

Web Semântica – 2º Semestre 2020-21

Relatório Trabalho Prático 1

Curso:

Mestrado em Engenharia Informática

Data:

6 de Maio de 2021

Docentes:

Professor Hélder Zagalo

Discentes:

João Magalhães - 79923

João Ferreira - 80041

Índice

Introdução	1
Dados, suas fontes e sua transformação	2
Operações sobre os dados (SPARQL)	2
Funcionalidades da aplicação (UI)	2
Conclusões	2
Configuração para executar a aplicação	2
Referências	2

Introdução

Este trabalho foi desenvolvido no âmbito da UC Web Semântica do Mestrado em Engenharia Informática da Universidade de Aveiro. Para o desenvolvimento desta aplicação foi utilizado um dataset em formato CSV que contém todos os títulos existentes na plataforma Netflix. [1]

Como pedido no enunciado fornecido, foram utilizadas as seguintes tecnologias, abordadas nas aulas: Python/Django, para a programação da aplicação; RDF, sendo que os dados utilizados estão em formato NT; Triplestore GraphDB, que funciona como repositório de dados e SPARQL para pesquisa e alteração dos dados existentes na triplestore.

Dados, suas fontes e sua transformação

Os dados foram recolhidos do link presente no ponto [1] das referências e foram, posteriormente, convertidos de CSV para NT. De modo a convertermos os dados para este formato, utilizamos um script Python a que chamamos `parse_csv`.

Neste script começamos por ler o ficheiro CSV linha a linha, sendo que, para cada linha, retiramos os campos *type*, *title*, *directed*, *cast*, *country*, *date*, *release year*, *duration* e *listed in*. Inicialmente, definimos também um URI base, sujeitos e predicados.

A partir destes dados construímos o ficheiro `netflix_titles.nt` que posteriormente será carregado para a GraphDB, num repositório chamado `netflix`.

Operações sobre os dados (SPARQL)

A aplicação desenvolvida permite consultar, adicionar e pesquisar dados na base de dados. As queries utilizadas para estas operações encontram-se no ficheiro `queries.txt`, sendo utilizadas nos ficheiros `repository.py` e `graphDBConnection.py`, ambos presentes no diretório `app`.

Uma das queries possíveis de utilizar é a da figura seguinte, sendo que esta query é utilizada para pesquisar por 'la casa'.

Esta query é case insensitive e procura por 'la casa' nos campos *title*, *directed by*, *cast* e *listed_in*. Não foram tidos em conta outros campos para a pesquisa uma vez que não os consideramos relevantes (por exemplo, não faria sentido pesquisar um filme pela sua duração em minutos).

```
PREFIX net:<http://netflix-titles.com/pred/>

Select ?film ?type ?title ?directed_by ?cast ?country ?date_added ?release_year ?duration ?listed_in
where{
    ?film net:type ?type .
    ?film net:title ?title .
    ?film net:directed_by ?directed_by .
    ?film net:cast ?cast .
    ?film net:country ?country .
    ?film net:date_added ?date_added .
    ?film net:release_year ?release_year .
    ?film net:duration ?duration .
    ?film net:listed_in ?listed_in .
    filter (regex (?title, 'la casa', 'i') || regex(?directed_by, 'la casa', 'i') || regex(?cast, 'la casa', 'i') || regex(?listed_in, 'la casa', 'i'))
}
```

Outra das queries possíveis de ser utilizada é a figura seguinte, sendo que esta query é utilizada para adicionar um filme. Outras queries foram utilizadas e encontram-se no ficheiro queries.txt.

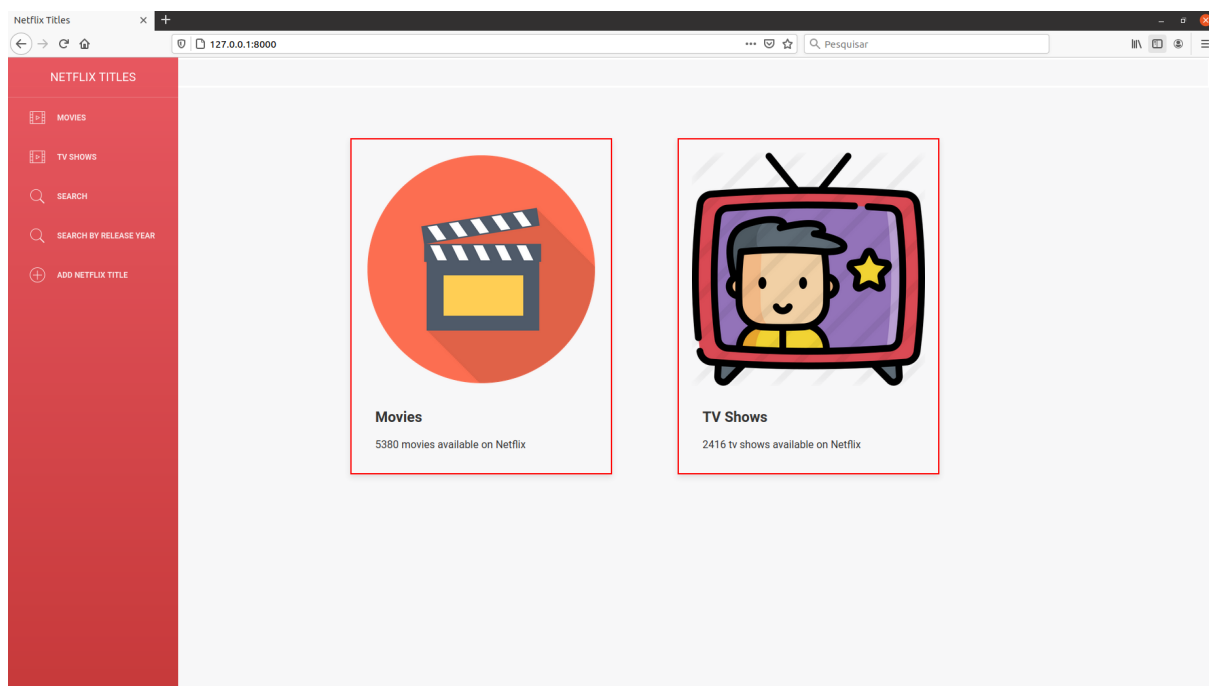
```
#add film

prefix n:<http://netflix-titles.com/>
PREFIX net:<http://netflix-titles.com/pred/>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

Insert data{
    n: net:type "Movie";
    net:title "FM";
    net:directed_by "FM";
    net:cast "FM";
    net:country "FM";
    net:date_added "FM";
    net:release_year "FM";
    net:duration "FM";
    net:listed_in "FM".
}
```

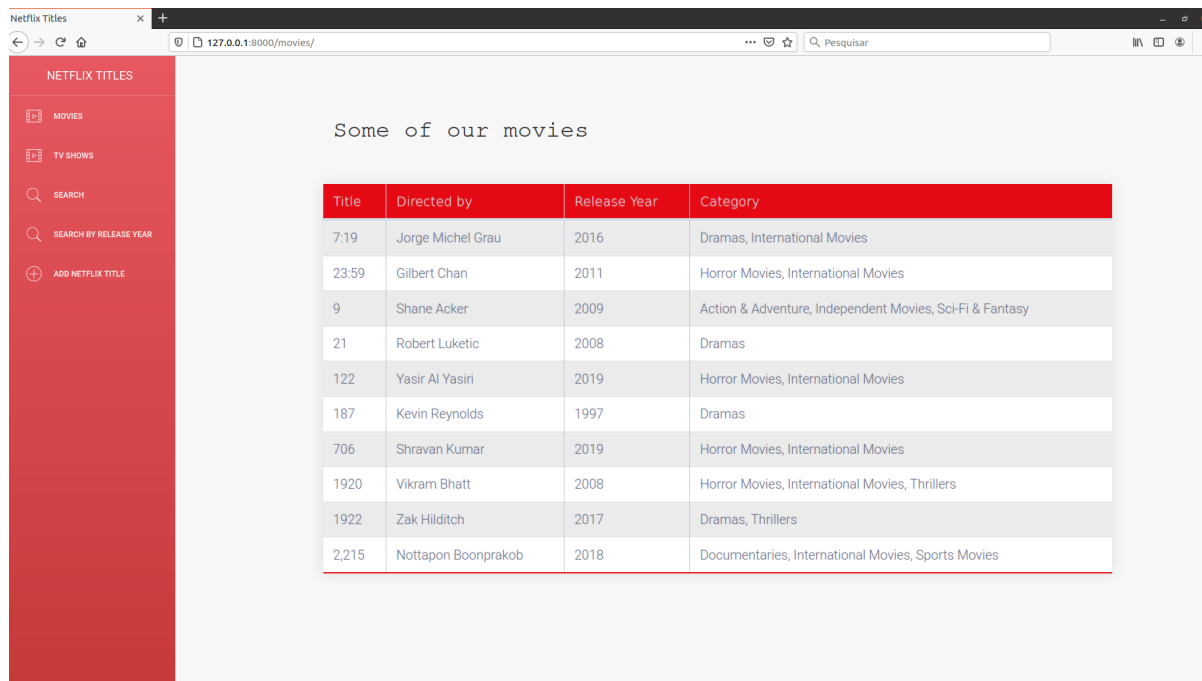
Funcionalidades da aplicação (UI)

Home



Quando o utilizador acede à aplicação, vê a imagem acima em que consegue saber quantos filmes e tv shows estão presentes no dataset, neste caso, tínhamos 5380 filmes e 2416 tv shows. A barra lateral presente nesta página, está presente no resto das páginas da aplicação e tem 6 botões para as páginas “Netflix titles” (página inicial), “Movies”, “TV Shows”, “Search”, “Search by release year”, “Add netflix title”.

Movies



Netfix Titles

127.0.0.1:8000/movies/

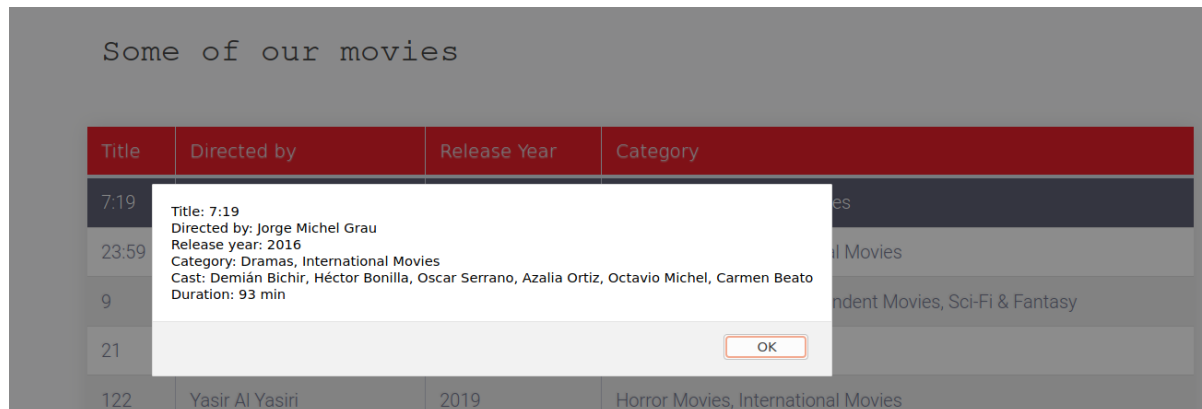
NETFLIX TITLES

- MOVIES
- TV SHOWS
- SEARCH
- SEARCH BY RELEASE YEAR
- ADD NETFLIX TITLE

Some of our movies

Title	Directed by	Release Year	Category
7:19	Jorge Michel Grau	2016	Dramas, International Movies
23:59	Gilbert Chan	2011	Horror Movies, International Movies
9	Shane Acker	2009	Action & Adventure, Independent Movies, Sci-Fi & Fantasy
21	Robert Luketic	2008	Dramas
122	Yasir Al Yasiri	2019	Horror Movies, International Movies
187	Kevin Reynolds	1997	Dramas
706	Shravan Kumar	2019	Horror Movies, International Movies
1920	Vikram Bhatt	2008	Horror Movies, International Movies, Thrillers
1922	Zak Hilditch	2017	Dramas, Thrillers
2,215	Nottapon Boonprakob	2018	Documentaries, International Movies, Sports Movies

Esta página é apresentada quando o utilizador vai para a página “Movies” e são-lhe apresentados alguns dos filmes, 10 dos 5380, que estão presentes no dataset. Se carregarmos num dos filmes irá aparecer um pop-up com a restante informação do filme como mostra na imagem abaixo.



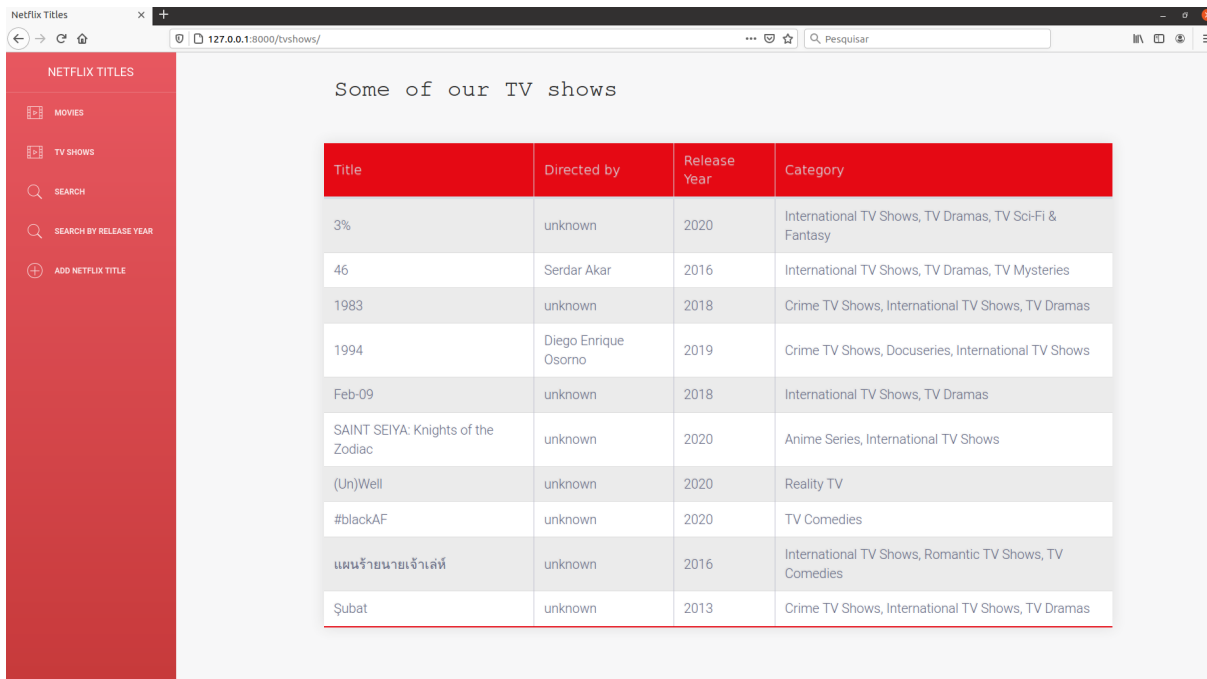
Some of our movies

Title	Directed by	Release Year	Category
7:19	Jorge Michel Grau	2016	Dramas, International Movies
23:59	Gilbert Chan	2011	Horror Movies, International Movies
9	Shane Acker	2009	Action & Adventure, Independent Movies, Sci-Fi & Fantasy
21	Robert Luketic	2008	Dramas
122	Yasir Al Yasiri	2019	Horror Movies, International Movies

Title: 7:19
Directed by: Jorge Michel Grau
Release year: 2016
Category: Dramas, International Movies
Cast: Demián Bichir, Héctor Bonilla, Oscar Serrano, Azalia Ortiz, Octavio Michel, Carmen Beato
Duration: 93 min

OK

TV Shows



Netfix Titles

127.0.0.1:8000/tvshows/

NETFLIX TITLES

MOVIES

TV SHOWS

SEARCH

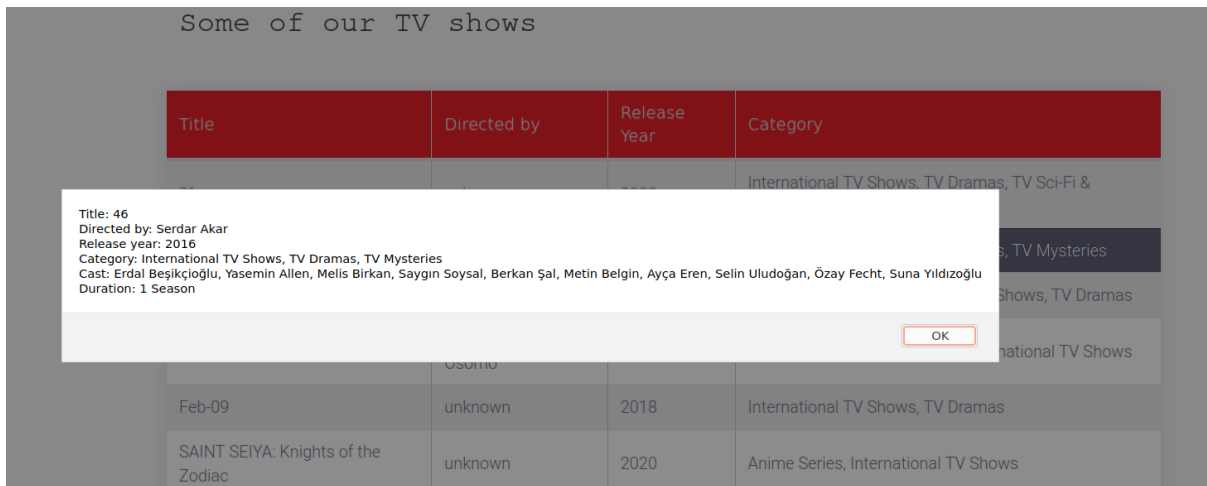
SEARCH BY RELEASE YEAR

ADD NETFLIX TITLE

Some of our TV shows

Title	Directed by	Release Year	Category
3%	unknown	2020	International TV Shows, TV Dramas, TV Sci-Fi & Fantasy
46	Serdar Akar	2016	International TV Shows, TV Dramas, TV Mysteries
1983	unknown	2018	Crime TV Shows, International TV Shows, TV Dramas
1994	Diego Enrique Osorno	2019	Crime TV Shows, Docuseries, International TV Shows
Feb-09	unknown	2018	International TV Shows, TV Dramas
SAINT SEIYA: Knights of the Zodiac	unknown	2020	Anime Series, International TV Shows
(Un)Well	unknown	2020	Reality TV
#blackAF	unknown	2020	TV Comedies
แผนร่ายนายเจ้าเล่ห์	unknown	2016	International TV Shows, Romantic TV Shows, TV Comedies
Şubat	unknown	2013	Crime TV Shows, International TV Shows, TV Dramas

Esta página funciona como a página “Movies” mas com os tv shows presentes no nosso repositório.



Some of our TV shows

Title	Directed by	Release Year	Category
46	Serdar Akar	2016	International TV Shows, TV Dramas, TV Sci-Fi & Fantasy
1983	unknown	2018	Crime TV Shows, International TV Shows, TV Dramas
1994	Diego Enrique Osorno	2019	Crime TV Shows, Docuseries, International TV Shows
Feb-09	unknown	2018	International TV Shows, TV Dramas
SAINT SEIYA: Knights of the Zodiac	unknown	2020	Anime Series, International TV Shows

Title: 46
Directed by: Serdar Akar
Release year: 2016
Category: International TV Shows, TV Dramas, TV Mysteries
Cast: Erdal Beşikçioğlu, Yasemin Allen, Melis Birkan, Saygın Soysal, Berkan Şal, Metin Belgin, Ayça Eren, Selin Uludoğan, Özay Fecht, Suna Yıldızoğlu
Duration: 1 Season

OK

Search

Netflix Titles

Search page

Search:

Search for: **la casa**

Found 2 results.

Type	Title	Directed by	Cast	Country	Date Added	Release Year	Duration	Category
TV Show	La casa de papel	unknown	Úrsula Corberó, Itziar Ituño, Álvaro Morte, Paco Tous, Enrique Arce, Pedro Alonso, María Pedraza, Alba Flores, Miguel Herrán, Jaime Lorente, Esther Acebo, Darko Peric, Kiti Mánver	Spain	April 3, 2020	2020	4 Seasons	Crime TV Shows, International TV Shows, Spanish-Language TV Shows
TV Show	The Mess You Leave Behind	unknown	Inma Cuesta, Bárbara Lennie, Tamar Novas, Arón Piper, Roberto Enríquez, Roque Ruiz, Isabel Garrido, Fede Pérez, Alfonso Agra, Susana Dans, Xavier Estévez, Xosé Antonio Touriñán, María Tasende, Camila Bossa, María Costas, Abril Zamora, Chelo Falcón, Ana Santos, Mel la casa , César Cambeiro	Spain	December 11, 2020	2020	1 Season	International TV Shows, Spanish-Language TV Shows, TV Dramas

No “Search” é realizada uma pesquisa de todos os filmes e tv shows que contenham as palavras que o utilizador inserir. Como explicado acima, esta pesquisa é case insensitive e procura por ‘la casa’ nos campos *title*, *directed by*, *cast* e *listed_in*. Neste caso podemos ver que o ‘la casa’ está presente no campo *title* do primeiro tv show encontrado e no campo *cast* do segundo.

Search by release year

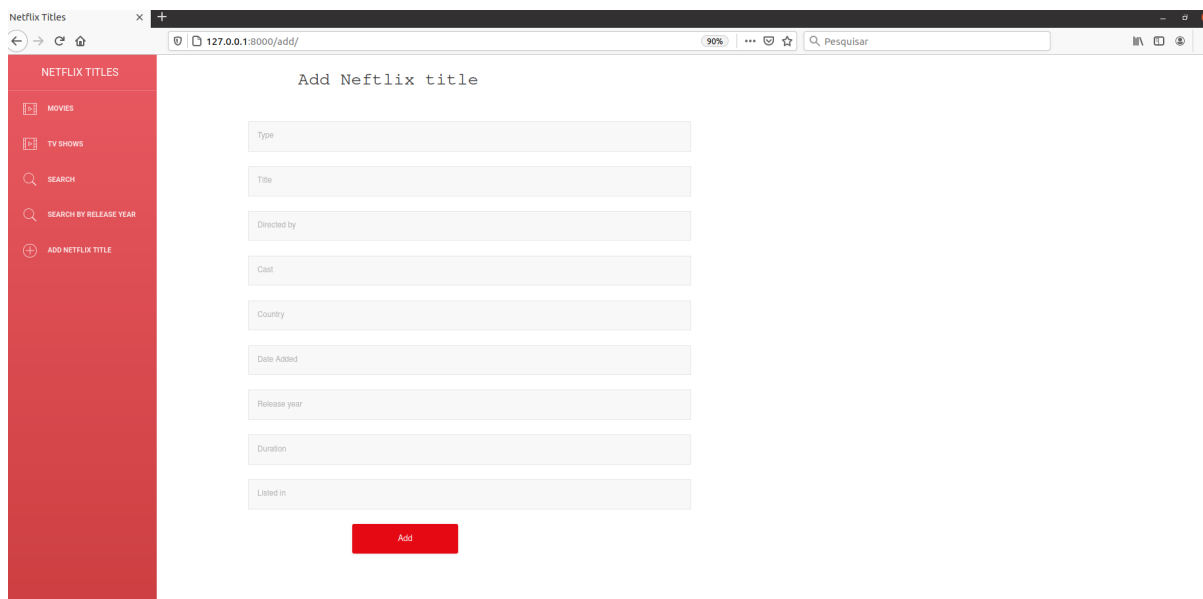


The screenshot shows a web browser window with the URL `127.0.0.1:8000/year/`. The page title is "Search page by release year". Below the title, it says "Search by release year from 1945 to 1949" and "Found 6 results.". A table displays the search results:

Type	Title	Directed by	Cast	Country	Date Added	Release Year	Duration	Category
Movie	Know Your Enemy - Japan	Frank Capra, Joris Ivens	Walter Huston, Dana Andrews	United States	March 31, 2017	1945	63 min	Classic Movies, Documentaries
Movie	Nazi Concentration Camps	George Stevens		United States	March 31, 2017	1945	59 min	Classic Movies, Documentaries
Movie	San Pietro	John Huston		United States	March 31, 2017	1945	32 min	Classic Movies, Documentaries
Movie	Let There Be Light	John Huston	Walter Huston	United States	March 31, 2017	1946	58 min	Classic Movies, Documentaries
TV Show	Pioneers of African-American Cinema	Oscar Micheaux, Spencer Williams, Richard E. Norman, Richard Maurice		United States	February 1, 2017	1946	1 Season	TV Shows
Movie	Thunderbolt	William Wyler, John Sturges	James Stewart, Lloyd Bridges	United States	March 31, 2017	1947	42 min	Classic Movies, Documentaries

Nesta página, o search é realizado de forma diferente. É feito através dos anos de lançamento que o utilizador inserir. Neste caso, os anos inseridos foram 1945 e 1949. Assim, aparecem todos os filmes lançados no intervalo desses anos, inclusivé.

Add Netflix title



O utilizador pode inserir dados referentes a um filme que pretende adicionar à netflix. Nesta inserção só os campos *type*, *title*, e *release_year* são obrigatórios. Quando esses campos são inseridos corretamente o filme é adicionado ao dataset, com um id gerado automaticamente.

Conclusões

Em conclusão, consideramos ter atingido todos os objetivos propostos pelo Professor, sendo que, entendemos que outras operações sobre os dados poderiam ter sido feitas (como o update, sendo que a query se encontra no ficheiro queries.txt) mas que no tema em que o trabalho se baseia não teriam qualquer relevância.

Outro dos aspectos que consideramos que poderiam ser melhorados estaria relacionado com a usabilidade da interface, nomeadamente tornar esta um pouco mais amigável para o utilizador. Pensamos também como trabalho futuro aprender como integrar (devidamente funcional e seguro) uma página de login, sendo que essa funcionalidade seria mais importante num momento em que a aplicação atingisse um ponto de escalabilidade em que fosse necessário que existisse.

Concluindo, consideramos este trabalho útil porque nos permitiu, sobretudo, aumentar o nosso conhecimento sobre modelação e interação com e entre dados mas também fomentar a nossa capacidade de criação de interfaces para o utilizador.

Configuração para executar a aplicação

Para executar a aplicação é necessário seguir os passos indicados abaixo:

1. Executar GraphDB na porta 7200 e abrir a interface do GraphDB no browser.
2. Criar um novo repositório com o ID “netflix”.
3. Importar o ficheiro netflix_titles.nt para o novo repositório.
4. Executar o projeto Django.
5. Aceder a <http://localhost:8000> (página da aplicação).

Referências

- [1] - Netflix Movies and TV Shows, Shivam Bansal
<https://www.kaggle.com/shivamb/netflix-shows>