
Introduction aux thesaurus, aux ontologies et au Web sémantique

[http ://www.univ-orleans.fr/Isabelle.Tellier](http://www.univ-orleans.fr/Isabelle.Tellier)

Plan

- Thesaurus
- Ontologies
- Web sémantique : RDF et OWL

Généralités

- origine linguistique : dictionnaire normalisé et hiérarchisé
- recense et organise les termes spécifiques d'un domaine
 - ⇒ dépendant d'une langue, d'un domaine

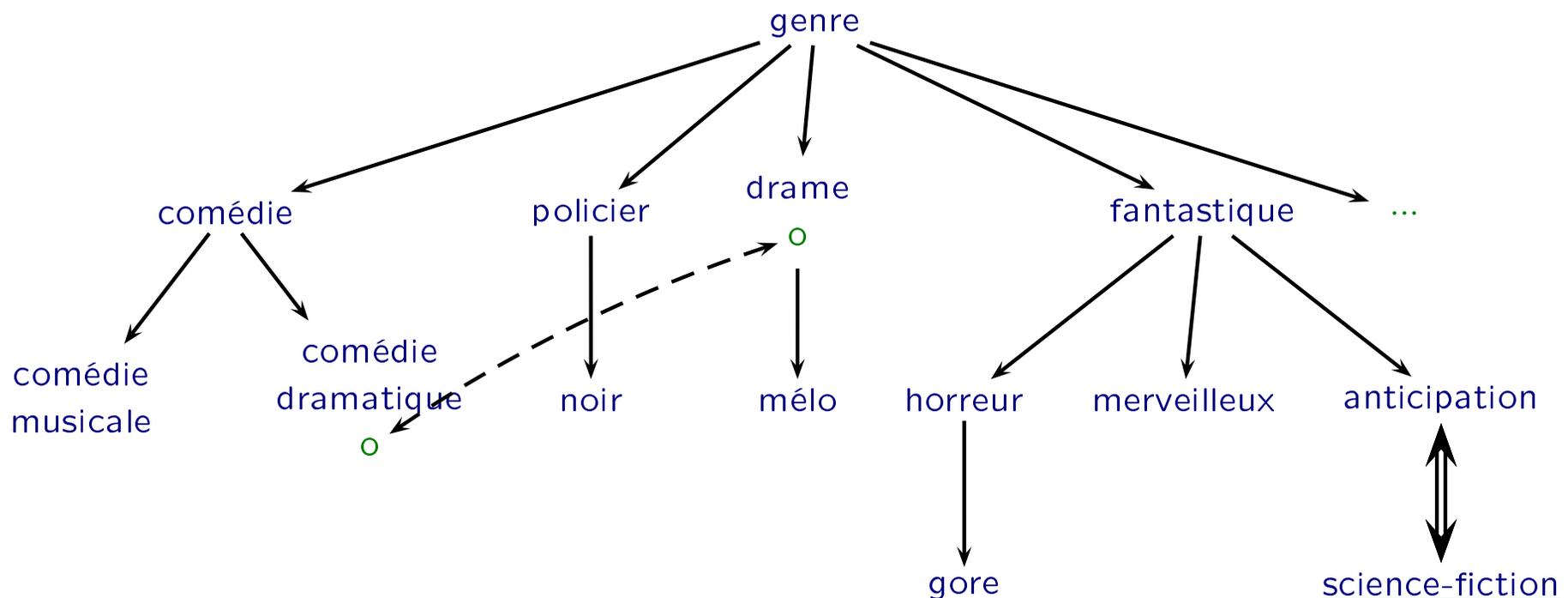
Propriétés

- contient des termes descripteurs et des termes non descripteurs
- un terme est un nom ou un groupe nominal pouvant servir de mot clé dans une indexation
- les termes sont liés entre eux suivant 3 types de relation
 - relation hiérarchique entre descripteurs : terme générique (TG)/terme spécifique (TS)
 - relation d'équivalence entre un terme descripteur et un terme non descripteurs : synonyme, employé pour (EP)
 - relation d'association entre descripteurs : terme associé (TA)

Utilisation

- en indexation de documents : on utilise les termes descripteurs du thesaurus les plus précis possibles
- en recherche d'information : on peut faire de l'expansion de requête en l'utilisant (par exemple en le "remontant" pour généraliser)

Exemple simplifié : genres cinématographiques



Exemples de thesaurus classiques

- thesaurus gouvernement du Québec :
<http://grds.ebsi.umontreal.ca/r/thesaurus-activite-gouvernementale/site-web/categories.htm>
- thesaurus de santé publique interrogeable en ligne :
<http://www.bdsp.tm.fr/TSP3/Default.asp>
- thesaurus multilingue de la communauté européenne :
<http://europa.eu/eurovoc/>

Exemples de thesaurus étendus

- visualisation et navigation dans l'espace :
<http://www.enekia.com/thesaurus.html>
- dictionnaire étendu le plus célèbre (en anglais) : Wordnet
<http://wordnet.princeton.edu/>

Généralités

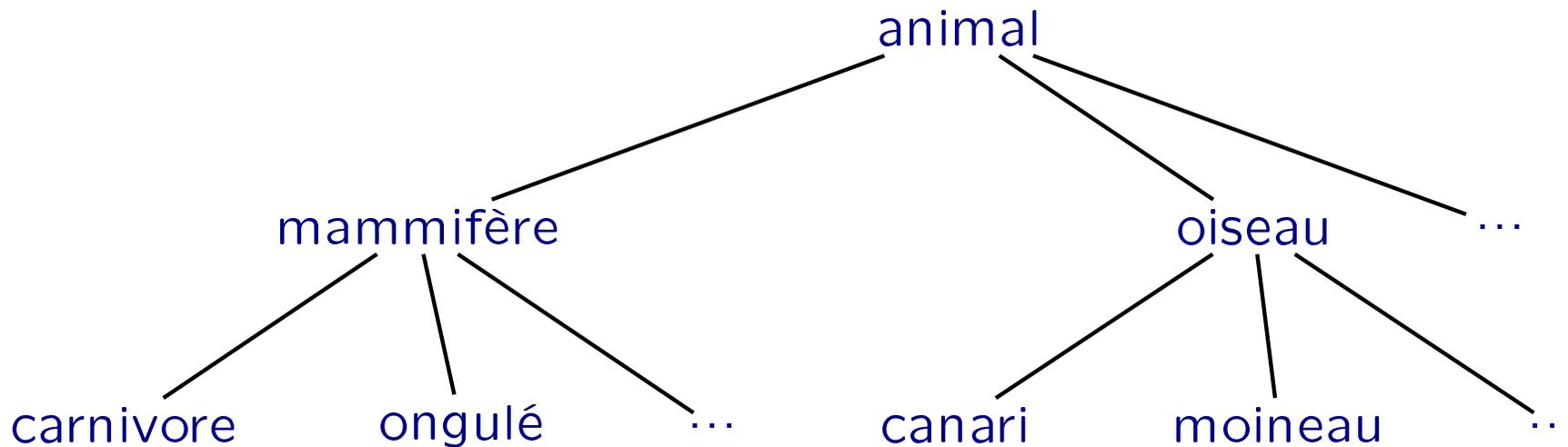
- origine philosophique : science de ce qui est
- censées représenter une théorie du monde
 - ⇒ en théorie indépendantes d'une langue, d'un domaine !

Propriétés

- définition opératoire la plupart du temps : hiérarchie de concepts
- en pratique : arbre dont les noeuds sont les concepts et les arcs une relation de subsomption
- un concept : notion abstraite générale non ambiguë, ne coïncide pas nécessairement avec un mot (ex : “ChickenMeat”)
- relation de subsomption : “is-a”, mais autres relations possibles (“part-of”, “causality”...)
- possibilité d'associer des propriétés aux noeuds, qui se propagent par héritage comme dans les langages orientés objets

- les ontologies de John Sowa sous forme de treillis :
<http://www.jfsowa.com/ontology/index.htm>
- une ontologie et sa logique
<http://virtual.cvut.cz/kifb/en/toc/229.html>
- une ontologie qui se veut universelle SUMO (Suggested Upper Merged Ontology) :
<http://ontologyportal.org/>
- plutôt des taxonomies, SHOE (Simple HTML Ontology Extensions)
<http://www.cs.umd.edu/projects/plus/SHOE/onts/>
- Framenet, un format de représentation des connaissances :
<http://framenet.icsi.berkeley.edu/>
- CYC, un projet encyclopédique commencé en 1985 et toujours en développement :
<http://www.opencyc.org/>

Exemple : ontologie sur les animaux



Traduction logique possible

- chaque concept est un prédicat à une place
- chaque lien de subsomption est une relation logique (base de règles) :
 $\forall x[mammifere(x) \implies animal(x)] \dots$
- chaque feuille est une instance (base de faits) : $canari(Titi)$
- propriété héritable au sous-arbre oiseau : $voler(x)$

Propriétés

- pas vraiment de consensus sur les concepts de base devant figurer dans une ontologie
- peut-on vraiment construire une ontologie indépendamment d'une langue ? (cf. thèse de Sapir-Whorf)
- influences sociales, culturelles...

Intérêt

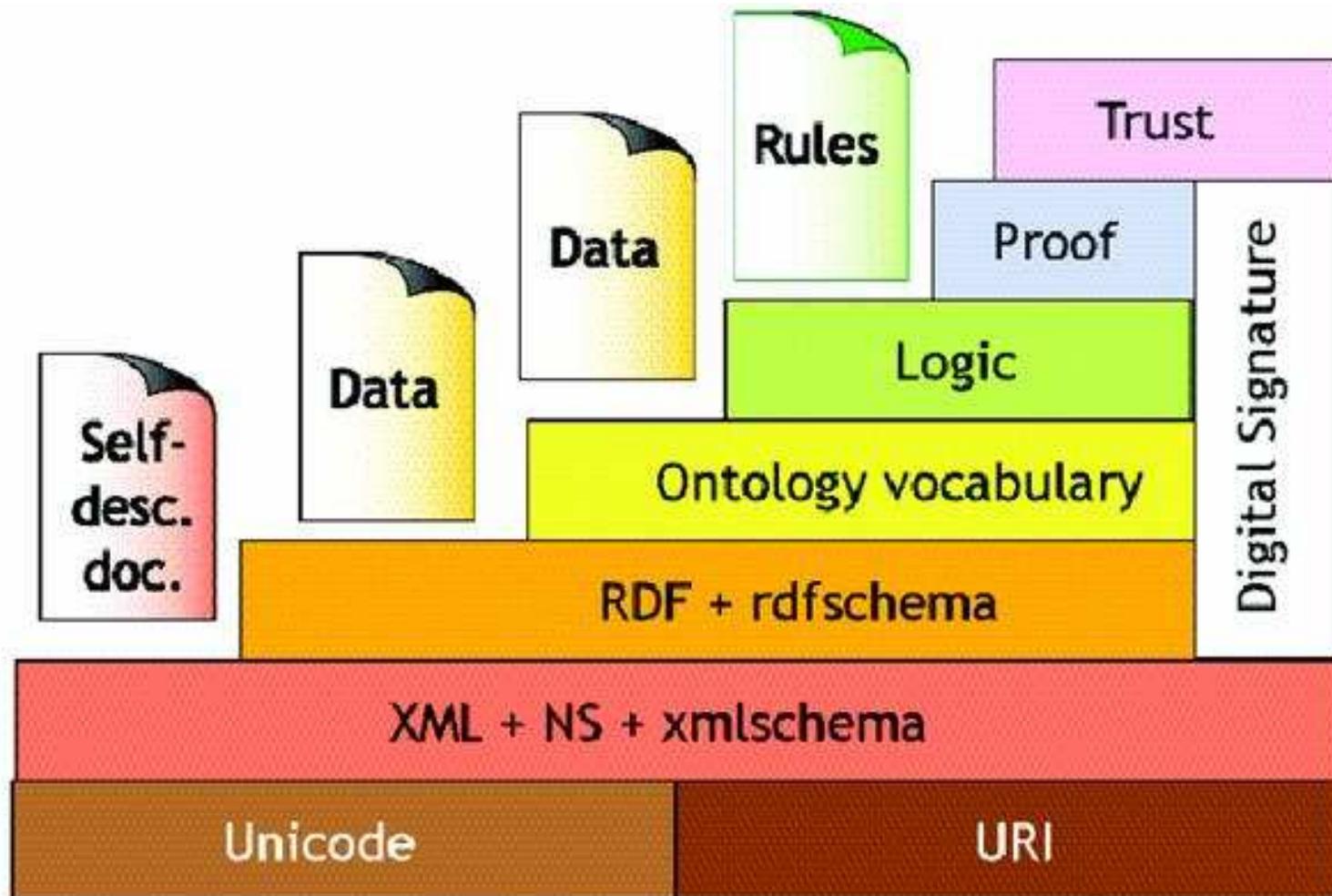
- modèle de formalisation des connaissances
- permet de faire des inférences (pas un thesaurus)

Généralités

- Web sémantique = Web de données
- des langages de représentation des connaissances
 - représentation de faits particuliers : RDF
 - représentation de règles générales : RDFS, OWL
 - une version simplifiée des deux : notation 3
- des outils pour les manipuler
 - des éditeurs de bases de connaissances : Protégé
 - des moteurs d'inférences : CWM, Jena...
 - un langage d'interrogation type SQL : SPAQL
 - un navigateur adapté : le Tabulator
- il ne reste plus qu'à s'en servir !

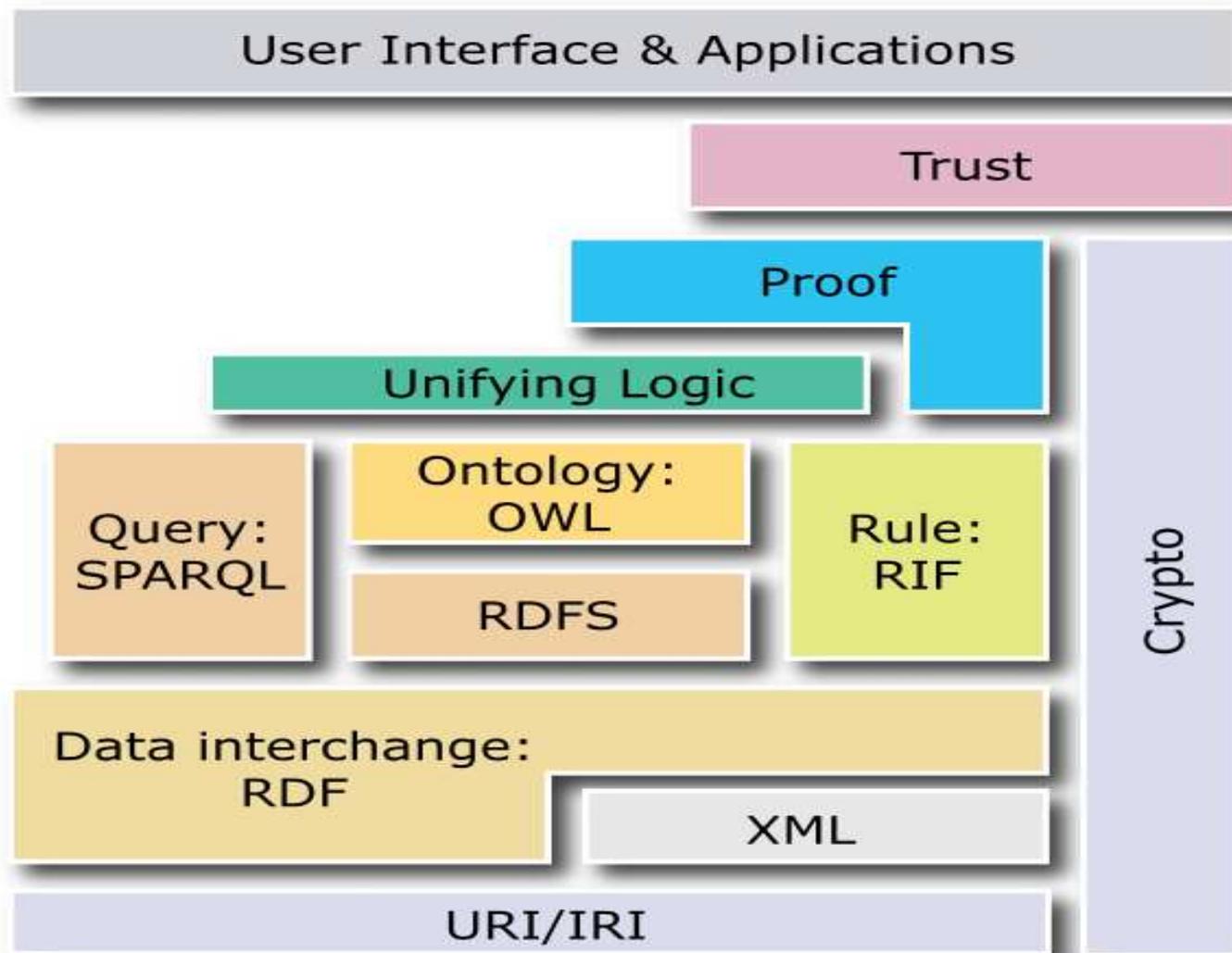
Les couches du Web sémantique

Représentation graphique



Les couches du Web sémantique

Représentation graphique mise à jour



Généralités

- principe de RDF : exprimer des faits à l'aide de triplets
- un triplet $\langle a \ P \ b \rangle$ s'interprète comme : $P(a, b)$
- domaine des éléments d'un triplet :
 - a et P sont des URI
 - b est une URI ou littéral (chaîne de caractères ou nombre)
- parmi les propriétés P possibles, on a intérêt à réutiliser certains standards :
 - prédicat is-a entre une instance et une classe
 - prédicats du DublinCore, de FOAF...
 - prédicats ensemblistes de rdfs : Class, SubClassOf, Property...
- on peut facilement faire appel à des entités ou des prédicats définis ailleurs (intégration de données)
- permet d'associer des métadonnées à des pages Web

Exemple (en notation 3)

définition de préfixes :

@prefix dc : <http://purl.org/dc/elements/1.1/> .

@prefix foaf : <http://xmlns.com/foaf/0.1/>

@prefix lifo : <http://www.univ-orleans.fr/lifo/Members/>

@prefix : <#> .

déclaration de faits :

lifo :Isabelle.Tellier dc :creator :moi

lifo :Isabelle.Tellier dc :title "I. Tellier : accueil"

:moi a foaf :Person

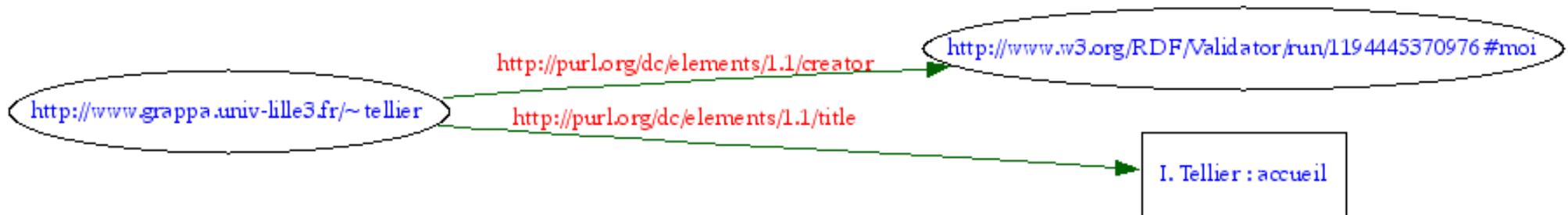
:moi foaf :name "Tellier"

Représentation graphique

déclaration de faits :

lifo :Isabelle.Tellier dc :creator :moi

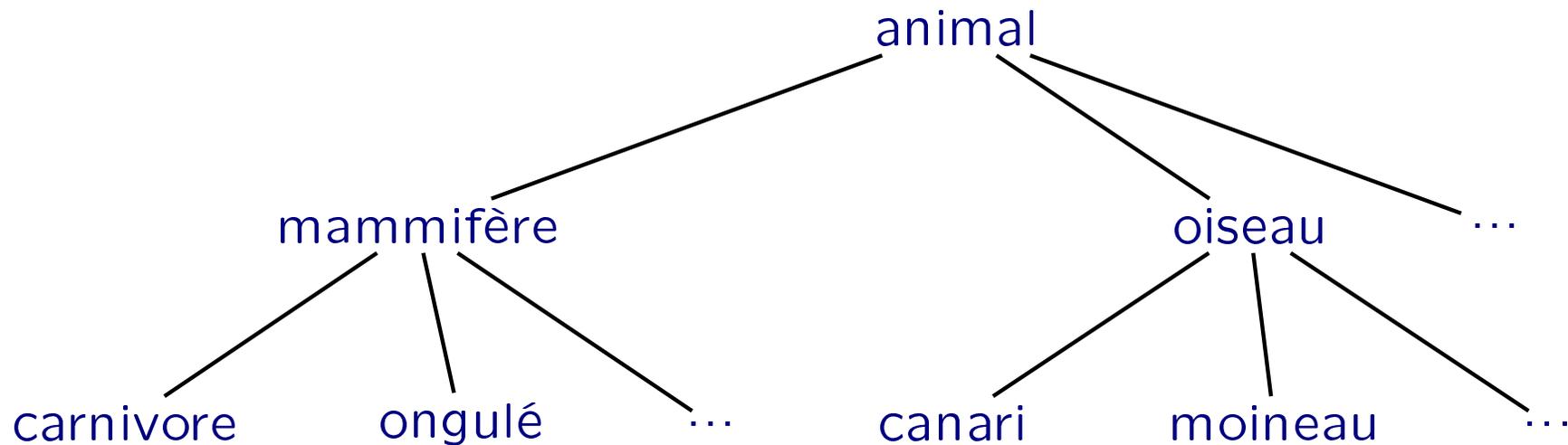
lifo :Isabelle.Tellier dc :title "I. Tellier : accueil"



Généralités

- OWL : surcouche de RDF/RDFS permettant la définition de règles, et donc l'encodage d'ontologies
- OWL inclut la définition de métaprédicats
ex : Parentde InverseOf Enfantde
- OWL permet de définir des contraintes d'intégrité
- différents niveaux d'expressivité :
 - OWL Lite (la moins expressive)
 - OWL DL pour Description Logic
 - OWL Full (version complète, mais tout n'y est pas décidable!)
- fichiers RDF et OWL peuvent être générées automatiquement à partir de triplets ou d'éditeurs (ex : Protégé)
- il existe des moteurs d'inférence (ex : CWM, Jena)
- exemples d'applications : FOAF, moteurs de recherche sémantiques...

Ontologie sur les animaux



Traduction approximative en OWL (RDFS en notation 3)

animal is-a Class

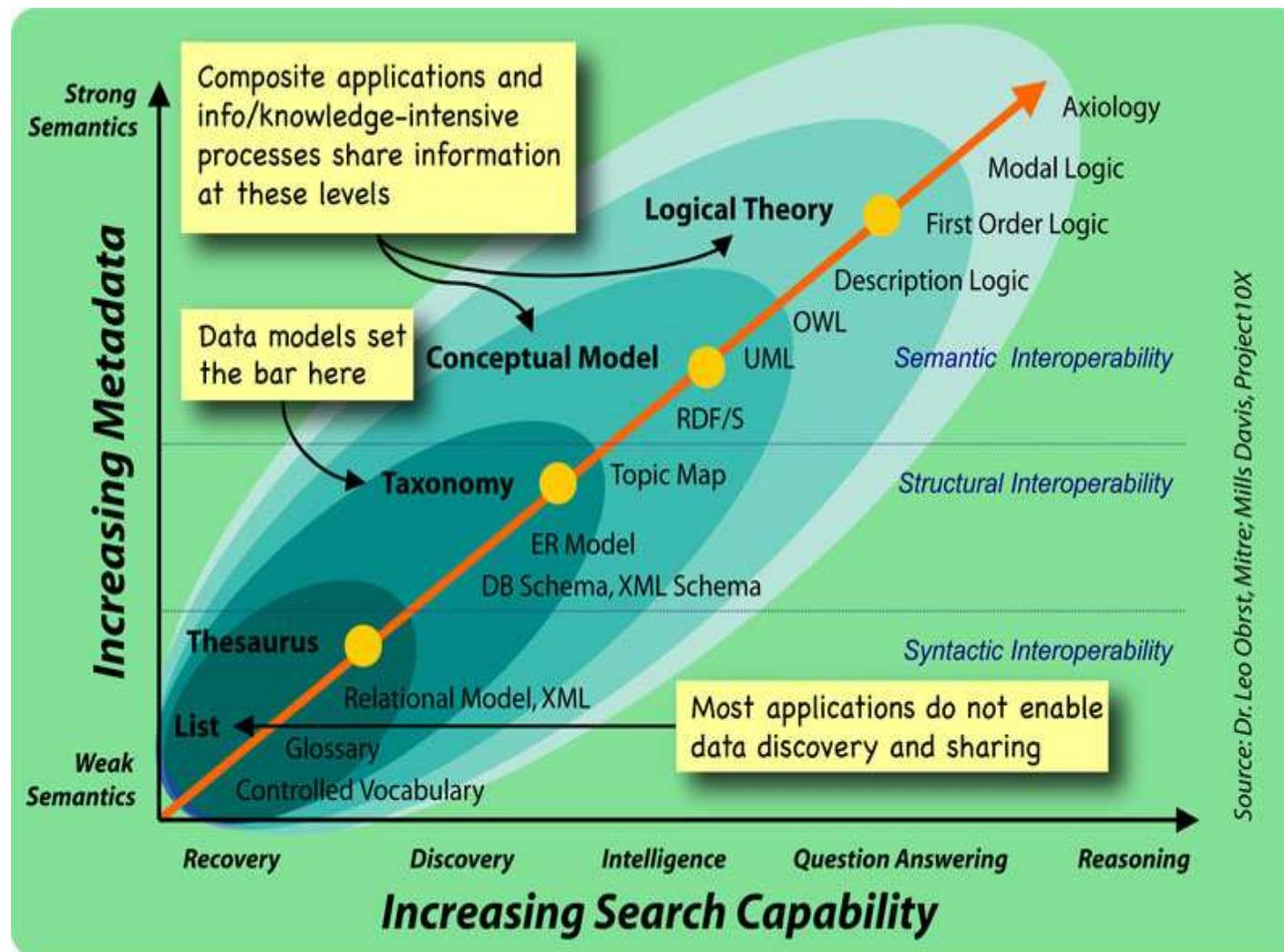
mammifere SubClassOf animal

Titi is-a canari

Le Web sémantique = Web 3.0 ?

- projet : mettre l'architecture du Web au service de l'intégration de données (et plus seulement de documents)
- RDF, OWL : standards de représentation des connaissances
- doivent être associés à un langage de requête : SPARQL
- doivent être associés à des systèmes d'inférences
- les ontologies joueront-elles le rôle de thesaurus du Web ?
- critiques sur l'annotation (RDF) : demande un effort supplémentaire ou des outils de transcription
- critiques sur les inférences (OWL) : problèmes d'efficacité, de cohérence

Le Web sémantique = Web 3.0 ?



Le Web sémantique = Web 3.0 ?

