

# Trabalho Final - Grafos de Conhecimento

Universidade Tecnológica Federal do Paraná

Profa. Dra. Rita Berardi e Prof. Dr. Cesar Tacla

Alunos: Alex Pinheiro das Graças, Guilmour Rossi, Vanessa Lazarin, Vitor Teixeira

## 1. Introdução

Esse trabalho tem como objetivo criar um grafo de conhecimento (KG, do inglês, Knowledge Graph) a partir da modelagem de uma ontologia, e em seguida, fazer consultas usando SPARQL nesse KG.

O domínio escolhido foi o do projeto em software livre Libreflix. O Libreflix é uma plataforma comunitária e gratuita de streaming, que reúne filmes independentes do Brasil e de outros países. O projeto Libreflix foi desenvolvido na UTFPR por um dos integrantes do grupo, Guilmour Rossi, durante a sua graduação.

## 2. Modelagem da Ontologia no Protégé

Para a criar a nossa ontologia, usamos o software Protégé (Fig. 1), indicado para a criação da ontologia através de um arquivo OWL.

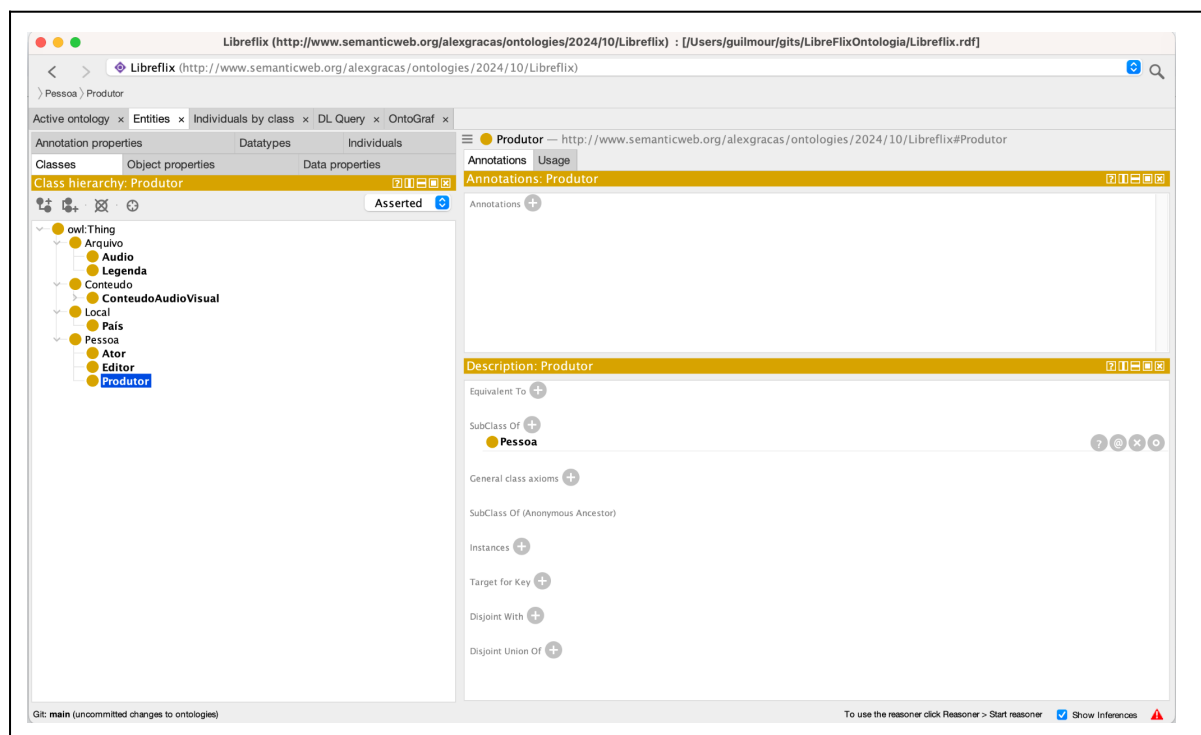
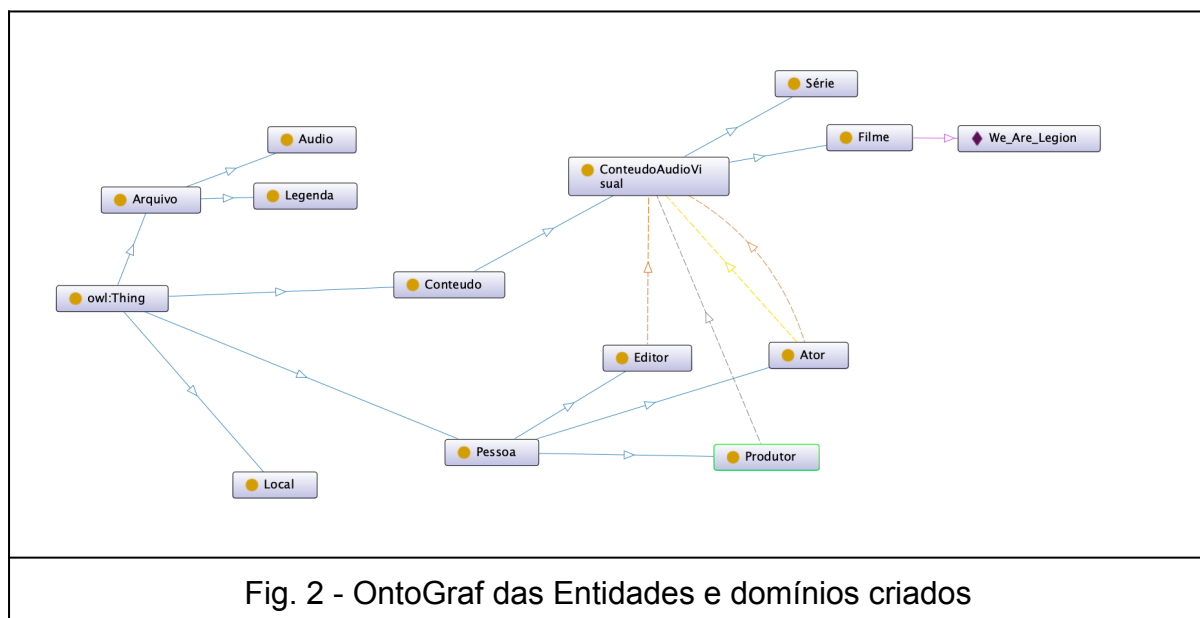


Fig. 1 - Captura de Tela do Protégé

Foi modelado o seguinte OWL (Fig. 2), com propriedades principais pertinentes a uma obra audiovisual do Libreflix e suas relações com as pessoas de atores e equipe. Localidade, Classificação Indicativa e outras classes serão usadas direto da DBPedia.



### 3. Triplificação dos Dados

Para a triplificação dos dados, exportamos alguns dados do Libreflix com a informação de cerca de 190 obras. O CSV foi exportado e passou por uma limpeza de dados (Fig. 3).

permalink	duration	classind	title	subtitle	sinopse
deus-e-o-diabo-em-cima-da-muralha	54	14	Deus e o Diabo em Cima da Muralha	Dráuzio Varella no Carandiru	Drauzio Varella e funcionários
quebrando-o-tabu	80	14	Quebrando o Tabu		O filme propõe um debate sér
tpb-afk	82	10	The Pirate Bay AFK	The Pirate Bay Além do Teclado	Documentário sobre a vida do
the-code	59	L	The Code	A história do GNU/Linux	The Code é um documentário
we-are-legion	95	14	We Are Legion	Nós Somos Anonymous	Este documentário nos leva p
earthlings	95	18	Terráqueos	Natureza, Animais e Humanidade	Earthlings é um documentário
life-in-a-day	94	14	Life in a Day	A Vida em Um Dia	O filme é formado por vídeos
planet-ocean	92	10	Planeta Oceano		Uma união única de pesquisac
the-internets-own-boy	105	12	O Menino da Internet	A História de Aaron Swartz	A história do ativista de intern
observar-e-absorver	70	12	Observar e Absorver	A Filosofia de Eduardo Marinho	"Eu sou extremamente ambici
hotel-laide	24	14	Hotel Laide		Hotel Laide foi um dos mais in
quem-matou-eloa	24	16	Quem Matou Eloá?		Em 2008, Lindemberg Alves d
10-anos-jean-charles-relatos	15	14	Jean Charles 10 Anos Depois	Cinco Relatos	Em 22 de julho de 2005, o bra
kungfury	31	16	Kung Fury		Um detetive de Miami dotado
revolution-os	85	L	Revolution OS		Revolution OS é um document
eu-nao-quero-voltar-sozinho	17	12	Eu Não Quero Voltar Sozinho		A vida de Leonardo, um adole

Fig. 3 - CSV exportado e limpadado dos dados da plataforma Libreflix

Como no projeto Libreflix é usado um banco de dados não-relacional, especificamente o MongoDB, uma exportação para um CSV não é diretamente simples de ser feita.

Após isso, os dados foram importados no Ontotext Refine, e começamos a mapear cada uma das entidades do CSV. Tivemos algumas dificuldades com o Ontotext travando por vezes, mas no final tudo funcionou.

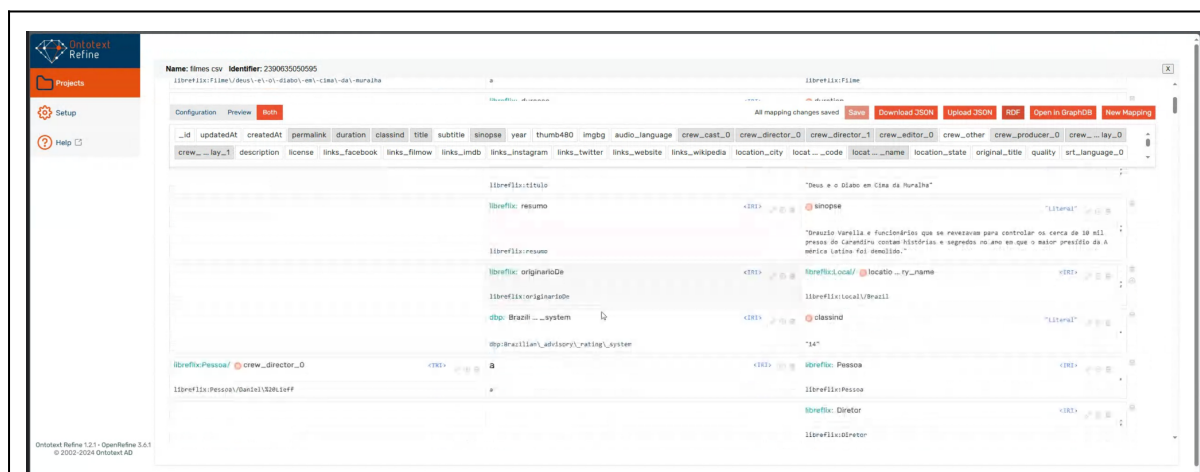


Fig. 4 - Mapeamento no Ontotext Refine

## 4. Perguntas de Competência no SPARQL

Escolhemos três perguntas para realizarmos as consultas no SPARQL.

1. Qual a classificação indicativa das obras que possuem “Dráuzio Varella” como ator/entrevistado no catálogo da Libreflix?



Fig. 5 - Query SPARQL e resposta da Pergunta 1

2. Qual o diretor dirigiu mais obras audiovisuais no catálogo da Libreflix?



3. Qual o título e o diretor do filme mais longo do catálogo da Libreflix?



## 4. Entregáveis

- 1 arquivo texto descrevendo brevemente o domínio que você modelou e com as perguntas e devidas queries em SPARQL. (o próprio documento)
- 1 arquivo da ontologia modelada (owl) - (ontologia.owl e libreflix.rdf)
- o arquivo CSV utilizado (filmes.csv)
- 1 arquivo RDF resultante da triplificação no GraphDB (arquivo de triplas) (triplificacaoLibreflix.ttl)
- 1 arquivo de mapeamento gerado pelo refine (mapeamentoLibreflix.json)
- 1 arquivo com prints dos resultados das consultas realizadas no GraphDB. (em anexo neste documento).