

ONTOLOGIES

Représentation de connaissance terminologiques

LRC - 2025/2026

Gauvain Bourgne

Représentation des connaissances

Système d'informations - stockage de connaissances

E.g. : Bases de données, programme Java, Web sémantique

- Connaissances sur des objets et leur relations
 - Schéma d'information (schéma de la base de donnée, diagrammes UML)
 - Données (informations sur individus/produits...)
- Raisonnement (permis par représentation formelle)
 - Vérifier l'intégrité des données
 - Dériver de nouvelles données
 - Trouver les individus tels que...

Traitement du Langage Naturel

Besoin de traduire du texte en des données structurées et intelligibles : basées sur un vocabulaire explicite, avec une sémantique claire et sans ambiguïté

Systèmes à base de connaissances

Exploiter une représentation formelle des connaissances pour résoudre des problèmes.

■ Représentation

- Des règles générales (e.g. telle maladie cause tel symptôme)
- Des faits spécifiques à des individus (e.g. le patient X a des rougeurs et de la fièvre).

■ Raisonnement

- Un système d'inférence (avec algo efficace)

Au delà de la logique classique

Formalisme de bases des bases de connaissances : logique classique (prop. ou LPPO).

Deux directions (compromis expressivité/complexité) :

- Définir des *extensions* pour manipuler des notions plus complexes (temps, modalités, préférences...)
- Définir des *fragments* plus simples. Pour 2 raisons :
 - Complexité : réduire la complexité des tâches de raisonnements associées
 - Acquisition des connaissances : facilité à formaliser, retrouver, partager et maintenir les connaissances

Les travaux de recherche autour des ontologies se concentre sur la définition de fragments décidables.

Type de connaissances

- Nature de l'objet sur laquelle porte la connaissance
 - ... ontique : connaissance qui se rapporte aux objets déterminés du monde, à ce qui est. Représentation d'entités et de relations comme modélisation d'une réalité objective
 - ... épistémique : connaissance sur le savoir. En informatique, représentation des états mentaux (qui sait/croit quoi).
 - ... déontique : connaissance qui se rapporte à ce qui devrait être.
- Temporalité
 - ... statique
 - ... dynamique
- Point de vue : structurel vs fonctionnel

Ontologie : *spécification formelle explicite d'une conceptualisation partagée d'un domaine donné*

- A minima, structurer et définir les termes du domaine : quels mots et avec quel sens ?
 - un vocabulaire : ensemble de termes (ou symboles) dotés d'un statut formel précis (classe, propriété...)
 - souvent, une notion de spécialisation/généralisation : organisation des termes du vocabulaire en une hiérarchie
 - des règles ou contraintes sur ces termes
 - utilisé par
 - Gestion des connaissances
 - Intégration sémantique des données
 - Recherche d'information
- ⇒ Web sémantique

Ontologie formelle : *vocabulaire conceptuel avec une sémantique précisé par un formalisme logiquement fondé*

■ Représentation

Formellement, une ontologie est un couple (V, O)

- V , le vocabulaire formel, est un ensemble typé de symboles non logiques (aussi appelé *signature*)
- O est une ensemble de formules logiques construites sur ce vocabulaire (à partir d'un ensemble connu d'opérateurs) qui vise à contraindre l'ensemble des modèles acceptables du domaine considéré

■ Raisonnement

- Limité aux modèles satisfaisant O
(ex. $G \models_{(V,O)} F$ devient $O, G \models F$)
- classiquement, *hypothèse du nom unique* : des constantes différentes doivent être interprétées par des éléments différent du domaine d'interprétation

Base de connaissance dirigées par une ontologie

Deux grands types de connaissances

- des connaissances ontologique proprement dites, les *axiomes ontologiques* : connaissances générales sur le domaine ayant vocation à fixer la sémantique du vocabulaire formel
- des connaissances particulières, les *assertions* (ou *faits*) : description des situations spécifiques relatives à des entités individuelles du domaine.

Le terme 'ontologie'

Dans les travaux sur OWL, on dénomme 'ontologie' l'ensemble de ces 2 types de connaissances, tandis que dans les autres domaines, 'ontologie' représente le vocabulaire formel et les axiomes ontologiques.

KB dirigé par ontologie vs BD

KB dirigée par ontologie

Composants

vocabulaire formel
et axiomes ontologiques

assertions

Hypothèses

Monde ouvert

Objectif

Inférer de nouvelles
connaissances à partir
de celles saisies

Travaux convergents (Web sémantique)

Inférence sur de
grandes bases de fait

Base de données

schéma de la base
et contraintes

données

Monde clos et Domaine clos

Stocker et
récupérer
des connaissances

Interrogation
en monde ouvert

Raisonnement avec ontologies

On considère un ensemble d'axiomes ontologiques O et un ensemble d'assertions A .

Raisonnement terminologique

Inférer des connaissances générales (compléter les règles).

Déduire de O un axiome ontologique o tel que

$$O \models o$$

Raisonnement assertionnel

Inférer des connaissances spécifiques à des individus (appliquer les règles aux situations).

Déduire de O et A une assertion a tel que

$$O, A \models a$$

Raisonnement avec ontologies

On considère un ensemble d'axiomes ontologiques O et un ensemble d'assertions A .

Interrogation d'une base d'assertion en présence d'ontologie

Trouver les individus qui correspondent à une requête.

Déterminer à partir de O, A et d'une requête q une substitution s des variables libre de q par des entités individuelles tel que

$$O, A \models s(q)$$

Ce raisonnement peut se ramener au raisonnement assertionnel (par énumération).

Deux grandes familles de formalismes pour représenter des bases de connaissances dirigées par des ontologies.

Logiques des descriptions

Focus Raisonnement par classification (recherche d'élément plus spécifiques ou généraux qu'un élément donné)

Idee Permettre des descriptions complexes de classes ou relations et construire bases d'axiomes et de faits. Famille de logiques selon les axiomes ontologiques utilisables (expressivité vs complexité)

Pt fort Riches expression d'axiomes ontologiques

Pt faible Ne permet que des requêtes d'instance (pas d'interrogation directe de données)

Ext. LD dites légères pour permettre interrogation efficace.

Deux grandes familles de formalismes pour représenter des bases de connaissances dirigées par des ontologies.

Représentation par graphes (graphes conceptuels)

Focus Dédutions entre assertions

Idée Représenter graphiquement les connaissances assertionnelles en leur associant une taxonomie

Pt fort Riche expression de requêtes (permet n'importe quelle requête existentielle conjonctive)

Pt faible Ne permet de représenter que des axiomes ontologiques simples (e.g. taxonomie)

Ext. Règles de graphe pour représenter les principaux axiomes ontologiques

Logiques de description – *ALC*

Logiques de descriptions

Qu'est-ce ?

Les logiques de descriptions sont

- des fragments (décidables) de la LPPO : toute 'formule' de LD peut se traduire en formule LPPO sur des prédicats unaires (\rightarrow concepts) et binaires (\rightarrow rôles).
- une famille de logiques plus ou moins expressives (et plus ou moins complexes), qui diffèrent par
 - les opérateurs disponibles
 - les formules autorisées
- la base théorique des standards du Web sémantique (en particulier OWL, langage de description d'ontologies)

KB en LD

Une bases de connaissance $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ en LD est divisée en deux parties :

- la Tbox \mathcal{T} : partie intensionnelle (axiomes ontologiques)
- la Abox \mathcal{A} : partie assertionnelle

La Tbox définit le modèle conceptuel des données de la Abox.

Les connaissances son exprimées à l'aide de trois type de symboles

- Concepts : classes d'entités (prédicats unaires)
- Roles : relations/propriétés (prédicats binaires)
- Individus : les éléments ou instances décrits (constantes)

Syntaxe des logiques de description

Signature et constructeurs

- La signature $(\mathbb{C}, \mathbb{R}, \mathbb{I})$ définit le vocabulaire (symboles non logiques)
 - \mathbb{C} : ensemble des concepts atomiques
 - \mathbb{R} : ensemble des rôles atomiques
 - \mathbb{I} : ensemble des constantes
- Les constructeurs, dépendant du type de LD, permettent de former des expression de concepts ou de rôles (exemples : $C \sqcap D, \neg C, R^{-1}$)

Exemple façon IMDb

Concepts : *Film, Série, Acteur, Réalisateur, Avis,...*

Rôles : *joueDans, réalise, épisodeDe, critiquéPar,...*

Individus : *fightclub, bradPitt.*

Syntaxe des logiques de description

Forme des connaissances

- Tbox : ensemble d'axiomes
 - inclusions entre concepts $C \sqsubseteq D$
 - équivalence entre concepts $C \equiv D$
 - possiblement, relations entre rôles ($R \sqsubseteq E$ ou $R \equiv E$)
- Abox : ensemble d'assertions
 - appartenance de constantes à des concepts $e : C$
 - appartenance à des rôles de couples de constantes $e_1, e_2 : R$

Langage

Le choix de la LD va restreindre

- les constructeurs (expressions de concepts ou rôles permises)
- la forme des axiomes autorisés
- la forme des assertions autorisées

La logique \mathcal{ALC} – Syntaxe

Alphabet

- Signature : ensemble des symboles non logiques
 - \mathbb{C} est l'ensemble des *concepts atomiques*
 - \mathbb{R} est l'ensemble des *rôles atomiques*
 - \mathbb{I} est l'ensemble des *constantes*
- Constructeurs :
 - (arité 0) \top le concept universel et \perp le concept vide
 - (arité 1) \neg la négation de concept
 - (arité 2) \sqcap , \sqcup permettent de combiner deux concepts, \forall , \exists construisent un concept à partir d'un rôle et d'un concept.

Expressions de concept

Soit A un concept atomique, C une expression de concept et R un rôle atomique. Une expression de concept est (forme de Backus-Naur) :

$$C ::= A \mid \top \mid \perp \mid C \sqcap C \mid C \sqcup C \mid \exists R.C \mid \forall R.C$$

Syntaxe

Une base de connaissances en \mathcal{ALC} est donné comme un couple $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ où

- la Tbox \mathcal{T} est un ensemble d'axiomes ontologiques qui peuvent être de deux types (C et D étant des expressions de concepts) :
 - équivalence : $C \equiv D$
 - inclusion : $C \sqsubseteq D$.
- la Abox \mathcal{A} est un ensemble d'assertions qui peuvent être de deux types (a et b étant des constantes, C un concept et R un rôle) :
 - assertion de concept : $a : C$
 - assertion de rôle : $a, b : R$

Syntaxe – Exemple IMDb (TBox)

TBox (ontologie simplifiée)

- $\text{Acteur} \sqsubseteq \text{Personne}$
- $\text{Réalisateur} \sqsubseteq \text{Personne}$
- $\exists \text{joueDans.Film} \sqsubseteq \text{Acteur}$
- $\text{Réalisateur} \sqsubseteq \exists \text{réalise.}(\text{Film} \sqcup \text{Épisode})$
- $\text{RéalisateurDeSérie} \equiv \text{Réalisateur} \sqcap \forall \text{réalise.Episode}$
- $\text{Avis} \sqsubseteq \text{Positif} \sqcup \text{Neutre} \sqcup \text{Négatif}$
- $\text{ActeurConsensuel} \sqsubseteq$
 $\text{Acteur} \sqcap \forall \text{joueDans.} \forall \text{critiquéPar.} \neg \text{Négatif}$
- $\text{Épisode} \sqsubseteq \exists \text{épisodeDe.Série}$
- $\text{Film} \sqcap \text{Épisode} \sqsubseteq \perp$

Syntaxe – Exemple IMDb (ABox)

ABox (assertions)

- `bradPitt : Acteur \sqcap ActeurConsensuel`
- `fightClub : Film`
- `friends : Série`
- `friendsS08E09 : Épisode`
- `(friendsS08E09, friends) : épisodeDe`
- `(bradPitt, fightClub) : joueDans`
- `(bradPitt, friendsS08E09) : joueDans`

Structures

- une structure (ou interprétation) $M = (\Delta^M, \cdot^M)$ définit
 - un domaine d'interprétation Δ^M non vide (ensemble d'individus)
 - une fonction d'interprétation associant à chaque élément de la signature $(\mathbb{C}, \mathbb{R}, \mathbb{I})$:
 - un individu $e^M \in \Delta^M$ pour $e \in \mathbb{I}$
 - un sous ensemble $C^M \subseteq \Delta^M$ pour $C \in \mathbb{C}$
 - une relation binaire $R^M \subseteq \Delta^M \times \Delta^M$ pour $R \in \mathbb{R}$
- Interprétation des expressions de concepts : \cdot^M étendu aux expression de concept à partir de interprétation des concepts et rôles atomiques ;
- Hypothèse du nom unique

$$\forall a, b \in \mathbb{I}, a \neq b \Rightarrow a^M \neq b^M$$

Lorsque cette hypothèse n'est pas faite, on peut avoir des assertions explicite dans la Abox ($a = b$ ou $a \neq b$).

Exemple IMDb – Domaine et individus

Domaine d'interprétation

$$\Delta^M = \{bradPitt, fincher, burrows, fightClub, s8E9, friends, a_1, a_2, a_3, a_4, a_5\}$$

Constantes

- $fightClub^M = fightClub$
- $friends^M = friends$
- $friendsS09E09^M = s8E9$
- $bradPitt^M = bradPitt$

Exemple IMDb – Domaine et concepts

Concepts principaux

$\text{Personne}^M = \{\text{bradPitt}, \text{fincher}, \text{burrows}\}$

$\text{Acteur}^M = \{\text{bradPitt}\}$ $\text{ActeurConsensuel}^M = \{\text{bradPitt}\}$

$\text{Réalisateur}^M = \{\text{fincher}, \text{burrows}\}$

$\text{RéalisateurDeSérie}^M = \{\text{burrows}\}$

$\text{Film}^M = \{\text{fightClub}\}$

$\text{Série}^M = \{\text{friends}\}$

$\text{Épisode}^M = \{\text{s8E9}\}$

$\text{Avis}^M = \{a_1, a_2, a_3, a_4, a_5\}$

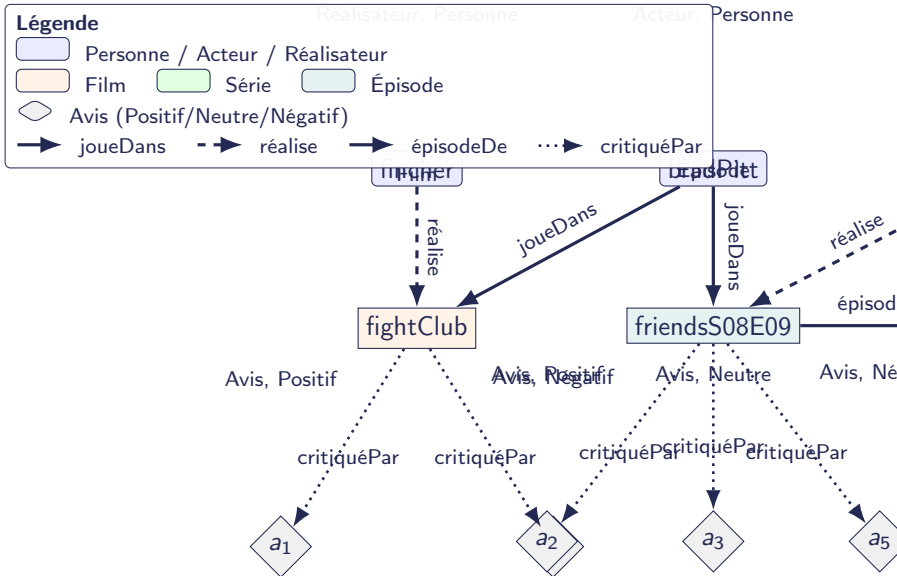
$\text{Positif}^M = \{a_1, a_2\}$ $\text{Neutre}^M = \{a_3\}$ $\text{Négatif}^M = \{a_4, a_5\}$

Exemple IMDb – Interprétation des rôles

Relations principales

- $\text{joueDans}^M = \{(bradPitt, fightClub), (bradPitt, friendsS08E09)\}$
- $\text{réalise}^M = \{(fincher, fightClub), (burrows, friendsS08E09)\}$
- $\text{épisodeDe}^M = \{(s8E9, friends)\}$
- $\text{critiquéPar}^M = \{(fightClub, a_1), (fightClub, a_4), (s8E9, a_2), (s8E9, a_3), (s8E9, a_5)\}$

Exemple IMDb – Représentation graphique



Sémantique (expressions de concept)

Étant donné une structure $M = \langle \Delta^M, .^M \rangle$, on étend la fonction d'interprétation de aux expressions de concepts comme suit

- $\top^M = \Delta^M$ et $\perp^M = \emptyset$
- $(\neg C)^M = \Delta^M \setminus C^M$
- $(C_1 \sqcup C_2)^M = C_1^M \cup C_2^M$
- $(C_1 \sqcap C_2)^M = C_1^M \cap C_2^M$
- $(\exists R.C)^M = \{x \in \Delta^M \mid \exists y \in \Delta^M, ((x, y) \in R^M) \wedge (y \in C^M)\}$
- $(\forall R.C)^M = \{x \in \Delta^M \mid \forall y \in \Delta^M, ((x, y) \in R^M) \rightarrow (y \in C^M)\}$

Exemple IMDb – Interprétation de concepts complexes

Expressions de concept

- $\exists \text{joueDans.Film}^M =$
- $(\exists \text{réalise.}(\text{Film} \sqcup \text{Épisode}))^M =$
- $(\forall \text{réalise.Série})^M =$
- $(\forall \text{réalise.Épisode})^M =$
- $(\text{Réalisateur} \sqcap \forall \text{réalise.Épisode})^M =$
- $(\text{Acteur} \sqcap \forall \text{joueDans.} \forall \text{critiquéPar.} \neg \text{Négatif})^M =$

Exemple IMDb – Interprétation de concepts complexes

Expressions de concept

- $\exists \text{joueDans.Film}^M = \{bradPitt\}$
- $(\exists \text{réalise.}(\text{Film} \sqcup \text{Épisode}))^M =$
- $(\forall \text{réalise.Série})^M =$
- $(\forall \text{réalise.Épisode})^M =$
- $(\text{Réalisateur} \sqcap \forall \text{réalise.Épisode})^M =$
- $(\text{Acteur} \sqcap \forall \text{joueDans.} \forall \text{critiquéPar.} \neg \text{Négatif})^M =$

Exemple IMDb – Interprétation de concepts complexes

Expressions de concept

- $\exists \text{joueDans.Film}^M = \{bradPitt\}$
- $(\exists \text{réalise.}(\text{Film} \sqcup \text{Épisode}))^M = \{fincher, burrows\}$
- $(\forall \text{réalise.Série})^M =$
- $(\forall \text{réalise.Épisode})^M =$
- $(\text{Réalisateur} \sqcap \forall \text{réalise.Épisode})^M =$
- $(\text{Acteur} \sqcap \forall \text{joueDans.} \forall \text{critiquéPar.} \neg \text{Négatif})^M =$

Exemple IMDb – Interprétation de concepts complexes

Expressions de concept

- $\exists \text{joueDans.Film}^M = \{bradPitt\}$
- $(\exists \text{réalise.}(\text{Film} \sqcup \text{Épisode}))^M = \{fincher, burrows\}$
- $(\forall \text{réalise.Série})^M = \{bradPitt, fightClub, s8E9, friends, a_1, a_2, a_3, a_4, a_5\}$
- $(\forall \text{réalise.Épisode})^M = \{bradPitt, burrows, fightClub, s8E9, friends, a_1, a_2, a_3, a_4, a_5\}$
- $(\text{Réalisateur} \sqcap \forall \text{réalise.Épisode})^M =$
- $(\text{Acteur} \sqcap \forall \text{joueDans.} \forall \text{critiquéPar.} \neg \text{Négatif})^M =$

Exemple IMDb – Interprétation de concepts complexes

Expressions de concept

- $\exists \text{joueDans.Film}^M = \{bradPitt\}$
- $(\exists \text{réalise.}(\text{Film} \sqcup \text{Épisode}))^M = \{fincher, burrows\}$
- $(\forall \text{réalise.Série})^M = \{bradPitt, fightClub, s8E9, friends, a_1, a_2, a_3, a_4, a_5\}$
- $(\forall \text{réalise.Épisode})^M = \{bradPitt, burrows, fightClub, s8E9, friends, a_1, a_2, a_3, a_4, a_5\}$
- $(\text{Réalisateur} \sqcap \forall \text{réalise.Épisode})^M = \{burrows\}$
- $(\text{Acteur} \sqcap \forall \text{joueDans.} \forall \text{critiquéPar.} \neg \text{Négatif})^M =$

Exemple IMDb – Interprétation de concepts complexes

Expressions de concept

- $\exists \text{joueDans.Film}^M = \{bradPitt\}$
- $(\exists \text{réalise.}(\text{Film} \sqcup \text{Épisode}))^M = \{fincher, burrows\}$
- $(\forall \text{réalise.Série})^M = \{bradPitt, fightClub, s8E9, friends, a_1, a_2, a_3, a_4, a_5\}$
- $(\forall \text{réalise.Épisode})^M = \{bradPitt, burrows, fightClub, s8E9, friends, a_1, a_2, a_3, a_4, a_5\}$
- $(\text{Réalisateur} \sqcap \forall \text{réalise.Épisode})^M = \{burrows\}$
- $(\text{Acteur} \sqcap \forall \text{joueDans.} \forall \text{critiquéPar.} \neg \text{Négatif})^M = \emptyset$

Sémantique – principes généraux

Modèles

M est un modèle de

- partie terminologique

- inclusion $C \sqsubseteq D$ (resp. $R \sqsubseteq E$) si $C^M \subseteq D^M$ (resp. $R^M \subseteq E^M$)
- équivalence $C \equiv D$ (resp. $R \equiv E$) si $C^M = D^M$ (resp. $R^M = E^M$)
- Tbox \mathcal{T} si M est un modèle de tous ses axiomes (d'inclusion ou d'équivalence)

- partie assertionnelle

- assertion $a : C$ (resp. $a, b : R$) si $a^M \in C^M$ (resp. $(a^M, b^M) \in R^M$)
- Abox \mathcal{A} si M est un modèle de toutes ses assertions
- KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ si M est un modèle de \mathcal{T} et de \mathcal{A} .

Sémantique – Restriction à ALC

Modèles

M est un modèle de

- partie terminologique
 - inclusion $C \sqsubseteq D$ si $C^M \subseteq D^M$
 - équivalence $C \equiv D$ si $C^M = D^M$
 - Tbox \mathcal{T} si M est un modèle de tous ses axiomes (d'inclusion ou d'équivalence)
- partie assertionnelle
 - assertion $a : C$ (resp. $a, b : R$) si $a^M \in C^M$ (resp. $(a^M, b^M) \in R^M$)
 - Abox \mathcal{A} si M est un modèle de toutes ses assertions
- KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ si M est un modèle de \mathcal{T} et de \mathcal{A} .

Exemple IMDb – Interprétation de concepts complexes

Exercice

- M est un un modèle de la Abox ? de la TBox ?

Sémantique – principes généraux

Conséquence

- Une Tbox \mathcal{T} entraîne logiquement un axiome α ($\mathcal{T} \models \alpha$) si tout modèle de \mathcal{T} est un modèle de α .
- Une KB \mathcal{K} entraîne logiquement un axiome ou une assertion α ($\mathcal{K} \models \alpha$) si tout modèle de \mathcal{K} est un modèle de α .

Satisfiabilité

- Une KB \mathcal{K} (ou une Tbox \mathcal{T}) est satisfiable (ou cohérente) si elle a au moins un modèle.
- Un concept C est satisfiable selon \mathcal{T} (C \mathcal{T} -SAT) si il existe un modèle M de \mathcal{T} tel que $C^M \neq \emptyset$
- Une Abox \mathcal{A} est cohérente selon \mathcal{T} (\mathcal{A} \mathcal{T} -SAT) si elle possède un modèle commun avec \mathcal{T} (ie si $\langle \sqcup, \mathcal{A} \rangle$ est cohérente).

La logique \mathcal{ALC} – Traduction en LPPO

Traduction en LPPO

- *Langage* : A chaque concept atomique C correspond un prédicat unaire $C(x)$, à chaque rôle atomique R un prédicat binaire $R(x, y)$ et à chaque constante a , une constante du même nom (a).
- *Expression de concept* : pour traduire des expressions de concept, il est nécessaire d'utiliser une fonction τ_x paramétrée par une variable x . Elle est définie inductivement comme suit :
 - $\tau_x(C) = C(x)$ si C concept atomique
 - $\tau_x(\perp) = \perp$, $\tau_x(\top) = \top$
 - $\tau_x(\neg C) = \neg \tau_x(C)$
 - $\tau_x(C \sqcap D) = \tau_x(C) \wedge \tau_x(D)$
 - $\tau_x(C \sqcup D) = \tau_x(C) \vee \tau_x(D)$
 - $\tau_x(\exists R.C) = \exists y(R(x, y) \wedge \tau_y(C))$
 - $\tau_x(\forall R.C) = \forall y(R(x, y) \rightarrow \tau_y(C))$

La logique \mathcal{ALC} – Traduction en LPPO

Traduction en LPPO

- *Expression de rôle* : que des rôles atomiques en \mathcal{ALC} , donc pas besoin de passer par une fonction de traduction $\tau_{x,y}$ (mais ce serait le cas pour les LD plus riches avec des rôles complexes - e.g. inversion de rôle)
- *Tbox* :
 - (inclusion) $C \sqsubseteq D$ se traduit par la formule
$$\forall X(\tau_X(C) \rightarrow \tau_X(D))$$
 - (équivalence) $C \equiv D$ se traduit par la formule
$$\forall X(\tau_X(C) \leftrightarrow \tau_X(D))$$
- *Abox* :
 - (assertion de concept) $a : C$ se traduit par la formule $\tau_a(C)$
 - (assertion de rôle) $a, b : R$ se traduit par la formule $R(a, b)$

Exemple IMDb – Traduction en LPPO

Exercice

- Traduire la TBox et Abox en LPPO