
THÈME

Construction d'une Ontologie pour la Recommandation de Films à un Utilisateur



Latifa Baba-Hamed, Soltani Réda et Sabri Kamel
Université d'Oran, Faculté des Sciences, Département
lbadahamed@yahoo.fr

Plan de travail

- Introduction
- Les systèmes de recommandation
- Les ontologies
- Architecture générale du système
- Construction de l'ontologie
- Définition des classes et de la hiérarchie des classes
- Définition des propriétés des classes
- Conclusion et perspectives



Introduction

Une **ontologie** est la spécification d'une conceptualisation d'un domaine de connaissances, c'est-à-dire, qu'il est possible de classier des éléments d'un domaine donné par concepts et sous-concepts et de créer tout type de relations entre les éléments et les concepts.

Nous construisons une **ontologie pour filtrer des requêtes utilisateur dans le domaine du cinéma.**

Les systèmes de filtrage

Définition

Un système de filtrage (ou système de recommandation), est un système qui collecte, sélectionne, classe et suggère à l'utilisateur les informations qui répondent vraisemblablement à ses intérêts à long terme.

Types de systèmes de filtrage

- Filtrage collaboratif
- Filtrage basé sur le contenu

Les systèmes de filtrage

Filtrage collaboratif

Il compare les utilisateurs entre eux sur la base de leurs jugements passés pour créer des communautés, et chaque utilisateur reçoit les documents jugés pertinents par sa communauté.

Filtrage basé sur le contenu

Il compare les nouveaux documents au profil de l'utilisateur, et recommande ceux qui sont les plus proches. Cette classe de systèmes utilise des **mesures de similarité** pour définir le degré de ressemblance entre concepts.

Les systèmes de filtrage

Classification des mesures de similarité

- Approches basées sur les espaces vectoriel

Mesure de Pearson :

$$\text{Sim}(x,y) = r_p = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Mesure de Liang & Conrath :

Distance (c1, c2) = $\frac{1}{2} (\text{CI}(c1) + \text{CI}(c2))$

Mesure de Jaccard : Avec : $\text{Sim}(X,Y) = \frac{|X \cap Y|}{|X \cup Y|}$


$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$

PPG : plus petit généralisant


$\text{CI}(c) = -\log(P(c))$

$P(c) = \frac{\text{Nbr d'instances}(c)}{\text{Nbr totale d'instances}}$

0.09




Comédie




Romance


0.12




Musical



Mystère



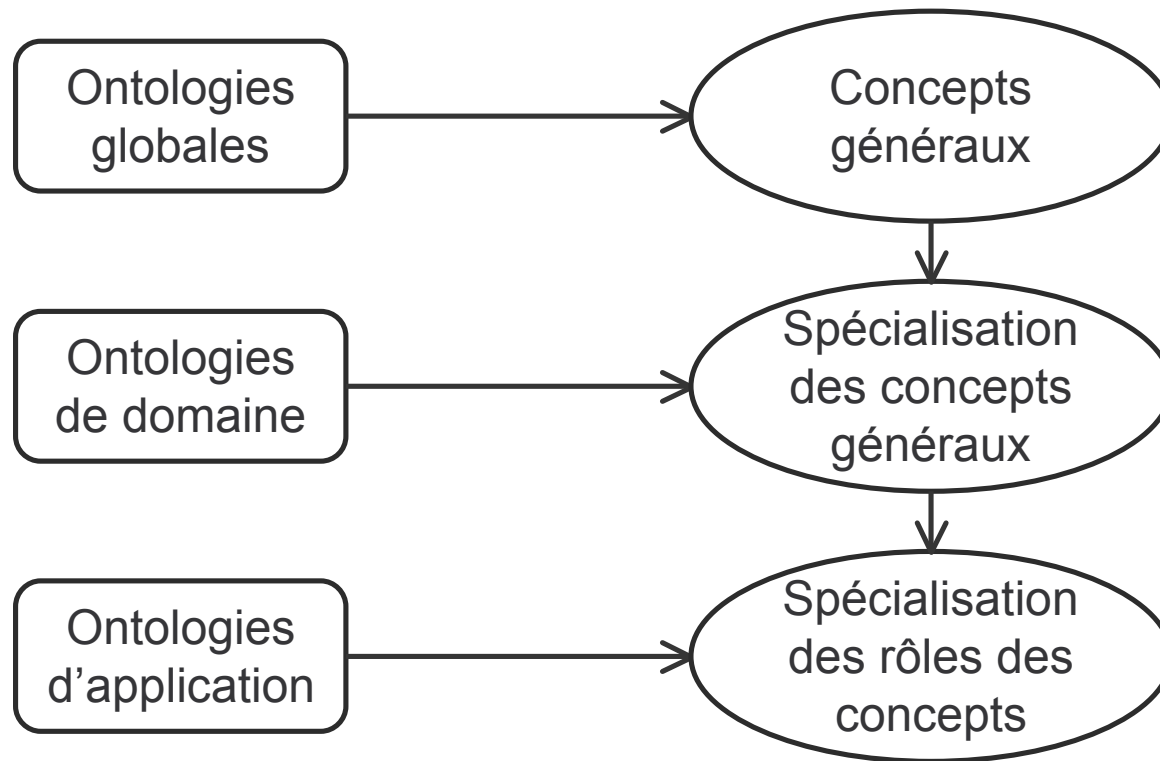
Horreur



Film-Noir

Les ontologies

Degré
d'abstraction



Types d'ontologies

Les ontologies

Méthodologie de construction

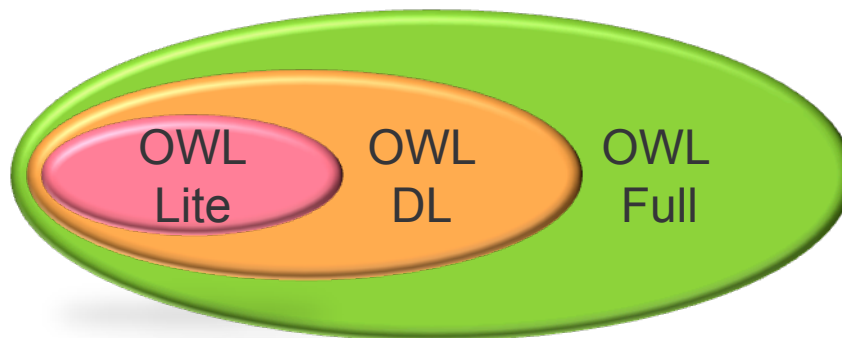
- Spécifier l'application
- Organiser les termes en utilisant les métas catégories
- Affiner l'ontologie et la structurer

Les ontologies

Le langage OWL

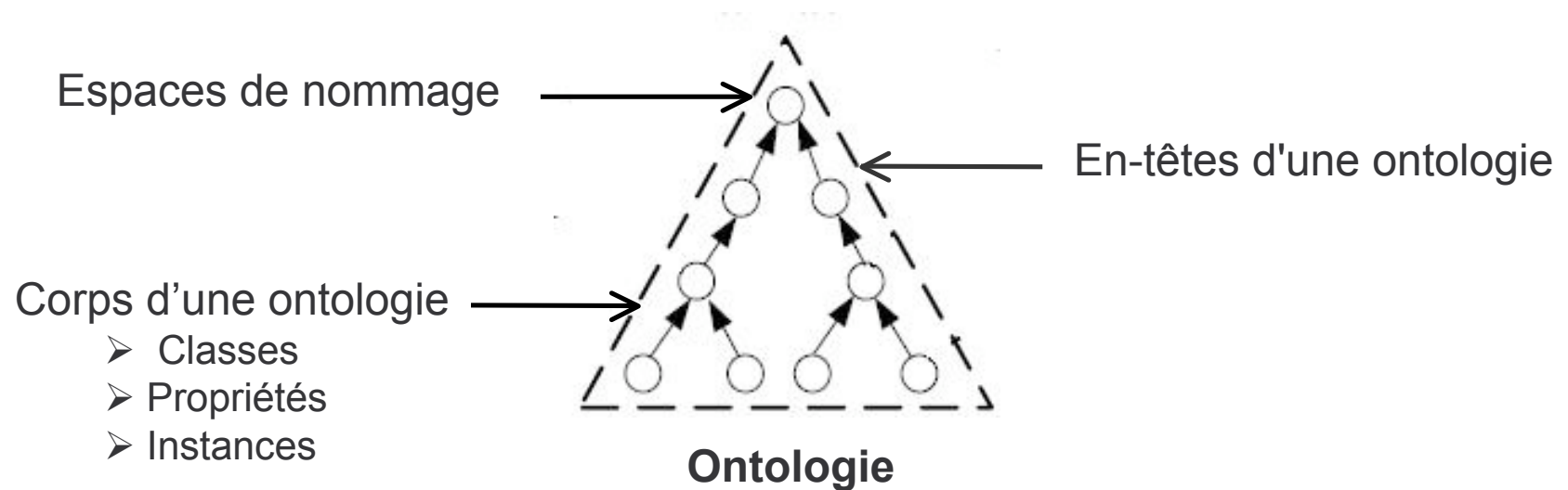
OWL (w3c : <http://www.w3.org>) est une représentation en XML basé sur la syntaxe RDF d'une ontologie. Il permet de définir des ontologies.

Sous-langages d'Owl

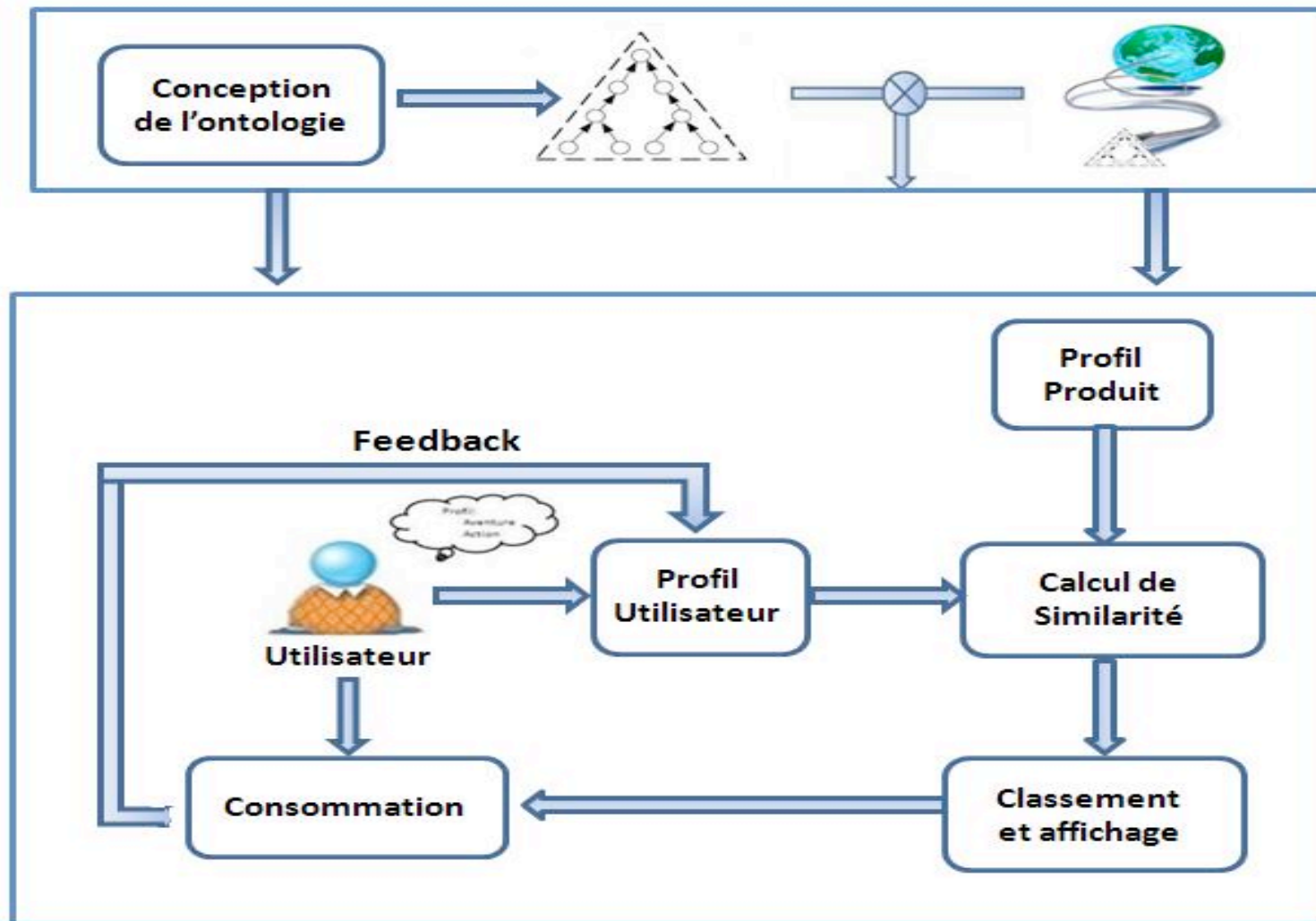


Les ontologies

Structure d'une ontologie écrite en OWL



Architecture Générale du Système



Architecture Générale du Système

Descriptif du profil utilisateur

- Le profil de l'utilisateur est représenté par un ensemble de prédicats pondérés: $\{(P1, w1), (P2, w2), \dots\}$.
- Un prédicat est de la forme $\langle \text{attribut } \theta \text{ valeur} \rangle$
où: attribut est un élément du profil,
 θ est un opérateur de comparaison,
valeur est une valeur ϵ au domaine de l'attribut du profil.
Par exemple: $P1 : \langle \text{Film.genre, '=', action} \rangle$
- Un utilisateur U1 qui a une grande préférence pour les films d'action peut attribuer à ce prédicat une pondération (poids) $w1=0,9$ par exemple.
- Le profil de l'utilisateur U1 sera donc constitué des paires d'éléments : $(P1, w1), (P2, w2), \text{ etc.}$
- Nous procédons de la même façon pour représenter le descriptif d'un film.

Construction de l'ontologie

Notre application nécessite une ontologie de films, qui doit nous permettre de calculer le degré de ressemblance entre des genres de films (ie faire le matching entre les genres de films préférés par un utilisateur et les genres associés à un film donné [Soltani & Sabri, 2009]) dans le but de recommander une liste de films à un utilisateur.



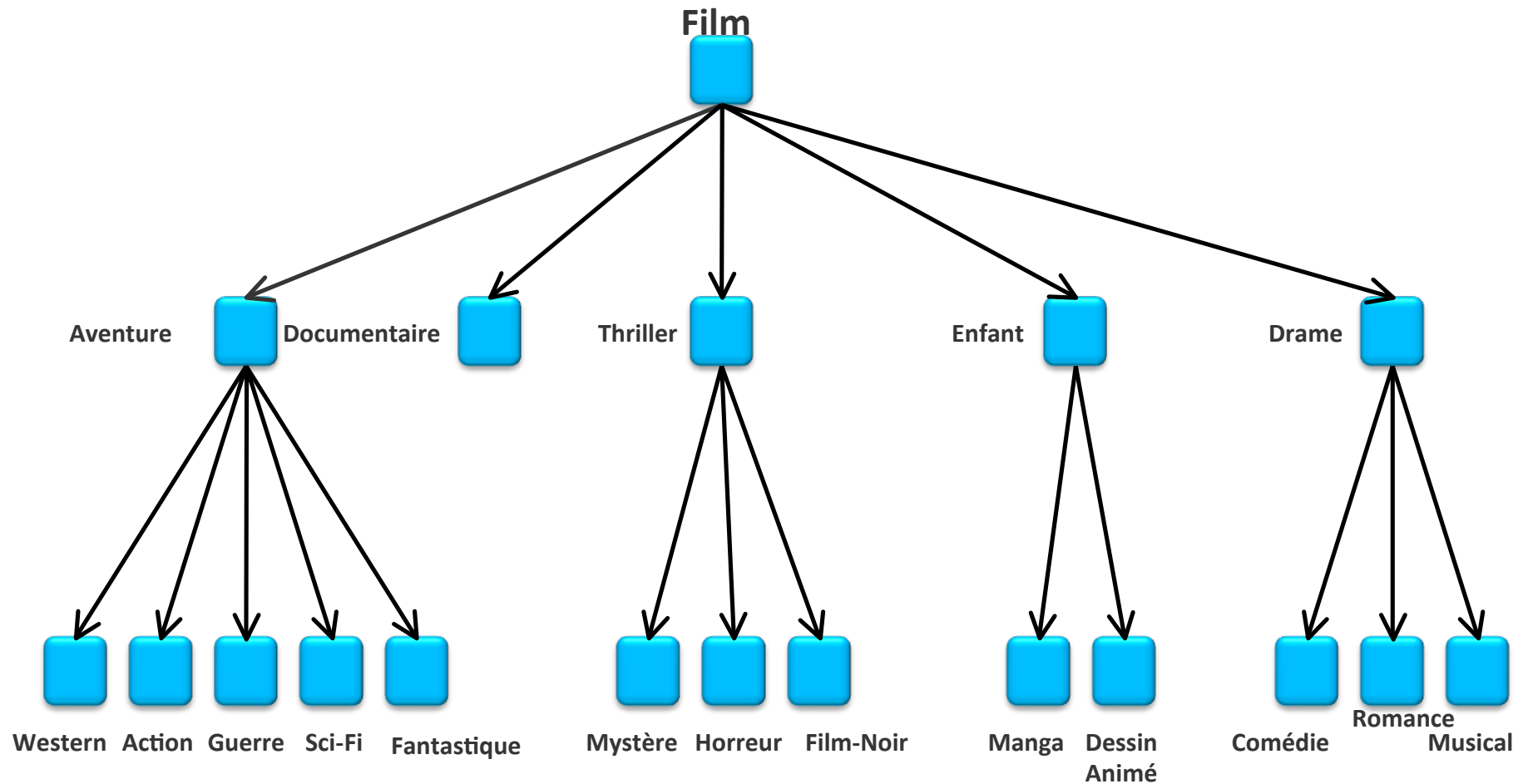
Nous pouvons limiter le domaine de notre ontologie aux genres de films, aux propriétés des films et leurs valeurs.

Construction de l'ontologie

Définition des classes et de la hiérarchie des classes

- Recensement de tous les genres de films en s'inspirant des données des sources [IMDb](http://www.imdb.com/) (site : <http://www.imdb.com/>) , et [MovieLens](http://movielens.umn.edu/) (site : <http://movielens.umn.edu/>).
- Classement des genres de films en classes et sous-classes formant ainsi une hiérarchie de classes ayant pour racine la classe Film.
- Ces classes constituent les concepts de notre ontologie.
- Nous avons retenu 18 genres de films : *Aventure, Action, Dessin-Animés, Enfant, Manga, Comédie, Documentaire, Drame, Fantastique, Film-Noir, Horreur, Musical, Mystère, Romance, Science-Fiction, Thriller, Guerre, Western.*
- Pour établir la hiérarchie des classes, nous avons procédé de **haut en bas** en commençant par les concepts les plus généraux et en terminant par la spécialisation des concepts.

Construction de l'ontologie



Représentation hiérarchique de l'ontologie des films

Construction de l'ontologie

Définition des propriétés des classes

- Il existe deux types de propriétés :
 - ❖ les **propriétés d'objet** (qui permettent de relier des instances à d'autres instances),
 - ❖ les **propriétés de type de donnée** (qui permettent de relier des individus à des valeurs de données).
- Une propriété d'objet est une instance de la classe `owl:ObjectProperty`.
- Une propriété de type de données est une instance de la classe `owl:DatatypeProperty`.
- Les deux classes `owl:ObjectProperty` et `owl:DatatypeProperty` sont des sous-classes de la classe **Property**. Classement des genres de films en classes et sous-classes formant ainsi une hiérarchie de classes ayant pour racine la classe `Film`.

Construction de l'ontologie

Définition des propriétés des classes

- Nous n'avons utilisé, dans notre cas, que des propriétés de type (pas de propriétés d'objet).
- Une propriété typée peut faire correspondre un type aux instances de la classe.

Par exemple, la propriété *Année-de-Production* (resp. *Acteurs*) fait correspondre aux instances de la classe Film le *type entier positif* (resp. le *type chaîne de caractères*) représentant *l'année de production* (resp. *la liste des acteurs*) du film.



Propriétés des concepts de l'ontologie

Construction de l'ontologie

Définition de restriction sur les propriétés

- Nous pouvons également exprimer des restrictions sur les propriétés.

Exemple:

- le fait qu'un film doit avoir une seule année de production (exprimée avec Protégé comme : *Année-de-Production exactly 1*).
- Le fait qu'un film doit avoir deux réalisateurs au maximum (exprimée par : *Réalisateur max 2*).

Construction de l'ontologie

Environnement de développement

- JBuilder 2005
- Protégé (<http://protege.stanford.edu/>)
- Jena (<http://jena.sourceforge.net/>)

Conclusion et Perspectives

- La méthode de construction de l'ontologie est manuelle
- La difficulté de l'accès à l'ontologie de films existante nous a empêchés de la réutiliser, ce qui nous a amenés à construire notre propre ontologie.
- Notre système est facilement adaptable à d'autres ontologies de produits différents de films, comme par exemple les livres.
- L'évaluation de notre système est en cours de validation.

**MERCI DE VOTRE
ATTENTION**