

Alunos: Enok Januário da Rocha e Orlando Enrico Liz Silvério Silva

Trabalho Prático 2

Relatório Final

1. Escolha do Tema

Em um primeiro momento o intuito era trabalhar com um grande dataset contendo informações sobre personagens de quadrinhos de diferentes editoras e universos. Porém, optou-se em momento posterior por realizar o projeto em cima somente do Universo Marvel, incluindo também o seu estúdio de filmes. Mesmo assim, é possível traçar aspectos interessantes relacionados aos integrantes desse mundo (heróis, vilões, anti-heróis, personagens secundários, etc.).



2. Datasets

Para a construção do presente trabalho selecionamos 4 datasets, os quais detalharemos a seguir:

Dataset 1 (Personagens): de forma geral, constitui as características dos personagens. Possui identificador, nome, gênero, cor dos olhos, raça, cor do cabelo, altura, cor da pele, alinhamento (do bem, do mal, etc.) e peso.

Dataset 2 (Poderes): os atributos são os poderes mais comuns manifestados pelos personagens (super força, agilidade, fator de cura, durabilidade, capacidade de voar, pontaria, longevidade, inteligência, super velocidade, dentre outros).

Dataset 3 (Filmes): esse dataset contém dados sobre o universo cinematográfico Marvel como nome do filme, ano de lançamento, bilheteria, fase e tempo de tela de cada personagem em cada filme.

Dataset 4 (Avaliações): além do identificador, os atributos são as notas de cada filme de dois portais renomados de crítica (IMDB e Rotten Tomatoes).

3. Modelagem Conceitual

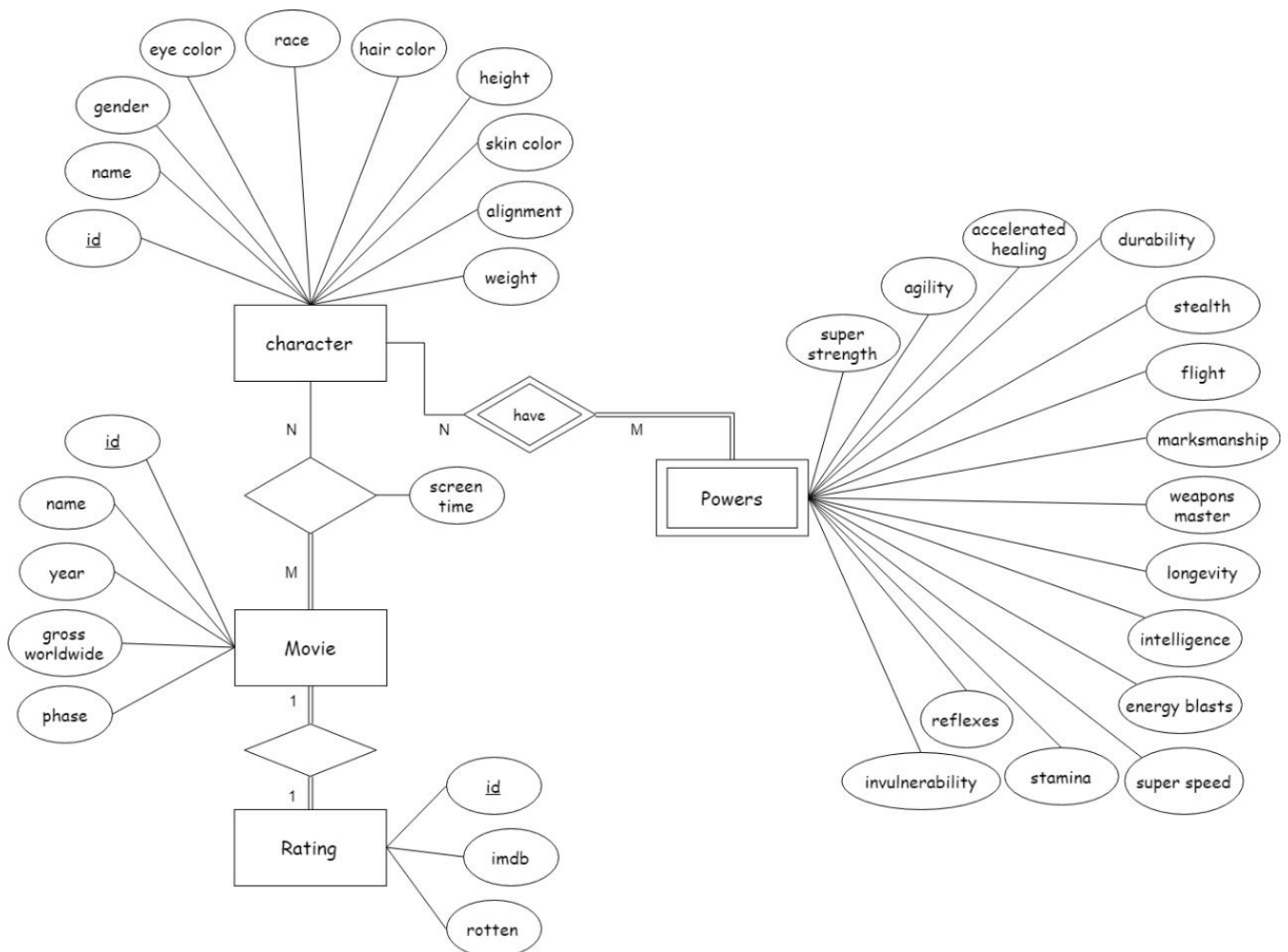


Figura 1: Diagrama ER

Através dos *datasets* mencionados na seção anterior, conseguimos estruturar um banco de dados com 4 entidades (representadas acima na figura 1, diagrama ER):

- **Character:**
- **Movie:**
- **Rating:**
- **Powers:**

Além disso, existem 3 relacionamentos estabelecidos durante a modelagem:

- **Character - Powers:**
- **Character - Movie:**
- **Movie - Rating:**

4. Modelagem Relacional

Character (idCharacter, nameCharacter, gender, eyeColor, race, hairColor, height, skinColor, alignment, weight)

PowersCharacter (characterID, superStrength, agility, acceleratedHealing, durability, stealth, flight, marksmanship, weaponsMaster, longevity, intelligence, energyBlasts, superSpeed, stamina, reflexes, invulnerability)

characterID REFERENCIA Character

Movie (idMovie, nameMovie, year, gross, phase)

CharacterMovie (character, movie, screenTime)

character REFERENCIA Character

movie REFERENCIA Movie

Rating (idRating, imdb, rotten)

MovieRating (movie, rating)

movie REFERENCIA Movie

rating REFERENCIA Rating

5. Consultas

Como proposto na descrição do trabalho, 10 consultas são realizadas com o propósito de aprender um pouco mais sobre o universo Marvel e de certa forma, descobrir certos fatos curiosos. A partir de agora iremos mostrar algumas das consultas realizadas e discutir um pouco sobre o trabalho que está sendo feito sobre os *datasets*.

A primeira pergunta que tentamos responder é: Qual o nomes dos filmes com bilheteria maior que a média de bilheteria dos filmes lançados. Para tal fizemos a seguinte consulta:

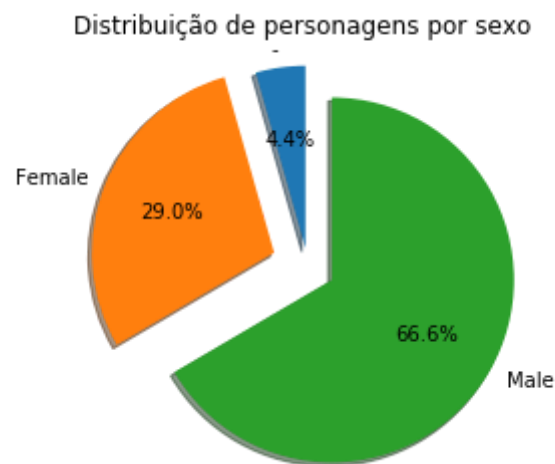
```
df = pd.read_sql("SELECT name From Movie WHERE GrossWorldwide > (SELECT AVG(GrossWorldwide) FROM Movie)", conn)
df
```

	name
0	The Avengers
1	Iron Man 3
2	Avengers: Age of Ultron
3	Captain America: Civil War
4	Black Panther
5	Avengers: Infinity War
6	Captain Marvel
7	Avengers: Endgame

Outro aspecto que achamos interessante de ser detalhado, foi o número de personagens por sexo. Para a maior parte das consultas assim como esta, optamos por gerar gráficos que auxiliam na visualização dos nossos resultados.

```
dfgender = pd.read_sql("SELECT Gender, count(P.Name) AS Personagens "+
                        "FROM Powers AS P LEFT JOIN character AS C on P.Name = C.name "+
                        "GROUP BY Gender",conn)
dfgender
```

	Gender	Personagens
0	-	17
1	Female	101
2	Male	239



Boa parte das consultas acompanham gráficos ou outros meios de visualização e estes serão adicionados a um infográfico que está em construção e se tornará disponível para consulta em breve. Abaixo segue outras consultas implementadas:

- Três personagens que mais aparecem no filme mais bem avaliado pelo IMDB
- Listar os poderes mais comuns dos vilões dos filmes
- Listar personagem com mais tempo de tela no filme mais bem avaliado de cada fase.
- Poderes do personagem que mais apareceu no filme mais mal avaliado pelo Rotten Tomatoes. Exibir também o nome desse filme.
- Os poderes do personagem que mais apareceu no filme de maior sucesso nas bilheterias.
- Perfil (características físicas e poderes) do herói com o menor tempo de tela no filme mais mal avaliado do universo cinematográfico.

6. Atividades

Boa parte das tarefas foram realizadas de forma conjunta, todavia, por tratar-se de um período difícil do semestre, acabamos encontrando dificuldades em nos encontrarmos e implementarmos nossas ideias juntos. Para sanar essas dificuldades, as tarefas do presente trabalho foram divididas de forma equivalente da seguinte forma:

As consultas foram pensadas em conjunto de modo que todas apresentem um nível de dificuldade mais ou menos parecido. A partir da definição das mesmas, cada um dos integrantes do grupo ficou responsável pela implementação de uma parcela de tais e, posteriormente, de explicá-la ao outro integrante de modo que ambos entendam como os dados estão sendo tratados e o que está sendo realizado.

A análise da complexidade das consultas e construção do infográfico foram atividades feitas na maior parte em conjunto, por se tratarem de atividades mais complexas e que demandam uma percepção um pouco mais aguçada.