

Ficha sobre os Jogos Olímpicos

2025-04-15 by jcr

Dataset

[Dataset](#) contendo os dados históricos sobre os jogos olímpicos de Atenas em 1896 até ao Rio de Janeiro em 2016.

Como o dataset tem várias dezenas de MB foram preparados [datasets mais pequenos](#) para poderem ser trabalhados na sala de aula. Nomeadamente com 100, 1000 e 10000 registos.

Plano de trabalho

Tenta resolver as questões que se colocam nas secções seguintes.

1. Modelo ontológico

Como podes ver facilmente analisando as primeiras linhas, a folha de cálculo não é modelo ideal para armazenar esta informação. Há várias dimensões que não cabem no limite de 2 dimensões.

Analisa o dataset e tenta modelar, com a ajuda de um diagrama, o modelo ontológico ideal para conter a informação do dataset.

Será promovida uma discussão na aula para se chegar a um modelo final mas é importante que consigas avançar com uma proposta tua.

2. Migração de Dados

Tendo o modelo estabilizado e especificado no Protégé, é preciso migrar a informação do dataset para o nosso modelo ontológico.

Depois de uma breve discussão com o docente sobre a abordagem a seguir, tenta criar a script de migração.

Após a migração, carrega a ontologia resultante no GraphDB, que deverás ter a correr num docker.

3. Queries

Vais agora tentar especificar várias queries e ser crítico em relação ao resultado obtido. Regista as queries que vais especificando num ficheiro de texto.

Testa todas as queries que vais fazendo no GraphDB.

1. Especifica uma query que produz a lista dos nomes de atletas que ganharam a medalha de ouro;
2. Especifica uma query que lista o nome dos atletas que ganharam medalhas e o respetivo número de medalhas (distribuição de medalhas por atletas), tem atenção à informação incompleta;
3. Especifica uma query que calcula a distribuição de atletas por equipa;

4. Especifica uma query que lista as equipas com a média do peso e da altura dos seus atletas;
 5. Especifica uma query que indica qual o atleta mais idoso que participou no **Judo** e em que ano;
 6. Especifica uma query que lista os nomes das modalidades por ordem alfabética;
 7. Especifica uma query que calcula a distribuição de atletas por modalidade;
-

4. Queries++

8. Especifica uma query que lista os atletas e respetivas alturas;
9. Especifica a mesma query que em vez de devolver valores, devolve triplos, usa a cláusula **CONSTRUCT**. Testa no GraphDB e analisa a forma do resultado comparando com a query anterior.
 - **Todas as variáveis** na cláusula **CONSTRUCT** têm de estar na cláusula **WHERE**.
10. Especifica uma query que produz uma lista dos eventos;
11. Especifica uma query que produz uma lista de eventos em que cada evento tem associada uma lista de atletas (é preciso utilizar a função de agregação **GROUP_CONCAT**);
12. Especifica uma query que calcula os triplos necessários para completar a ontologia com a simetria da relação **éParticipaçãoDe** (**temParticipação**), usa a cláusula **CONSTRUCT**;
 - Vamos **inserir** esta informação na BD:
13. Cria uma nova query, baseada na anterior, onde substituis a cláusula **CONSTRUCT** pela cláusula **INSERT**. Testa e observa o que acontece no GraphDB;
14. Com os novos triplos inseridos na BD, cria uma nova query para a distribuição de atletas por evento, onde usas **property chains** para simplificar;

Revisitando as medalhas

- Uma participação relaciona-se com os atletas pela relação **temParticipante**;
 - Vamos completar a simetria desta relação:
 1. Testando com uma query **CONSTRUCT**;
 2. Acrescentando os triplos em falta com **INSERT**.
15. Vamos voltar a especificar agora a distribuição de medalhas por atleta, tirando partido da nova relação e usando novamente **property chains**;
 - Desafio: **Lista as medalhas do atleta mais medalhado**
-

App Web de exploração

- Pensa na interface de uma Aplicação Web para explorar este dataset e as queries desenvolvidas;
- Cria o dashboard principal.