



Web sémantique et Big data  
Projet de réalisation d'une ontologie du cinéma  
RAPPORT  
Pour le 03 mai 2021

Morgann Sabatier et Julien Bezançon  
MASTER 1 « Langue et informatique »

# Table des matières

<b>1</b>	<b>Etat de l’art</b>	<b>2</b>
1.1	Introduction . . . . .	2
1.2	Ontologies et généralités . . . . .	3
1.2.1	Différentes ontologies . . . . .	3
1.3	Cinéma et ontologies . . . . .	4
1.3.1	Movie Labs, une ontologie de l’industrie du film et de la télévision . . . . .	4
1.3.2	Une ontologie représentant le déroulement d’une histoire transmédia . . . . .	5
<b>2</b>	<b>Projet</b>	<b>8</b>
2.1	Création de l’ontologie . . . . .	8
2.1.1	Classes et individus de l’ontologie . . . . .	8
2.1.2	Propriétés de l’ontologie . . . . .	10
2.1.3	Consistance de notre ontologie . . . . .	13
2.2	Requêtes SPARQL . . . . .	15
2.3	Conclusion . . . . .	19

# Chapitre 1

## Etat de l'art

### 1.1 Introduction

Ce projet a pour but de comprendre les enjeux des ontologies, apprendre à structurer, créer et utiliser une ontologie et ses outils. Nous expliquerons le déroulé de notre projet, ce que nous avons appris ainsi que les difficultés que nous avons rencontrées.

Pour ce projet de Web Sémantique, nous avons choisi de nous concentrer sur la création d'une ontologie liée au septième art. Ce rapport concernera l'ensemble de notre démarche de création, de nos recherches et applications. Dans un premier temps, nous allons faire un bref état de l'art des ontologies, puis nous préciserons nos recherches avec des ontologies du cinéma. Ensuite, nous expliciterons la construction de notre ontologie et ses applications.

Dans un premier temps nous présenterons un bref aperçu des travaux effectués sur les ontologies et l'ingénierie des connaissances. Ensuite, nous focaliserons notre propos afin de donner une représentation d'ensemble des ontologies liées au cinéma.

## 1.2 Ontologies et généralités

Les ontologies naissent d'un besoin profond de représenter les connaissances qui puissent être utilisées et interpréter à la fois par l'homme et par la machine[Camille Mouhamadou Saliou Diallo, 2013]. Ainsi l'ontologie va permettre de représenter de façon simple et compréhensible un domaine qui peut être extrêmement complexe à l'aide d'une construction logique basée sur une structure Objet-relation-objet. Cette représentation demande donc rigueur et mises à jour régulières afin d'être pertinente. C'est d'ailleurs pour cette raison que bon nombre d'ontologies sont participatives. Les contributeurs s'impliquent dans la construction par des outils qui, aujourd'hui, sont particulièrement adaptatifs[Giboin, 2018].

### 1.2.1 Différentes ontologies

#### Ontologies unifiées

Elles ont pour but de donner une immense représentation d'un concept. Elles se concentrent à la fois sur des sujets abstraits et concrets dans toute une variété de domaine[Hoekstra, 2009]. On citera notamment SENSUS<sup>1</sup>, immense ontologie lexicale ayant pour but de servir à des systèmes de traduction.

#### Ontologies de haut niveau

Elles représentent des concepts très généraux. Il s'agit des ontologies les moins spécifiques. Il est raisonné de dire que ces ontologies devront être unifiés pour une communauté immense d'utilisateurs[Guarino, 1997].

#### Ontologies de tâche et de domaine

Ces dernières sont spécifiques à un domaine général ou encore une activité. Par exemple, le domaine médical ou encore le domaine du cinéma comme nous l'avons fait pour l'ontologie pour ce projet. Elles se spécialisent par rapport aux ontologies de haut niveau.

---

1. <https://www.isi.edu/natural-language/resources/sensus.html>

## Ontologies d'application

Elles sont l'intersection de l'ontologie de domaine et de tâche. On parlera souvent d'activités spécifiques telle que remplacer une pièce de mécanique.

Le domaine du cinéma offre un véritable terrain de jeu dans pour la création d'ontologie de domaines. C'est ce dernier auquel nous nous intéressons. Nous allons tout d'abord l'explorer à travers des travaux qui ont déjà été fait, puis vous présenter notre démarche de construction de l'ontologie.

## 1.3 Cinéma et ontologies

Afin de réaliser une ontologie cohérente, il est évident que nous devons voir ce qui avait été fait par la communauté scientifique. En effet, cela nous a permis tout d'abord de comprendre les enjeux d'une ontologie du cinéma, de nous inspirer de différentes méthodologies et structures d'ontologies existantes mais surtout de modeler notre travail en fonction des usages que nous souhaitions mettre en valeur.

### 1.3.1 Movie Labs, une ontologie de l'industrie du film et de la télévision

L'ontologie créée par MovieLabs met en valeur la nécessité d'organiser, de mettre en relation des entités et d'uniformiser cet ensemble. L'objectif est de diminuer la duplication des données et d'adopter un modèle unique pour l'ensemble de l'industrie via une ontologie pertinente et collaborative. Afin de permettre une réutilisation des données pour l'industrie audiovisuelle de manière aisée et structurée.

L'architecture des données audiovisuelle s'est souvent concentrée sur la description des entités plutôt qu'une mise en relation et une organisation permettant leur exploitation pérenne.

Une ontologie commune permet d'avoir à disposition un ensemble à la fois robuste et flexible de données afin de les intégrer à différents systèmes. Le projet est drastique mais offre un socle afin de laisser libre cours aux chercheurs de concentrer leurs efforts sur l'innovation et la création de nouveaux systèmes plutôt que de perdre du temps à créer une multiplicité d'ensemble de données.

Cette économie de temps permet de focaliser les efforts sur l'application des données et non par la collection et l'organisation de ces dernières. Les applications peuvent être d'ordre analytiques ou marketing. Le Machine

Learning est un outil particulièrement puissant pour construire des modèles et analyser des données correctement structurées[MovieLabs, 2018].

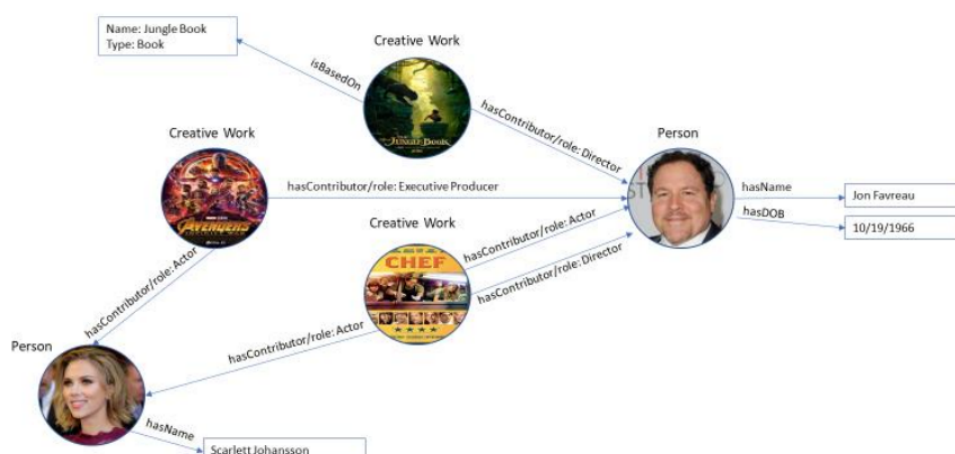


FIGURE 1.1 – Extrait exemple de la structure de l'ontologie MovieLabs

La création de cette ontologie met en valeur à la fois la complexité de rassembler un grand nombre de données et de les structurer mais nous offre une opportunité d'observer les possibilités liées à cette structure. Elle rend également compte de la force d'un travail collaboratif efficace lorsque le squelette est déjà déterminé et qu'il suffit d'y ajouter des mises à jour.

Cette ontologie nous a permis de nous rendre compte de la puissance des ontologies, leurs capacité à donner une photographie précise d'un domaine de travail. Nous avons donc motivé notre projet autour de la pertinence de la structure de nos données.

### 1.3.2 Une ontologie représentant le déroulement d'une histoire transmédia

Parmi la pluralité d'ontologies liées au cinéma, la plupart s'évertuent à proposer une séquence objective d'entités et de relations afin de proposer une description concrète du milieu du cinéma.

Ici, nous allons parler du projet d'utiliser l'ontologie comme représentation non pas d'une oeuvre mais bien d'un univers ayant plusieurs adaptations tout en conservant et mettant en valeur des trames narratives, leurs liens selon le média employé.

Cette idée de représentation d'un univers et de mise en valeur des ressources narratives permet d'avoir à la fois une vision d'ensemble, objective, d'un univers mais également des ressources mises en contexte. On a alors un double emploi qui rassemble informations objectives et narratives [Sánchez and Pérez, 2018].

Ce concept est particulièrement intéressant et nous démontre qu'une ontologie peut ne pas s'arrêter à un ensemble de données immuables. Qu'il est possible de les modéliser en fonction de ce que l'on recherche. Ainsi, nous avons décidé d'ajouter des propriétés permettant d'identifier le type de rôle incarné par un acteur via des propriétés booléennes de protagonistes ou d'antagonistes.

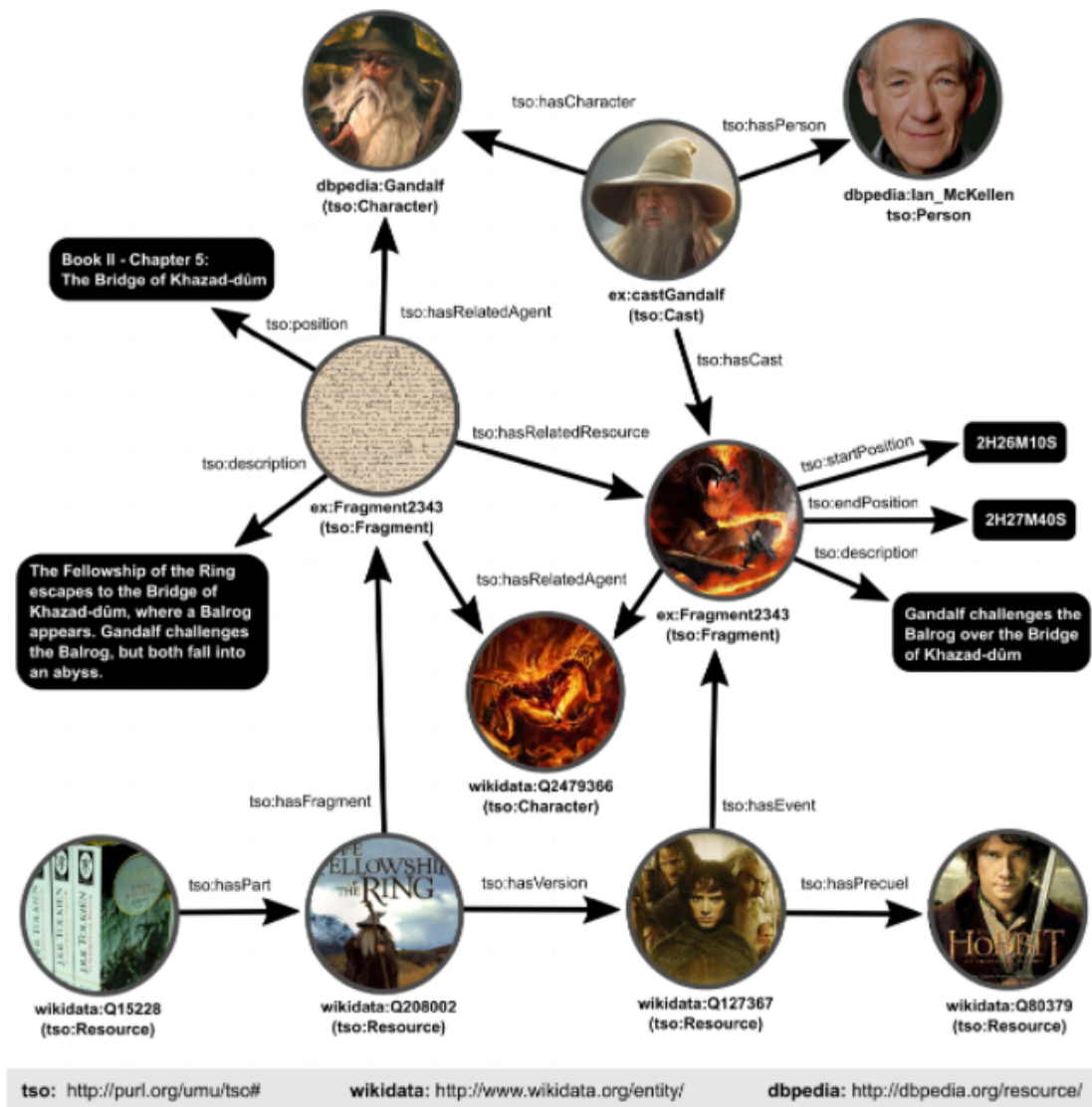


FIGURE 1.2 – Exemple d'application de l'ontologie citée



# Chapitre 2

## Projet

### 2.1 Création de l'ontologie

#### 2.1.1 Classes et individus de l'ontologie

Dans cette ontologie, nous retrouvons de nombreuses classes et de nombreux individus, tous étant liés au domaine du cinéma. Nous avons dans un premier temps établis quelques consignes à respecter :

- Les individus de l'ontologie correspondent principalement aux acteurs, personnages et autres personnes ayant contribué aux films.
- Les classes de l'ontologie correspondent à des informations techniques autour des films : leur classification, leur genre ainsi que leur pays d'origine.
- Les films, pièces centrales de l'ontologie, doivent être relié à autant de classes et d'individus que possible, afin de permettre de varier les requêtes SPARQL à leur sujets.

Maintenant, il nous faut préciser quelque chose d'important : créer une ontologie sur le cinéma s'est révélé ardu, de par la quantité d'aspects et d'informations entourant un film. Il était pratiquement impossible pour nous de faire le tour de tous les aspects d'un film en un temps limité, mais aussi de tous ses personnages, de toutes les personnes qui ont joué un rôle lors de sa création, etc... Nous avons donc dû faire des choix. De ce fait, notre ontologie peut sembler légère par rapport à ce qui aurait pu être fait. Nous avons voulu cibler un certains nombre d'aspects, tout en faisant attention à ne pas s'enfoncer trop dedans. C'est pour ça que les listes des acteurs,

personnages et des producteurs peuvent sembler petites : nous avons limité le nombre d'individus dans chacune d'entre elles, afin de nous assurer que nous notre ontologie sera un minimum variée, et surtout que l'on puisse la finir dans les temps.

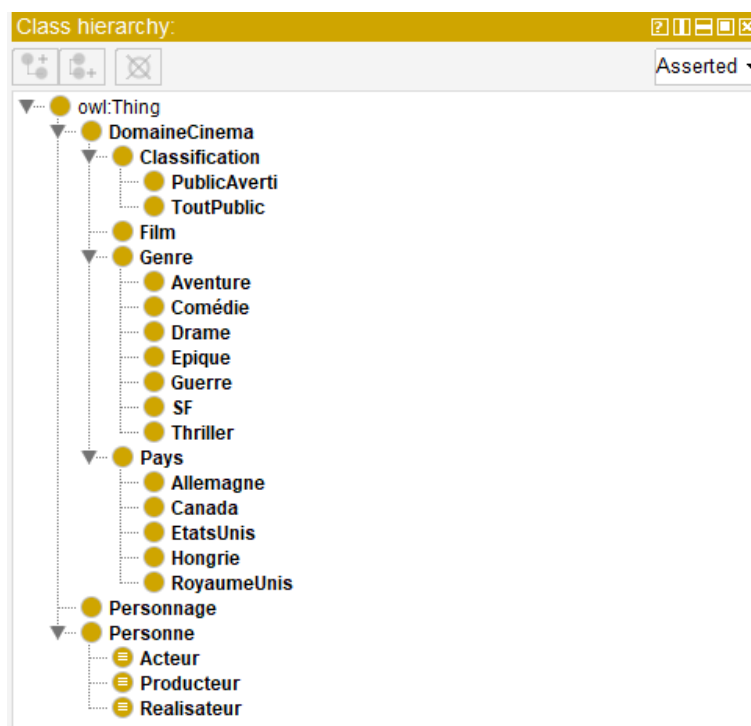


FIGURE 2.1 – Les différentes classes présentes dans l'ontologie

Nos classes sont visibles dans la figure 2.1. Pour obtenir des informations supplémentaires sur ces classes, il suffit de consulter les annotations de ces classes dans l'ontologie, sous l'annotation `rdfs:comment`. Parmi ces classes, on retrouve Acteur, Producteur et Realisateur, qui sont toutes trois des classes avec un paramètre `Equivalent To`, ce qui signifie qu'elles contiennent des individus ayant certaines propriétés.

Prenons Acteur : comme les deux autres, c'est une sous-classe de Personne, et elle contient toutes les personnes qui ont une relation `AjoueDans`, c'est-à-dire toutes les personnes qui ont déjà joué dans un film (some Film). Les classes Classification, Genre et Pays ont en individus les films qui ont la classification / le genre / qui viennent du pays adéquat. Cela permet de donner différents types aux individus films, tout en les regroupant logiquement dans des classes explicitant leur nature.

Certaines classes sont donc des étiquettes que l'on peut attribuer aux films, en leur donnant comme individu les films en question. Ce choix a été fait pour bien délimiter les individus, étant tous des personnes ou des personnages, des attributs techniques des films. Cependant, on peut s'interroger : pourquoi ces éléments concernant les films, ces attributs, ne sont-ils pas dans les data properties, alors que d'autres attributs le sont (comme l'année de sortie ou encore la note).

Là encore, nous avons différencié les attributs des films en deux catégories : les attributs pouvant être facilement énuméré, possédant déjà des réponses "par défaut" (les genres correspondent dans le milieu du cinéma à une liste établie, tout comme les pays de notre monde, ou encore les classifications des films) des attributs plus spécifiques, et possédant trop de valeurs différentes pour être énumérées (comme les années de sortie, les notes, les durées...).

En d'autres termes, cela permet de fixer les choix possibles pour les attributs mis en classe, tout en laissant plus de liberté aux valeurs des attributs mis en tant que data properties. Ce choix paraît logique, car on aura beaucoup plus de chances de pouvoir regrouper des films dans des classes en fonction de leur genre qu'en fonction de leurs dates de sortie, par exemple, et cela parce que les genres sont bien plus limités en nombre que les dates que les dates.

La 2.2 contient les individus créés dans notre ontologie. Il n'y a pas besoin de faire de remarques supplémentaires sur cette liste d'individus, si ce n'est qu'il ne s'agit, comme dit plus haut, qu'elle n'est constituée que de personnes, personnages et films.

## 2.1.2 Propriétés de l'ontologie

### Data properties

Une fois les classes et individus définis, il nous faut leur créer des propriétés. Commençons par les Data properties, les propriétés données aux individus, présentes dans la figure 3.

- **Propriétés propres aux individus étant des personnes :**
  - **age (int)** : Précise l'âge de la personne.
  - **origine (string)** : indique l'origine de la personne.
  - **sexe (string)** : Donne le sexe de la personne.
  - **vivant (bool)** : Indique si la personne est en vie ou morte.

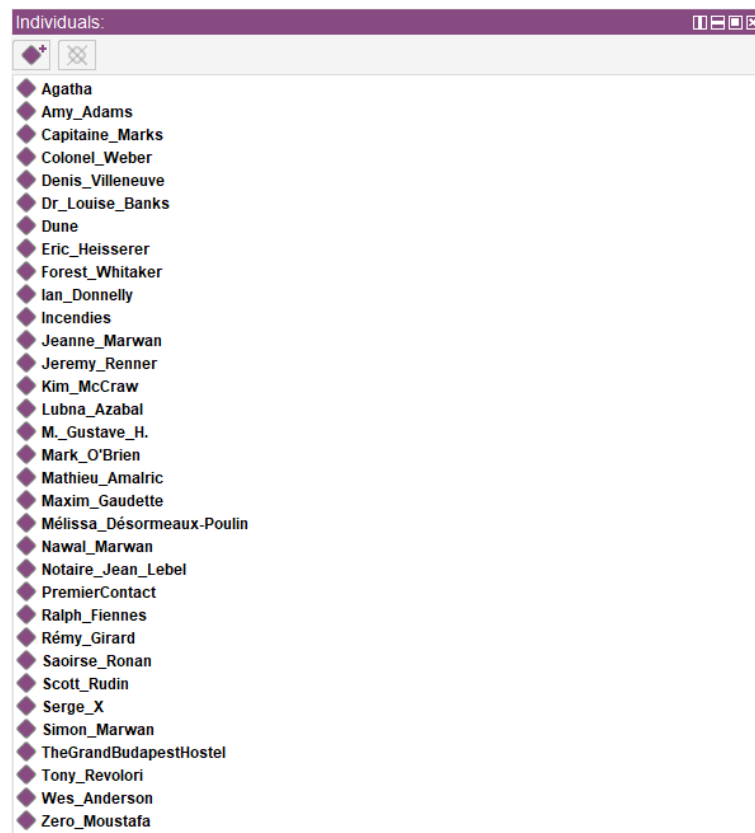


FIGURE 2.2 – Les individus de notre ontologie

- **Propriétés propres aux individus étant des personnages :**
  - **antagoniste (bool)** : indique si le personnage est antagoniste.
  - **protagoniste (bool)** : indique si le personnage est protagoniste.
- **Propriétés propres aux individus étant des films :**
  - **duration (int)** : Donne la durée d'un film en minutes.
  - **langue (language)** : Précise la langue d'un film.
  - **note (decimal)** : Précise la note du film sur Allociné.
  - **revenus (int)** : Renvoie en euros le revenu du film.
  - **sortie (int)** : Donne l'année de sortie du film.

Lors de la création de l'ontologie, nous avons réalisé qu'il n'était pas forcément évident de trouver toutes les informations désirées pour les personnages : en effet, leur âge et leurs origines ne sont pas forcément indiqués. De plus, indiquer si ces personnages sont vivants ou non ne nous semblait pas être une bonne idée.

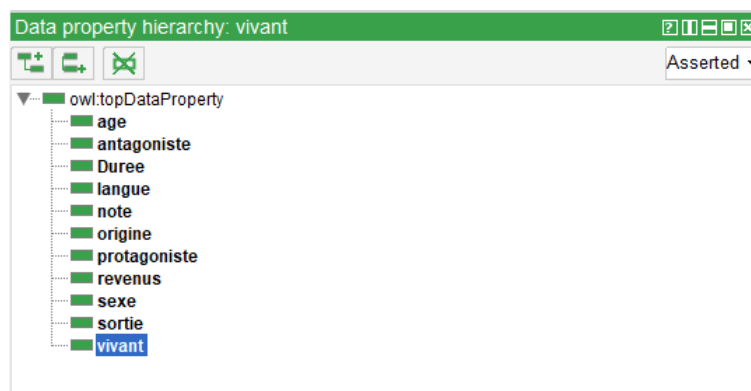


FIGURE 2.3 – les data properties de notre ontologie

De ce fait, nous avons décidé de dresser d'autres attributs pour les personnages, protagoniste et antagoniste. Ces deux attributs sont complémentaires, et non opposés, comme on pourrait d'abord le penser. Il arrive qu'un personnage soit nuancé, et rentre aussi bien dans la catégorie d'antagoniste que dans celle de protagoniste.

De plus, Un personnage peut évoluer au cours d'un film, et donc indiquer seulement si il est protagoniste ou antagoniste ne suffit pas. C'est des attributs que nous avons ajouté parce qu'ils permettent de faire des recherches intéressantes.

Imaginons que nous voulions chercher un film où Jean Reno apparaît comme antagoniste, lui qui est souvent le protagoniste dans les films. Il suffira de chercher un personnage qui est à la fois antagoniste et joué par Jean Reno.

On pourrait même dire que cette possibilité de recherche est se qui démarque notre ontologie du cinéma des autres : dans le milieu des ontologies, le cinéma est très bien représenté, et pratiquement tous ses aspects ont été abordés, mais celui là, bien que minime, reste nouveau.

Ce type de recherche sera possible grâce aux object properties, que nous nous apprêtons à aborder. Aussi, toutes les data properties ont le caractère Functional sauf langue et origine, qui peuvent avoir plusieurs occurrences pour un individu (contrairement aux data properties avec le caractère Functional, donc).

## Object properties

Nous disposons maintenant de classes, d'individus et de data properties. Afin de finaliser notre ontologie, il nous reste à créer des liens entre nos individus d'un côté, et nos classes de l'autre. Comme pour les classes, il suffit de regarder les annotations dans l'ontologie pour obtenir toutes les

informations nécessaires sur les différents object properties, qui sont affichés dans la figure 2.4.

Ces propriétés ont été créées de manière à pouvoir aller dans les deux sens (d'un objet A à un objet B, mais aussi de l'objet B à l'objet A). Chaque object property a donc son object property inversant la relation, grâce au paramètre *Inverse Of*. Aussi, l'une des propriétés (*RealisePar*, qui pour un film indique son réalisateur) bénéficie du caractère *Functional*, parce qu'elle ne peut être utilisée une seule fois par individu. Bien qu'un film puisse avoir plusieurs réalisateurs, nous avons voulu outrepasser ce fait afin d'offrir plus de diversité dans notre ontologie. De ce fait, la propriété inverse de *RealisePar*, nommée *ARealise*, a le caractère *Inverse Functional*, ce qui signifie que l'inverse de cette propriété ne peut être utilisée qu'une fois par individu.

Aussi, nos object properties bénéficient des caractères *Asymmetric* et *Irreflexive*, ce qui signifie que les individus n'ont pas de rapports équivalents avec les object properties (acteur A *AJoue* personnage B, mais B n'a pas joué A) et que les individus ne sont pas reliés à eux même via les propriétés.

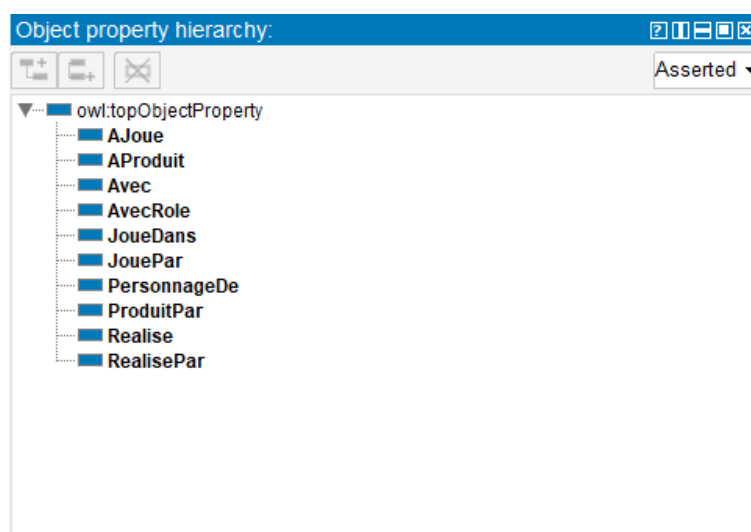


FIGURE 2.4 – Les object properties de notre ontologie

### 2.1.3 Consistance de notre ontologie

Afin de vérifier que notre ontologie est bien cohérente nous avons vérifié sa consistance à l'aide d'Hermit (vu en cours) et de PELLET (demandé dans la fiche de projet).

Aucune erreur de quelque sorte n'a été renvoyée, on peut en conclure que notre ontologie est consistante.

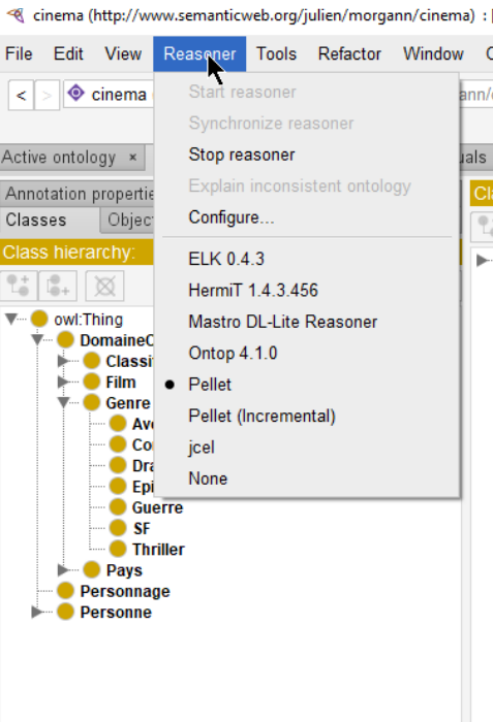


FIGURE 2.5 – Exécuter le reasoner Pellet

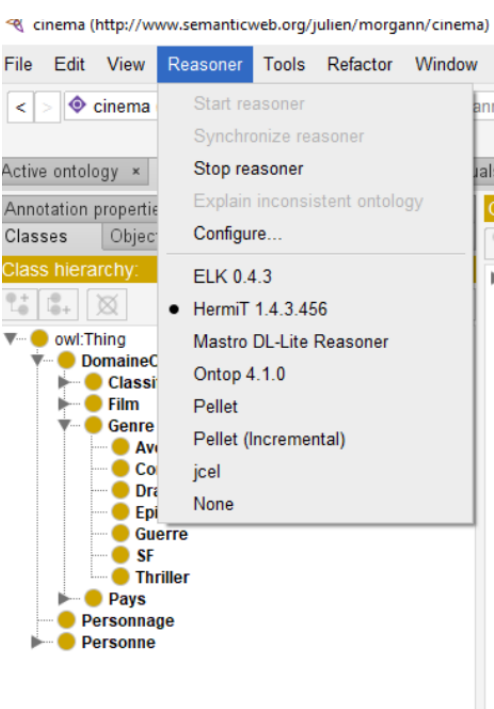


FIGURE 2.6 – Exécuter le reasoner HermiT

## 2.2 Requêtes SPARQL

Les requêtes SPARQL permettent de faire des recherches au sein de notre ontologie. Pour ce faire, on utilise des préfixes, des variables et des conditions. Dans un système de recommandations, elles permettent à l'utilisateur de rechercher un film, un acteur, un réalisateur, selon des critères désirés.

Dans cette partie, nous listerons les requêtes SPARQL que nous avons effectué et leur fonction.

Afin d'observer des données, nous allons commencer par des requêtes très simples.

Nous recherchons dans un premier temps toutes les femmes de notre ontologie. Nous avons d'abord effectué des requêtes très simples qui nous ont permis de vérifier le bon fonctionnement de notre ontologie.

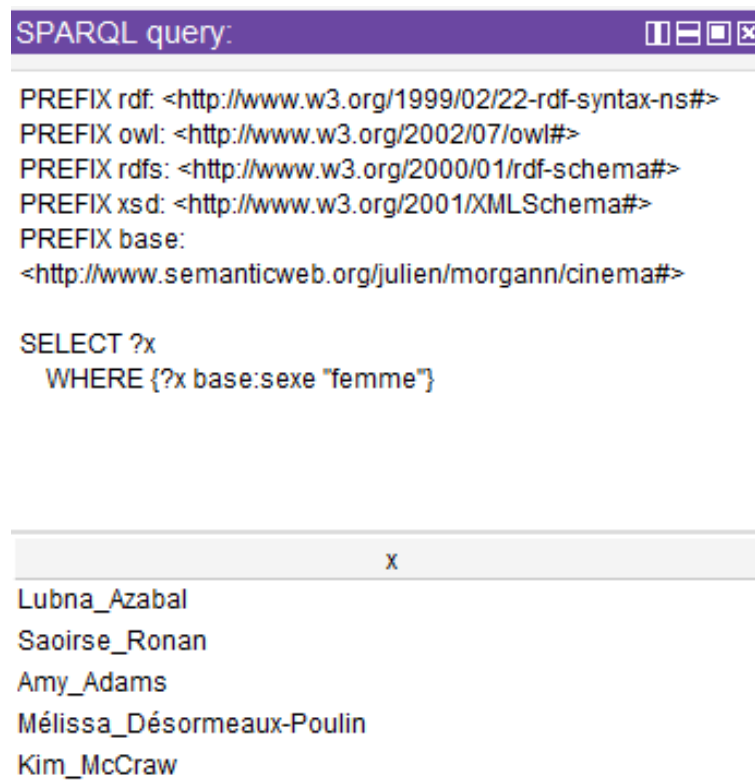


FIGURE 2.7 – Requête et résultat de la requête permettant de trouver toutes les femmes de notre ontologie

On a pu notamment vérifier l'intégralité des classes de nos instances. On fait le test avec toutes les classes de Dune.



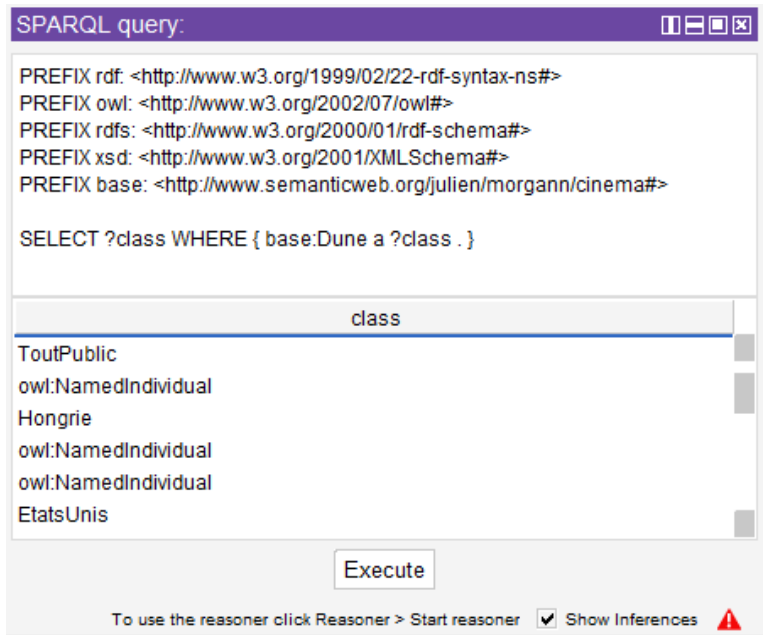


FIGURE 2.8 – Requête et résultat de la requête permettant de trouver toutes les classes attribuées à l’instance Dune

Ensuite, nous pouvons complexifier les requêtes en recherchant tous les films postérieurs à 2010, on choisit de les classer par date dans l’ordre croissant. On obtient alors :

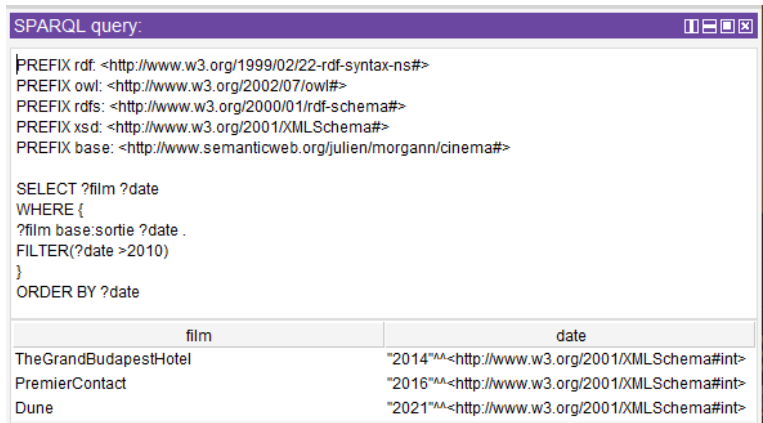


FIGURE 2.9 – Requête et résultat de la requête permettant de trouver les films sortis après 2010

La méthode FILTER nous permet de choisir des films en fonction d’une caractéristique. Ici, la date doit être supérieure à 2010. ORDER BY va per-

mettre de classer les résultats. Cette requête pourrait être effectuée par un utilisateur en quête de films récents.

Il est possible de rechercher un film en fonction de son genre. On peut renvoyer de nombreuses informations le concernant. Ici, nous avons choisi de renvoyer le nom de son réalisateur.

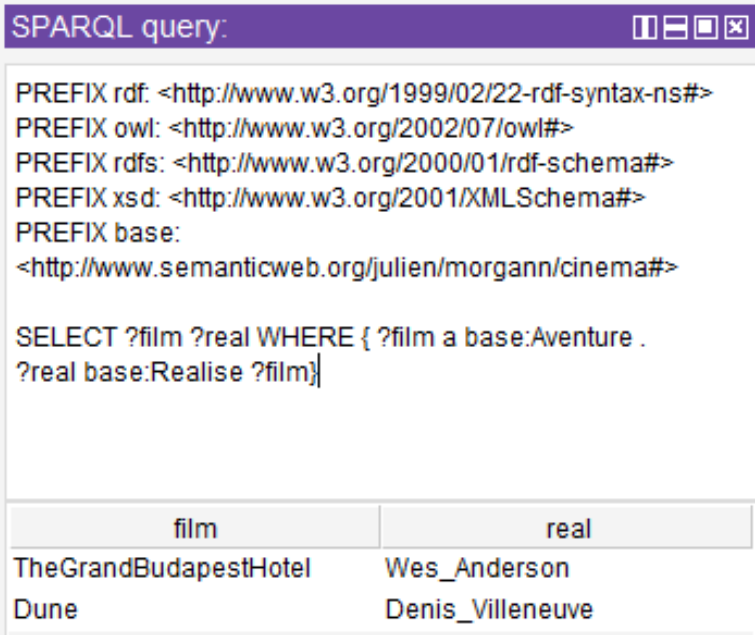


FIGURE 2.10 – Requête et résultats de la requête permettant de trouver les films d’aventure et leur réalisateur

Un autre outil intéressant des requêtes SPARQL réside dans la méthode ASK qui permet de vérifier si une affirmation est vraie ou fausse. Par exemple, si l'utilisateur souhaite voir le film Dune, il peut s'informer à savoir si ce film convient à son genre favori (Aventure) et pour sa famille (ToutPublic).



FIGURE 2.11 – Requête et résultats de la requête permettant de vérifier l'affirmation "Dune est un film d'aventure tout public"

La fonction ASK pourrait donc permettre à un potentiel utilisateur de vérifier que ce qu'il pense est juste ou une information. On peut évidemment vérifier toutes sortes d'information, qu'il s'agisse de la date de sortie d'un film ou de la nationalité d'un réalisateur.

## 2.3 Conclusion

Ce projet nous a permis de nous rendre compte de l'étendu des avantages de structurer une grande quantité d'informations sous forme d'ontologie. Nous avons pu mettre en valeur les aspects essentiels des films de nos deux réalisateurs et des relations qui les lient.

Notre ontologie est consistante et pertinente mais pourrait cependant être améliorée. En effet, nous avons limité d'entrée le nombre d'instances, de classes, et de propriétés, pour ne pas nous perdre dans l'étendue titanesque d'informations entourant généralement un film, tout en essayant de créer quelque chose de simple à comprendre et d'utile. Nous avons peiné à essayer de démarquer cette ontologie, et la mise en place d'un système de tri des personnages selon leur caractère antagoniste / protagoniste est le résultat que nous avons obtenu.

# Bibliographie

- [Camille Mouhamadou Saliou Diallo, 2013] Camille Mouhamadou Saliou Diallo, Moussa Lo, C. T. D. F. K. S. (2013). *Etat de l'art sur ontologies et extraction de connaissances*. PhD thesis, Université Gaston Berger de Saint-Louis(UGB).
- [Giboin, 2018] Giboin, A. (2018). Étude de l'évolution du modèle de l'utilisateur des systèmes de construction collaborative d'ontologies. In Ranwez, S., editor, *29es Journées Francophones d'Ingénierie des Connaissances, IC 2018*, 29es Journées Francophones d'Ingénierie des Connaissances, IC 2018, pages 197–212, Nancy, France. AFIA.
- [Guarino, 1997] Guarino, N. (1997). Semantic matching : Formal ontological distinctions for information organization, extraction, and integration.
- [Hoekstra, 2009] Hoekstra, R. (2009). Ontology representation : design patterns and ontologies that make sense. *IOS Press*.
- [MovieLabs, 2018] MovieLabs (2018). A creative works ontology for the film and television industry. MovieLabs.
- [Sánchez and Pérez, 2018] Sánchez, J. A. P. and Pérez, T. S. (2018). A conceptual model for an owl ontology to represent the knowledge of trans-media storytelling. "Conference : 15th International ISKO Conference (2018)".