

DCC011

Introdução a Banco de Dados Definição do Trabalho Final

Mirella M. Moro

Departamento de Ciência da Computação

Universidade Federal de Minas Gerais

mirella@dcc.ufmg.br

INTRODUÇÃO

- O objetivo deste trabalho é projetar e implementar um banco de dados relacional para realizar consultas interessantes.
- O projeto do banco de dados deverá seguir o processo apresentado em aula.
- Deverão ser apresentados
 - Diagrama entidade-relacionamento
 - Esquema relacional
 - Dados inseridos
 - Consultas realizadas



Exemplos
ao final

Características básicas (até 20%)

- Cada grupo deverá escolher entre:
 - i. um conjunto de dados do portal <http://dados.gov.br>
 - ii. um conjunto de dados real que possa ser coletado da Web (e.g, twitter, imdb, github, dblp, etc)
 - iii. uma aplicação atual ou inovadora para a qual dados sintéticos possam ser gerados
- Diagrama ER com requisitos:
 - Pelo menos **4 tipos de entidade**, cada tipo com ao menos **2 atributos** (além de atributo identificador)
 - Pelo menos **3 tipos de relacionamento**, ao menos um com cardinalidade **M:N**
 - Pode ser necessário alterar os dados originais para atender a esses requisitos

Consultas (até 30%)

Deverão ser especificadas e executadas um total de 10 consultas em SQL, sendo:

- 2 operações de seleção e projeção
- 3 junção de duas relações
- 3 junção de três ou mais relações
- 2 funções de agregação sobre o resultado da junção de pelo menos duas relações

Descrição textual + comando SQL (+resultado)

Características avançadas (até 50%)

A segunda metade da avaliação deste trabalho considera **três dessas quatro opções**:

- i. avaliação de eficiência das consultas (cada uma das 10 consultas especificadas é formulada de pelo menos 2 formas)
- ii. dados e consultas mais interessantes, com dados de tipo não-convencional
- iii. mais três consultas do tipo relatório
- iv. interface gráfica (e.g., um website) para visualização *interativa* dos dados

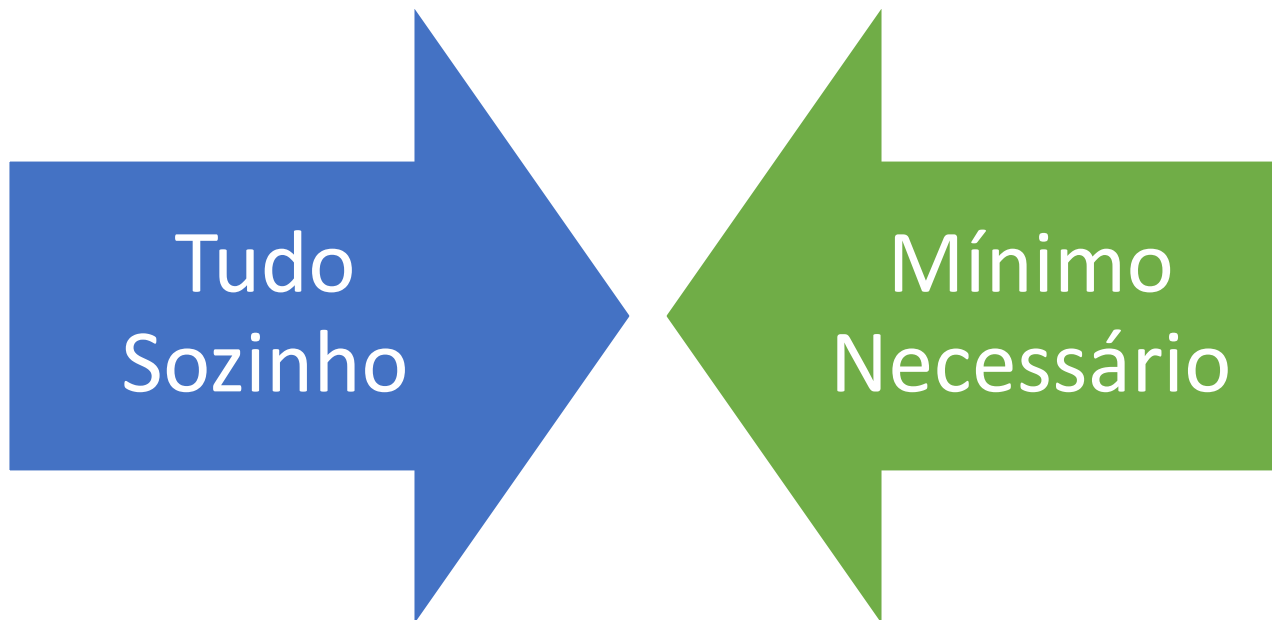
NOTA: pontos extras se o grupo optar por realizar as quatro opções

Observações

- O trabalho deverá ser feito em **grupo de 2 a 4 pessoas, onde uma é líder** (responsável por avaliações individuais dos integrantes do grupo, os quais também a avaliarão) → → →
- DIVIDIR as atividades para evitar desperdício de tempo
- A avaliação é *individual*
- O trabalho é proporcional ao grupo
- SGBD MySQL ou PostgreSQL
- **Avaliação: o relatório final + apresentação**
- Além da funcionalidade básica prevista, a avaliação irá considerar a **criatividade** e a **qualidade** do banco de dados desenvolvido

Sobre a Escolha do Grupo

- Escolher por **expectativa** de nota a ser alcançada
- Escolher por **habilidades** dos integrantes
- Em semestres anteriores



Sobre a Escolha do Grupo

Neste semestre

INOVAR É PRECISO

Orientar a turma pela expectativa da nota

Três casos

- Fazer tudo possível para a nota máxima com extras
- Fazer um bom trabalho
- Fazer o mínimo possível para ser aprovado (por vários motivos, incluindo dificuldade em outras disciplinas)

Habilidades

- IBD dominado
- Criatividade
- IHC

Calendário

30/10: Proposta: grupo e tema (máx. 1 página; via moodle)

A proposta deverá descrever o conjunto de dados escolhido. Em particular, deverão ser descritas as entidades e relacionamentos presentes, e possíveis consultas a serem formuladas sobre esses dados.

26/11: Relatório Final (pelo líder via moodle)

O relatório final deve conter: descrição textual do banco de dados, diagramas ER e Relacional, consultas realizadas com explicação textual e respectivo comando SQL, e a descrição das atividades realizadas individualmente por participante do grupo.

27 e 29/11: Apresentação

Cada grupo terá em torno de 10 min (a confirmar, dependendo do número de grupos) para apresentar o tema do trabalho, a modelagem desenvolvida (conceitual e relacional), uma seleção das várias consultas formuladas e de seus resultados, e demonstração da interface (se tiver sido desenvolvida).

04/12: Última chance para entregar o relatório (com possíveis correções e complementações, e.g., interface).

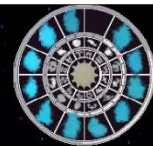
Exemplos de Apresentações

Introdução



Introdução

O Banco



- Foi pensado para servir como um compilado sobre o universo de Cavaleiros do Zodíaco
- As informações que encontrávamos eram sempre muito espalhadas. Queríamos algo mais compacto e direto.
- Os dados tiveram que ser coletados à mão
- Utilizamos uma FANDOM para realizarmos a nossa coleta

INTRODUÇÃO

- Base de dados contém informações sobre os atletas medalhistas das Olimpíadas desde 1896 até 2016;
- Conjunto de dados reais retirado da Internet apresenta várias informações, inclusive de desempenho dos atletas, quando este se aplica.



INTRODUÇÃO DO TEMA

Análise do funcionamento de uma **plataforma de cashback**, inserida no contexto do e-commerce brasileiro;



Introdução

Porque Vídeo Games



Introdução: Dados Reais



Tema: Telefonia no Brasil

Estudo demonstra o que se sabe: telefonia no Brasil é uma das mais caras no mundo

Jornal GGN - Telefonia, no estudo demonstra que trabalhador brasileiro paga mais caro por serviços de telefonia do que em tantos outros países avaliados. Custo alto e um panorama do s

Telefonia móvel perdeu 574 mil linhas em fevereiro aponta a Anatel

Oi tem pior avaliação do consumidor de telefonia do Brasil

A pesquisa englobou as principais operadoras brasileiras de banda larga fixa, telefonia fixa, telefonia móvel e TV por assinatura

Por **Estadão Conteúdo**
© 13 mar 2018, 22h04

Sobre Dados Reais Utilizados

Internet Game Database



- Objetivo de reunir informações relevantes sobre jogos digitais
- Construir aspectos sociais e exploratórios sobre essas informações
- Aproximar as comunidades de jogadores e desenvolvedores

Aplicações Reais

Contexto



- Clínica odontológica situada no bairro Lagoinha, em Venda Nova
- 13 dentistas
- 10 especialidades: Cirurgia, Dentística, Endodontia, Estética, Gerais, Implantodontia, Odontopediatria, Ortodontia, Periodontia, Prótese
- Cada dentista atende uma especialidade



Aplicações Reais: Contribuição Social

Motivação

Utilização de um banco de dados atualizaria o modo de armazenamento atual em asilos garantindo eficiência, segurança, facilidade tanto para acesso quanto para modificação, além da diminuição de espaço físico necessário para armazenar a mesma quantidade de dados.

A possibilidade dos dados serem usados na vida real para ajudar nos asilos.

Visita a dois asilos para saber das dificuldades enfrentadas e necessidades que eles têm para melhorias.



Aplicações Reais: Contribuição Social

Cenário Nacional



Fonte: Sistema de Informação sobre mortalidade -
Ministério da Saúde



Consultas

Consulta simples



nomearm	classificacao	imagem
Fênix	Bronze	
Sagitário	Ouro	
Balron	Sapuris	

Consultas

Consultas

```
SELECT DISTINCT nomeP, ator, habilidade FROM (pp natural join personagem)  
NATURAL JOIN PH  
NATURAL JOIN habilidades,  
(SELECT COUNT(*) as cont, Ator as AA FROM (SELECT DISTINCT nomeP, Ator FROM pp) AS A GROUP BY Ator HAVING Cont>'1') as A  
WHERE pp.ator = A.aa;
```





Tabela de golpes de um determinado Pokemon (selecionado pelo Id, 63 por exemplo), incluindo o level em que aprende cada habilidade e o tipo do golpe

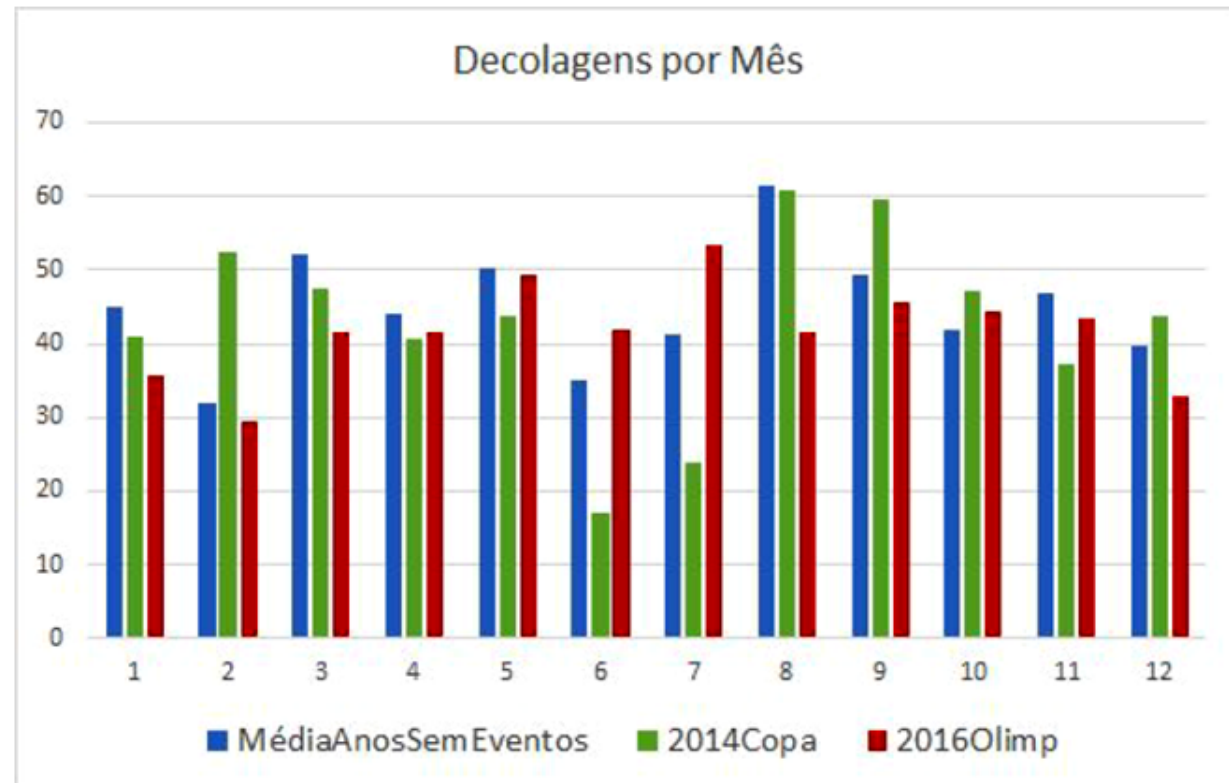
```
SELECT move.*, pokemonMove.level, pokeType.name AS type FROM    pokemon JOIN pokemonMove
ON pokemon.id = pokemonMove.pokeld JOIN move ON pokemonMove.moveId = move.id JOIN
pokeType ON move.typeId = pokeType.id WHERE pokemon.id = 63 ORDER BY pokemonMove.level;
```

id	typeId	name	class	power	accuracy	pp	ailment	description	level	type
271	14	trick	status	0	100	10	none	User and target swap items.	0	psychic
356	14	gravity	status	0	0	5	none	Disables moves and immunities that...	0	psychic
351	13	shock-wave	special	60	0	20	none	Never misses.	0	electric
347	14	calm-mind	status	0	0	20	none	Raises the user's Special Attack an...	0	psychic
324	7	signal-beam	special	75	100	15	confusion	Has a \$effect_chance% chance to c...	0	bug
290	1	secret-power	physical	70	100	20	none	Has a \$effect_chance% chance to i...	0	normal
289	17	snatch	status	0	0	10	none	Steals the target's move, if it's self-t...	0	dark
285	14	skill-swap	status	0	0	10	none	User and target swap abilities.	0	psychic
282	17	knock-off	physical	65	100	20	none	Target drops its held item.	0	dark

Consultas e Gráficos

Média de Passageiros e Decolagens nos anos de 2012, 2013, 2015, 2017 comparadas aos anos de 2014 e 2016

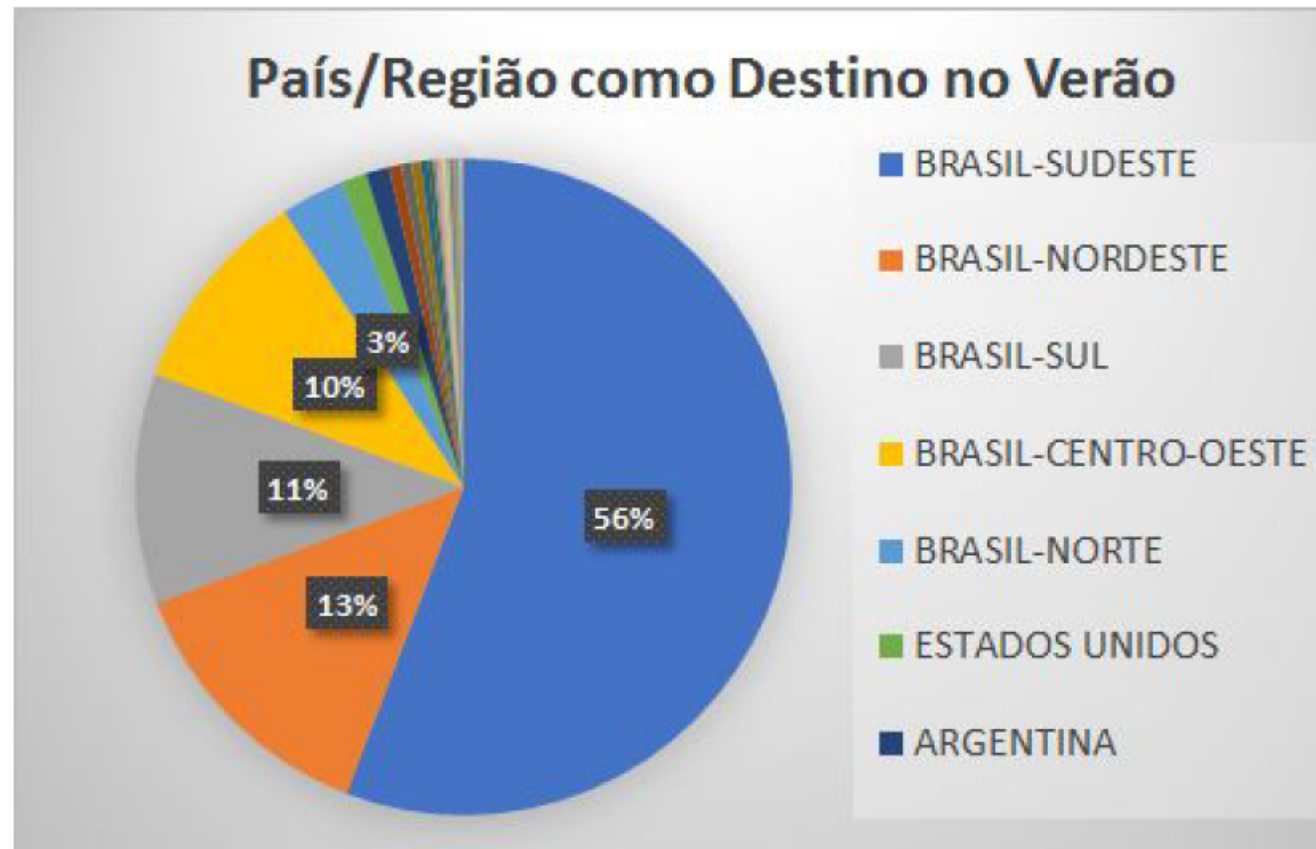
```
SELECT mes, media_pass_sem_evento,  
media_pass_COPA, media_pass_Olimp,  
media_deco_sem_evento, media_deco_COPA,  
media_deco_Olimp  
FROM(((SELECT mes, AVG(pass) AS  
media_pass_sem_evento, AVG(deco)  
AS media_deco_sem_evento  
FROM infovoomensal  
WHERE ano = 2012 OR ano = 2012 OR  
ano = 2013 OR ano = 2015 OR ano =  
2017 GROUP BY mes) AS semevento  
NATURAL JOIN (SELECT mes,  
AVG(pass) AS media_pass_COPA,  
AVG(deco) AS media_deco_COPA  
FROM infovoomensal WHERE ano =  
2014 GROUP BY mes) AS copa)  
NATURAL JOIN (SELECT mes,  
AVG(pass) AS media_pass_Olimp,  
AVG(deco) AS media_deco_Olimp  
FROM infovoomensal  
WHERE ano = 2016  
GROUP BY mes) AS olimp);
```



Consultas e Gráficos

No verão, para qual país/região há mais passageiros?

```
SELECT pais, regioao,  
SUM(pass) AS  
nro_passageiros  
FROM( SELECT aerod, pass  
FROM (infovooaero NATURAL  
JOIN infovooemsa1)  
NATURAL JOIN data  
WHERE estacao = 'verao')  
AS T1  
JOIN aeroportos ON aerod  
= sigla  
GROUP BY pais, regioao  
ORDER BY nro_passageiros  
DESC;
```



Consultas e Resultados

1 Mulheres na Política

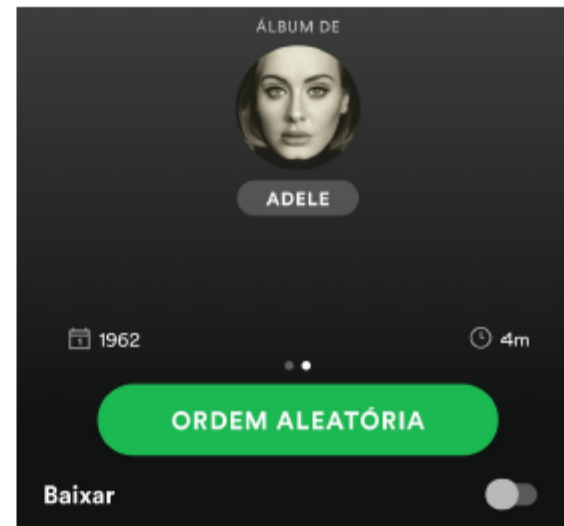
- Partidos com mais deputadas

```
select NomePartido, count(SiglaPartido)
from(partidos2017 join deputados2017 on SiglaPartido =
SiglaPardido)
where Genero = 'F'
Group by SiglaPartido
Order By SiglaPartido;
```

NomePartido	count(SiglaPartido)
Democratas	4
Partido Comunista do Brasil	5
Partido da República	5
Partido da Social Democracia Brasileira	7
Partido Democrático Trabalhista	1
Partido dos Trabalhadores	7
Partido Humanista da Solidariedade	1
Partido Popular Socialista	2
Partido Progressista	1
Partido Republicano Brasileiro	2
Partido Republicano da Ordem Social	2
Partido Social Cristão	2
Partido Social Democrático	1
Partido Socialismo e Liberdade	1
Partido Socialista Brasileiro	2
Partido Trabalhista Brasileiro	1
Partido Verde	1
Podemos	2

BUG DO SPOTIFY

```
mysql> SELECT A.name, Al.name  
-> FROM (Artist A JOIN ArtistAlbum AA ON A.id = AA.artist_id) JOIN Album Al ON AA.album_id = Al.id  
-> WHERE Al.release_date = '1962';  
+-----+-----+  
| name | name |  
+-----+-----+  
| Adele | Skyfall |  
+-----+-----+  
1 row in set (0.08 sec)
```



Exemplos de Interfaces

Três tipos mais frequentes de interface

- (1) Mostrar apenas as consultas que estão no trabalho
- (2) Mostrar as do trabalho + campo para digitar comando SQL
- (3) Interface interativa

(1) Apenas para consultas do trabalho

[Retornar ao Menu](#)[Escolha uma das consultas básicas](#)[Escolha uma das consultas tipo relatório](#)[Dados Não-Convencionais](#)

A consulta escolhida foi:

Mostrar título da palestra, nome e quantidade do rating das top 15 palestras.

Essa consulta é do tipo:

Envolvendo a junção de duas relações.

A consulta retornou:

3 colunas
15 linhas

```
1 SELECT Rating.rating_name, Talk.title, Talk_Rating.count
2 FROM Talk NATURAL JOIN Talk_Rating NATURAL JOIN Rating
3 GROUP BY Rating.rating_id, Talk.talk_id, Talk_Rating.talk_id
4 ORDER BY Talk_Rating.count DESC
5 LIMIT 15;
6
```

rating_name	title	count
Inspiring	Do schools kill creativity?	24924
Inspiring	How great leaders inspire action	21564
Inspiring	The power of vulnerability	21444

Consultas

Nº	Descrição	Ações
01	Nome e escudo dos times do estado de Minas Gerais	Executar
02	Clássicos Internacional X Grêmio e seus placares	Executar
03	Nome de cada campeão e quantos títulos possui	Executar
04	Quantas vezes o Internacional jogou no Mineirão	Executar
05	Times que já disputaram a série B	Executar
06	Maiores pontuações alcançadas nos campeonatos	Executar
07	Partidas com placar "6x1" ocorridas no estado de Minas Gerais	Executar
08	O primeiro jogo que ocorreu em Minas Gerais no ano de 2009	Executar
09	Times ordenados por quantidade de vice-campeonatos	Executar
10	Número total de pontos conquistados por cada time durante todos os campeonatos	Executar

Consultas Relatório

Nº	Descrição	Ações
01	Time que mais ganhou em clássicos CRU x CAM - EXTRA	Executar
02	Número de títulos por estado	Executar
03	Número de times de cada estado que participaram de todas as edições na série A	Executar
04	Jogos realizados na Região Norte do Brasil	Executar

(2) Consultas do trabalho + Espaço para escrever SQL qualquer

← → ⓘ localhost:3000 🔍 spotify bar ☆ 📄 ⬇



Album(id, name, image_url, album_type, release_date, popularity)

Artist(id, name, image_url, followers, popularity)

Track(id, name, duration_ms, explicit, popularity)

User(id, display_name, followers)

Playlist(id, name, owner, description, followers, collaborative, public)
owner referencia User, por bloqueio

AlbumTrack(album_id, track_id, disc_number, track_number)
album_id referencia Album, por propagação
track_id referencia Track, por propagação

ArtistGenre(artist_id, genre)
artist_id referencia Artist, por propagação

ArtistAlbum(artist_id, album_id)
artist_id referencia Artist, por propagação
album_id referencia Album, por propagação

TrackArtist(track_id, artist_id)
track_id referencia Track, por propagação
artist_id referencia Artist, por propagação

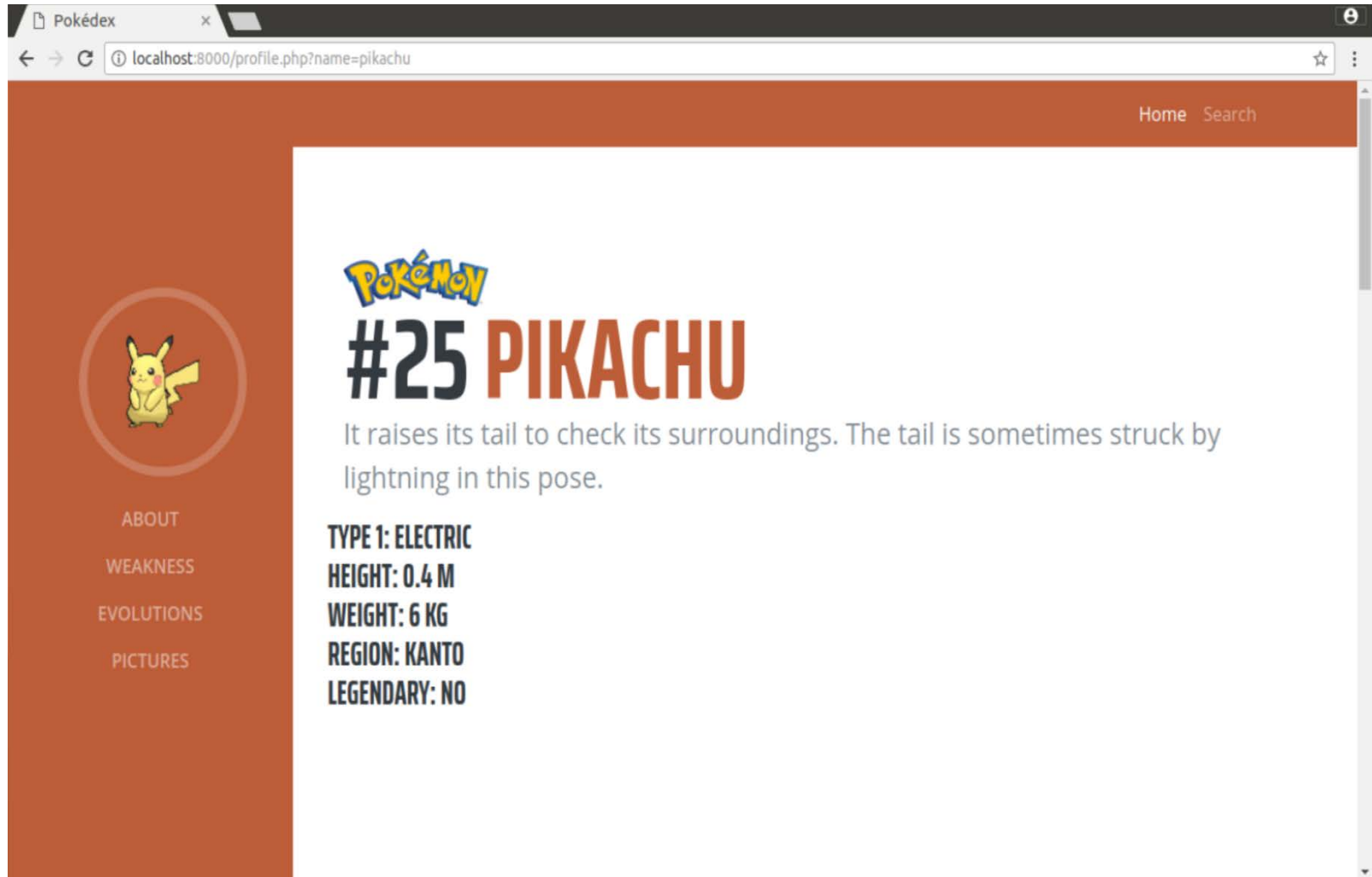
PlaylistTrack(playlist_id, track_id)
playlist_id referencia Playlist, por propagação
track_id referencia Track, por propagação

Selecione uma das consultas abaixo:

--Selecione a consulta-- ▼

Escreva a sua consulta e aperte enter

(3) Consultas interactivas



Pokémon List

Try querying for a pokémon!

Search for pokémon...

Search



1

Bulbasaur

2

Ivysaur

3

Venusaur

4

Charmander

5

Charmeleon

6

Charizard

7

Squirtle

8

Wartortle

9

Blastoise

10

Caterpie

11

Metapod

12

Butterfree

13

Weedle

14

Kakuna

15

Beedrill

16

Pidgey

17

Pidgeotto

18

Pidgeot

19

Rattata

20

Raticate

21

Spearow

22

Fearow

23

Ekans

24

Arbok

Dicas Finais

Dicas Finais

```
SELECT {consulta}  
INTO OUTFILE "{caminho/arquivo.csv}"  
[FIELDS {opções}]
```

Exemplo:

```
mysql> SELECT * FROM cadastro  
> INTO OUTFILE ' /tmp/cadastro.csv '  
> FIELDS TERMINATED BY ';'   
> ENCLOSED BY '"'  
> LINES TERMINATED BY '\n';
```

Dicas Finais

- A partir do CSV, é possível importá-lo facilmente para o Excel, ou outras planilhas
- Também é possível importá-lo para R, por exemplo com os comandos

```
variavel=read.table( "cadastro.csv" ,  
header=T, dec=" , " , sep=" ; " )
```

Ou

```
read.csv2( "cadastro.csv" , sep=" ; " ,  
dec=" , " )
```

Dicas Finais

- Não confundir editor de diagrama ER com o de esquema relacional
- Exemplo de editor de **diagrama ER**

<http://sis4.com/brModelo/>

(com notação do livro do prof Heuser, pode ser utilizado na disciplina sem problemas; este é o único que testei)

- Outras possibilidades (e várias mais na Web)
 - <https://www.lucidchart.com/pages/er-diagrams>
 - <https://erdplus.com>
 - Power Point 😊
 - Gimp 😊

Dicas Finais

O grupo pode criar um gerador de dados sintéticos (artificiais) ou utilizar um existente

Exemplos:

- <https://sourceforge.net/projects/spawner>
- <https://www.generatedata.com>
- <http://www.freedatagenerator.com/sql-data-generator>