

Projet/TP à rendre pour le 13/01/2025
Modèle Sémantique simple pour la prédiction de notes de film

L'objectif de ce devoir est de créer un modèle sur Protege permettant de décrire des FILMS, des individus et des avis sur ces films, de peupler ce modèle avec les données issues de données tabulaires et de définir des règles applicables sur la ABox permettant de prédire l'avis d'une personne sur un film, règles que vous évalueriez.

Les données sont issues de MovieLens <https://grouplens.org/datasets/movielens/100k/> ,

Le jeu de données contient :

- * 100,000 ratings (de 1à5) de 943 personnes sur 1682 films.
- * Chaque personne a noté au moins 20 films.
- * Des informations démographiques sur les personnes users (age, gender, occupation, zip)

- 1) Représenter graphiquement ce modèle de données sur papier : classes, propriétés (object property pour des propriétés entre classes ou datatype property pour les propriétés d'une classe ayant des valeurs littérales) et axiomes. Vous pourrez utiliser les identifiant comme URI pour les individus et créer des URI pour les éléments du modèle). Le réaliser sur Protege.
 - 2) Réaliser un programme permettant de Transformer un extrait des données de u1.base en RDF/XML conformément avec votre modèle (a peu près 500 notations, trop de données vous poserait un problème sur protege).
Attention : Utiliser les bonnes URI pour les classes, et utiliser les identifiants des films et individus dans les données tabulaires pour créer automatiquement les URI (et les recréer facilement quand elles sont mentionnées dans un autre tableau).
 - 3) Importer ce fichier dans Protege (active ontology, importation à partir d'un fichier local).

Ajouter les déclarations de propriétés OWL (datatypeproperty ou objectproperty voir les sources de la mini ontologie du tutoriel si vos propriétés apparaissent en annotation property sous protege).
 - 4) Ajouter les liens de subsumption si besoin, ajouter d'autres axiomes si besoin (disjonction entre classes par exemple).
 - 5) Requêter vos données en SPARQL : Descriptifs complet des films et des individus qui ont noté ces films par 5.
- a) Ajouter dans le modèle des classes et propriétés permettant de modéliser des notations prédites (pour les séparer des notes réelles).
 - b) Créer un ensemble de règles de la logique du premier ordre (3 ou 4) permettant de prédire la note d'un film pour un individu et évaluer la qualité des règles (confiance, support).

Annexe : syntaxe swrl sur <https://dior.ics.muni.cz/~makub/owl/#swrl>

Vous veillerez à introduire dans votre modèle quelques contraintes de base permettant de garantir une certaine cohérence de vos données (disjonctions, propriétés fonctionnelles – i.e. propriétés ayant une seule valeur ...).

Vous déposerez : le fichier OWL/RDF de protégé, et un petit document présentant une brève explication de vos données choix (modèle, règles), les résultats obtenus (qualité des règles).