:0	ru	@ Ru	Desenvolviment	Python, JavaScri	Sim, possuo/ten	E
7	07/03/2024 18:2	ra	ca Ra	Desenvolvimento back-end, Segu	Não, programo v	M	
8	07/03/2024 18:2	sa	ac Sa	Desenvolviment	Desenvolviment	Sim, possuo/ten	E
9	07/03/2024 18:2	la	@ La	Desenvolvimento front-end (web	Sim, possuo/ten	E	
10	12/03/2024 17:1	fr	ac Fr	Segurança da Informação	Sim, possuo/ten	E	
11	12/03/2024 18:0	le	il. P	Desenvolviment	Todas	Sim, possuo/ten	E
12	12/03/2024 18:0	pa	ca Pa	Segurança da In	Inteligência artifi	Sim, possuo/ten	E
13	12/03/2024 18:1	j.	ca P	Desenvolviment	react	Sim, possuo/ten	M

Figura 3.1.3 - Planilha originada do Google Forms sem colunas renomeadas

	A	B	C	D	E	F	G
1	q1	q2	q3	q4	q5	q6	q7
2	07/03/2024 18:0	fe	ac Mo	Desenvolviment	HTML, CSS, Jav	Sim, possuo/ten	Estruturas d
3	07/03/2024 18:0	he	@ He	Desenvolviment	JavaScript	Não, programo v	Estruturas d
4	07/03/2024 18:0	ju	ci Jo	Desenvolviment	React, TailwindC	Sim, possuo/ten	Algoritmos
5	07/03/2024 18:0	gu	ca Zi	Desenvolviment	SQL, Python par	Sim, possuo/ten	Estruturas d
6	07/03/2024 18:0	ru	@ Ru	Desenvolviment	Python, JavaScri	Sim, possuo/ten	Estruturas d
7	07/03/2024 18:2	ra	ca Ra	Desenvolvimento back-end, Segu	Não, programo v	Matemática	
8	07/03/2024 18:2	sa	ac Sa	Desenvolviment	Desenvolviment	Sim, possuo/ten	Estruturas d
9	07/03/2024 18:2	la	@ La	Desenvolvimento front-end (web	Sim, possuo/ten	Estruturas d	
10	12/03/2024 17:1	fr	ac Fr	Segurança da Informação	Sim, possuo/ten	Estruturas d	
11	12/03/2024 18:0	le	il. P	Desenvolviment	Todas	Sim, possuo/ten	Estruturas d
12	12/03/2024 18:0	pa	ca Pa	Segurança da In	Inteligência artifi	Sim, possuo/ten	Estruturas d
13	12/03/2024 18:1	j.	ca P	Desenvolviment	react	Sim, possuo/ten	Matemática

Figura 3.1.4 - Planilha originada do Google Forms com colunas renomeadas

Com os metadados da pesquisa salvos, podemos renomear as colunas cujo nomes se referem às questões da maneira que quisermos. Nesse caso, optei por usar 'q1' até 'qn' de maneira igual para todas as pesquisas. É importante utilizar **identificadores** que tenham

algum significado, como a escolha de nome para uma variável. Esses identificadores serão utilizados posteriormente nas consultas SQL.

3.2 Exportando os arquivos '.csv'

Tendo garantido que os três passos anteriores foram concluídos, exporte os arquivos '.csv' através de:

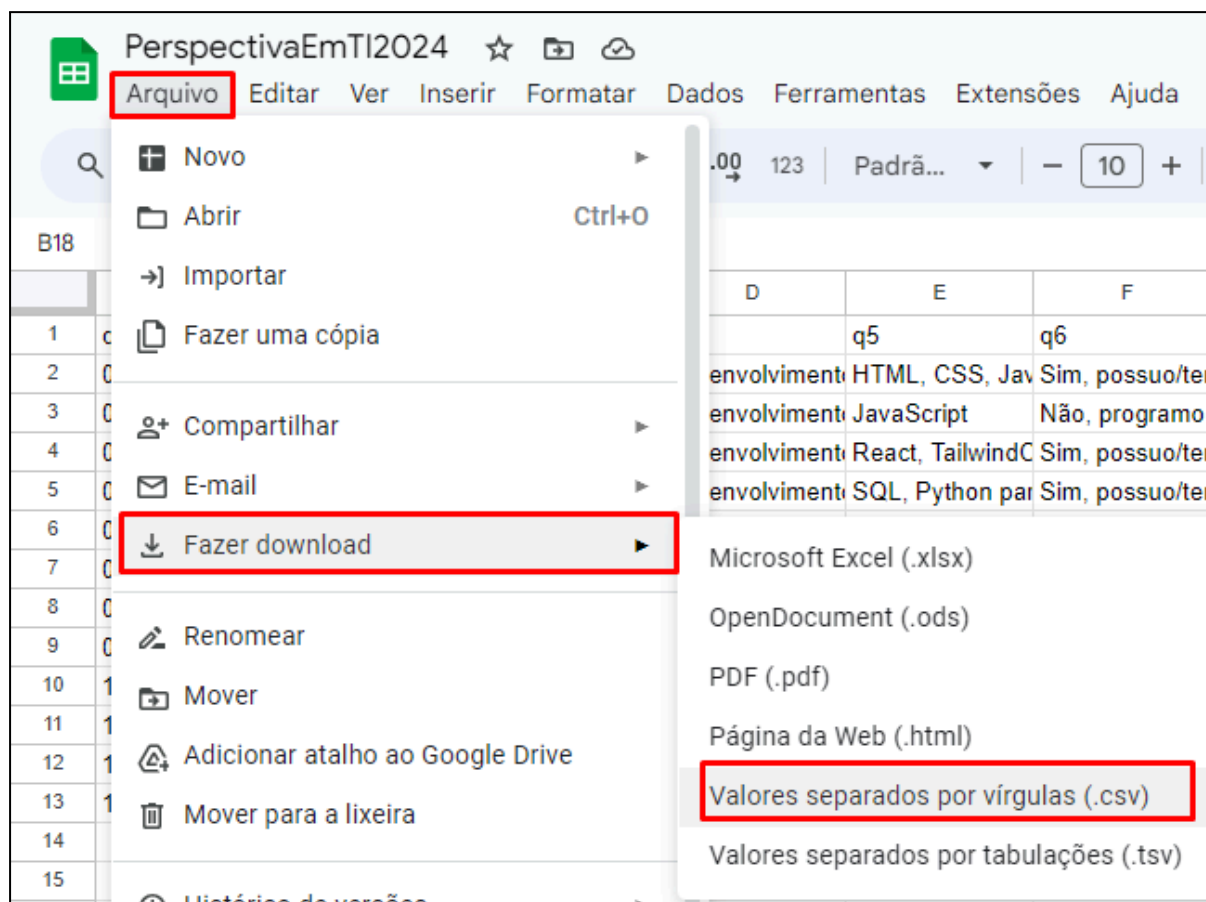


Figura 3.2.1 - Exportação dos dados em arquivos '.csv'

Garanta, ao final, que sejam extraídos dois arquivos: um para os dados obtidos pela pesquisa e outro para os dados da pesquisa.

OBS: Exportamos, até aqui, dados de apenas uma única pesquisa.

4. Importando dados

Importar os dados através de arquivos '.csv', então, se torna mais fácil após o tratamento realizado no capítulo anterior. Tendo o modelo de banco de dados proposto ou qualquer outro, **importe o arquivo '.csv' da pesquisa como uma tabela** ao seu próprio banco de dados através da ferramenta **'Table Data Import Wizard'**:

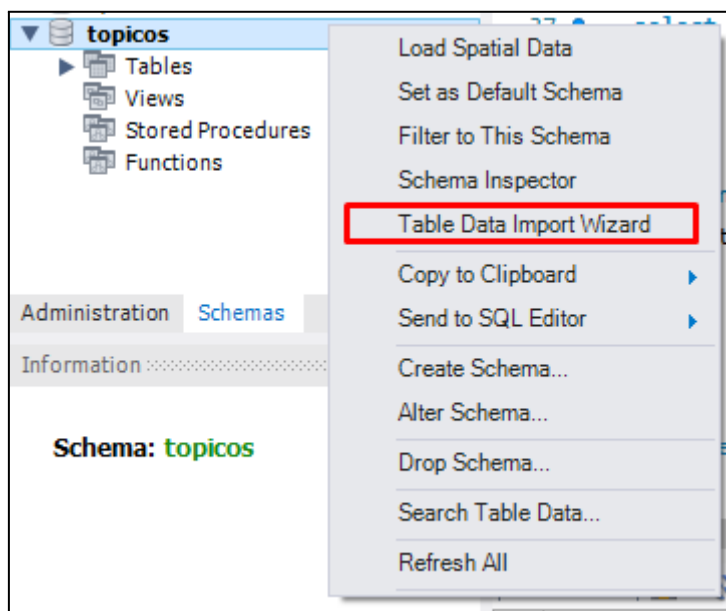


Figura 4.1 - Importação dos dados da pesquisa.

Como dito anteriormente, é esperado que o aluno esteja habituado com a plataforma MySQL Workbench e consiga localizar o seu próprio banco de dados no navegador (aba 'Schemas'). Garanta, então, que seja criada uma nova tabela com a opção **'Create new table'** marcada. Dessa maneira, as pesquisas ficam separadas umas das outras:

 A screenshot of the 'Select destination table and additional options' dialog box. It contains three radio buttons: 'Use existing table:' (unselected), 'Create new table:' (selected), and 'Drop table if exists' (unselected). The 'Use existing table:' option has a dropdown menu showing 'topicos.pesqfran'. The 'Create new table:' option has two input fields: the first contains 'topicos' and the second contains 'pesqjj'.

Figura 4.2 - Opções de tabela

Caso seja necessário, faça configurações em relação à maneira que os dados serão importados. Após isso, prossiga:

Configure Import Settings

Detected file format: csv

Encoding:

Columns:

<input checked="" type="checkbox"/>	Source Column	Dest Column
<input checked="" type="checkbox"/>	q1	<input type="text" value="q1"/>
<input checked="" type="checkbox"/>	q2	<input type="text" value="q2"/>
<input checked="" type="checkbox"/>	q3	<input type="text" value="q3"/>
<input checked="" type="checkbox"/>	q4	<input type="text" value="q4"/>
<input checked="" type="checkbox"/>	q5	<input type="text" value="q5"/>
<input checked="" type="checkbox"/>	q6	<input type="text" value="q6"/>

< q1 q2 q3 q4 q5

Figura 4.3 - Opções de importação

Detected file format: csv

Options:

Field Separator

Line Separator

Enclose Strings in

null and NULL word as SQL keyword

Figura 4.4 - Opções de importação (após clicar na chave)

De acordo com a figura 4.4, garanta que pelo menos as opções de 'Field Separator' e 'Enclose Strings in' estejam de acordo com a maneira que seu arquivo .csv está formatado (você pode abri-lo pelo Visual Studio Code ou pelo bloco de notas, para conferir). Por padrão, o Google Sheets exporta os dados separados por vírgula e entre aspas duplas.

Até aqui, parte essencial do processo está feito: **a pesquisa foi realizada e os dados obtidos já estão no banco de dados**. Entretanto, **ainda nos resta inserir os metadados da pesquisa ao banco de dados** a partir do outro arquivo '.csv' exportado. O procedimento é semelhante: ao invés de inserir diretamente ao banco, esses dados serão inseridos à tabela 'questões_pesquisas', após registrarmos a pesquisa. Isso pode ser feito de acordo com o exemplo abaixo, no editor SQL:

```

23      -- PESQUISA JOSÉ JÚNIOR
24
25      -- 1) Registrando a pesquisa no BD (MANUAL)
26 •   insert into pesquisas values (1, 'pesqjj', 'Pesquisa do aluno José Júnior');
27      -- 2) Metadados da pesquisa (INSERIDOS POR ARQUIVO .CSV)
28 •   select idpesquisa, idquestao, descquestao from questoes_pesquisas where idpesquisa = 1;
29      -- 3) Dados da pesquisa (INSERIDOS POR ARQUIVO .CSV)
30 •   select * from pesqjj;

```

Figura 4.5 - Comandos para registrar uma pesquisa

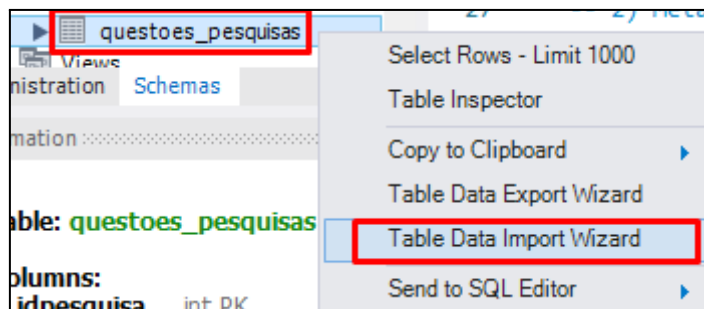


Figura 4.6 - Importação de metadados da pesquisa

De acordo com os passos da Figura 4.5:

1. Registre a pesquisa que foi adicionada através do comando **INSERT INTO**.
 - a. Importe o arquivo '.csv' contendo os metadados da pesquisa diretamente à tabela 'questoes_pesquisas' (Figura 4.6). O 'id' registrado precisa ser o mesmo do 'idpesquisa' dentro do arquivo '.csv'.
2. Confira se os metadados da pesquisa foram importados com sucesso.
3. Confira os próprios dados obtidos pela pesquisa.

Feito isso, já é possível manipular os dados da pesquisa e saber a qual questão cada coluna da pesquisa se refere, através dos metadados da pesquisa.*

*: Existe a opção de adicionar comentários às colunas, evitando a necessidade de uma tabela 'questões_pesquisas' para armazenar perguntas, de maneira exclusiva. Cabe a você optar pela maneira que armazena esse dado, de acordo com a necessidade. De qualquer maneira, não o perca. É imprescindível que estejam no banco de dados.

	idpesquisa	idquestao	descquestao
▶	1	1	Carimbo de data/hora
	1	2	Nome de usuário
	1	3	Qual o seu curso de T.I?
	1	4	Qual sua instituição de ensino?
	1	5	Em qual semestre você está?
	1	6	Por quais áreas de T.I você se interessa?
	1	7	Quais são as habilidades que você considera mais importantes para a sua área de interesse?
	1	8	Qual o seu nível de satisfação com o seu curso?

< questionnaires_pesquisas 8 x

Figura 4.7 - Metadados da pesquisa

	q1	q2	q3	q4	q5
	07/03/2024 18:01:50	M. Desenvolvimento back-end, Desenvolvimento M...	HTML, CSS, JavaScript	Sim, pos	
	07/03/2024 18:21:31	R.. Desenvolvimento back-end, Segurança da Inf...		Não, p	
▶	07/03/2024 18:07:02	Z.. Desenvolvimento de Software, Segurança da ...	SQL, Python para análise de dados	Sim, pos	
	07/03/2024 18:03:46	H.. Desenvolvimento front-end (web e afins), Dese...	JavaScript	Não, p	
	12/03/2024 18:16:04	P.. Desenvolvimento front-end (web e afins), Dese...	react	Sim, pos	
	12/03/2024 18:00:38	P.. Desenvolvimento front-end (web e afins), Dese...	Todas	Sim, pos	
	07/03/2024 18:24:00	S.. Desenvolvimento front-end (web e afins), Dese...	Desenvolvimento de redes neurais	Sim, pos	
	07/03/2024 18:08:33	R.. Desenvolvimento front-end (web e afins), Dese...	Python, JavaScript, HTML, CSS, JavaScript, Excel	Sim, pos	

< pesqj 9 x Read Only

Figura 4.8 - Dados obtidos pela pesquisa

5. Manipulando dados

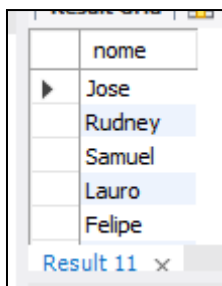
A partir daqui, a manipulação de dados se dá de maneira natural, visto que garantimos uma chave primária, seja qual for, às pesquisas. Por exemplo, a partir das tabelas *pesqjj* e *pesqfran*, que se referem a pesquisas de alunos diferentes, temos, por exemplo:

```

50  -- Nome de pessoas que programam fora de horário de aula/trabalho
51  -- e estão satisfeitas com o curso.
52
53  •  select nome from pessoa as p
54         inner join pesqjj as p1 on p.email = p1.q9
55         inner join pesqfran as p2 on p1.q9 = p2.q2
56  WHERE p1.q5 LIKE 'Sim%' AND p2.q8 > 7;
```

Figura 5.1 - Exemplo de consulta correlacionando pesquisas diferentes

Observe que, com nomes simples para as colunas que contém os dados das pesquisas, como *q1* e semelhantes, declarar uma consulta torna-se extremamente mais fácil. Novamente, a escolha quanto aos nomes das colunas é arbitrário. Caso tenha uma ideia que considere mais eficiente para o seu caso, não hesite em usá-la.



nome
José
Rudney
Samuel
Lauro
Felipe

Figura 5.2 - Resultado da consulta em SQL apresentada na figura 5.1.

Se preferir, utilize a tabela 'pessoa', criada desde o modelo apresentado na Figura 1.1, para guardar dados importantes que podem ser utilizados no futuro, como e-mails, datas de nascimento e até mesmo matrículas em âmbito acadêmico (que podem ser utilizadas como chave primária em outras tabelas, adicionando uma nova coluna)

6. Contribua com esse guia

Como dito anteriormente, todas as informações referentes ao processo de padronização foram coletadas, principalmente, em sala de aula, **a fim de solucionar uma questão específica abordada na disciplina de Tópicos em Gestão**. Sendo assim, é evidente que muitos tópicos aprofundados relacionados a banco de dados relacional, manipulação de arquivos '.csv' e planilhas, além de ferramentas diversas que ainda são desconhecidas por mim, não foram abordados aqui. Cabe a você, aluno, contribuir com esse guia.

Entre em contato através de:

- GitHub: github.com/jjuniorbrasil
- E-mail: junior.brasil@acad.ifma.edu.br

para sugestões e dúvidas.