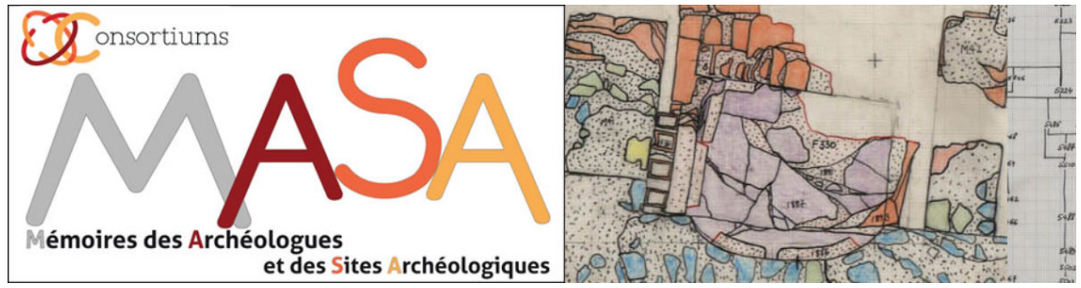


Journées DONIPAT

Thérèse LIBOUREL- Université de Montpellier

Modéliser une ontologie

Espace DEV

Objectif

Comment modéliser une ontologie ?

Ontologies

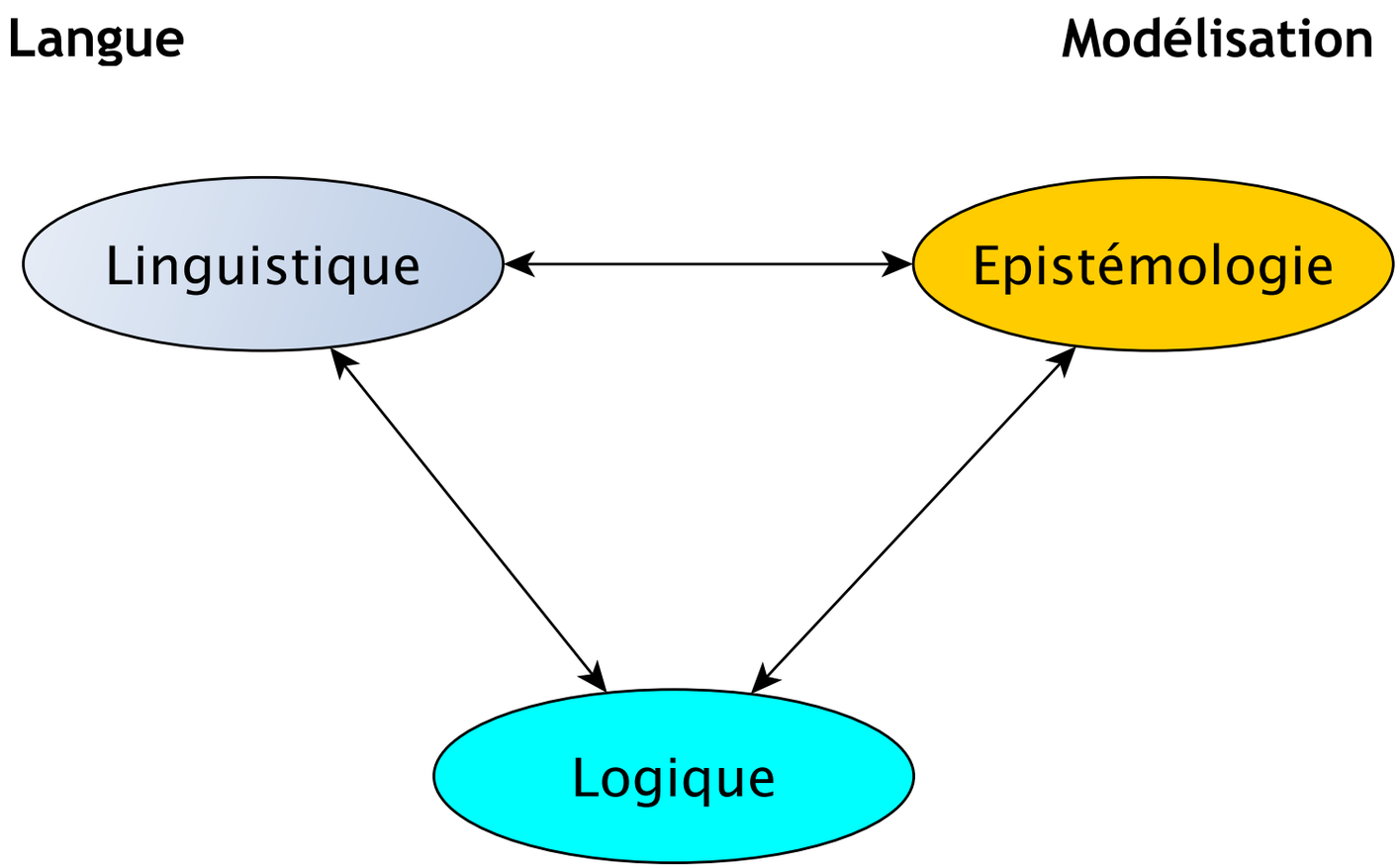
Le terme « ontologie », apparaît il y a une dizaine d'années, dans le domaine des sciences cognitives et celui de l'informatique. Pris dans son sens le plus large, le terme ontologie est plus ou moins synonyme de : théorie ou conception du réel.

Une ontologie est une compréhension partagée d'un domaine d'intérêt» [Uschold, 98]

Une ontologie peut prendre différentes formes, mais elle inclura nécessairement un *vocabulaire de termes* et une *spécification de leur signification*. Cette dernière inclut des définitions et une indication de la façon dont les concepts sont reliés entre eux, les liens imposant collectivement une structure sur le domaine et contraignant les interprétations possibles des termes."

Une telle caractérisation rend compte d'objets divers tels des glossaires, des terminologies, des thesaurus et des ontologies (au sens strict), mis en œuvre par différents professionnels et se distinguant suivant que l'accent est mis sur les termes ou leur signification.

Ingénierie des connaissances



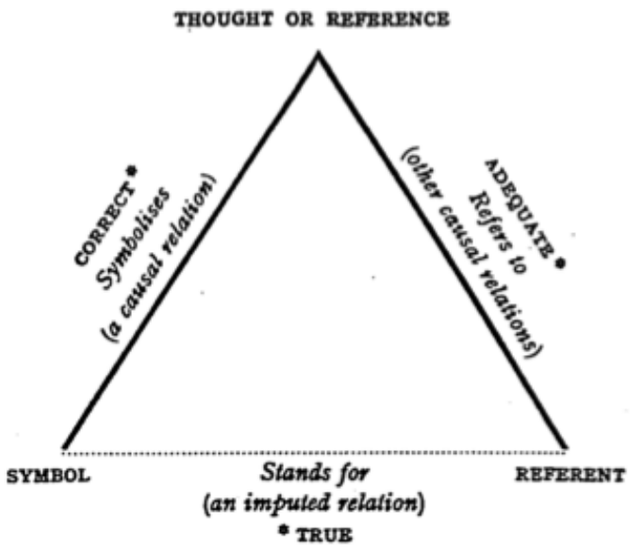
Représentation

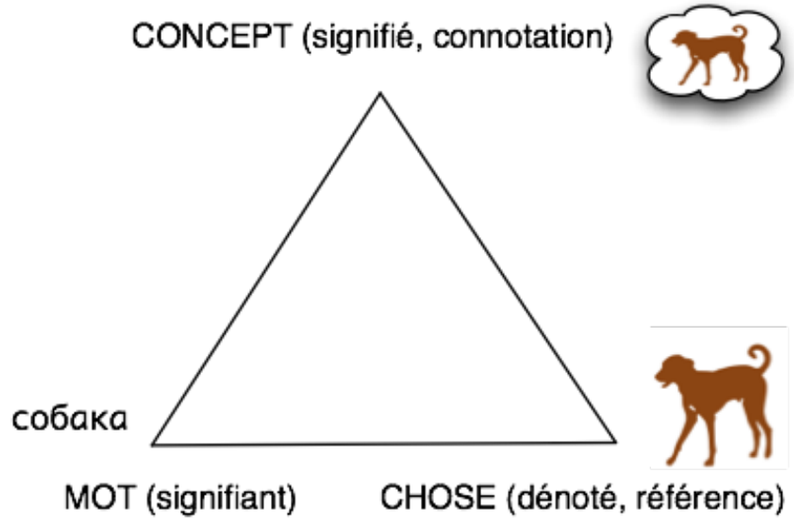
(Approche computationnelle)

Ontologies

Les recherches et écrits s'expriment dans la langue de spécialité sous la forme d'une terminologie et d'une conceptualisation associée spécifiques.

L'existence de tels « jargons » entraîne des problèmes de compréhension et des difficultés à partager des connaissances entre les acteurs.





Linguistique

Un **vocabulaire contrôlé** est une liste de termes convenus par une communauté ou une organisation.

Une **taxonomie** est un vocabulaire contrôlé organisé comme une hiérarchie. Par exemple, on peut avoir les termes « véhicule », « voiture » et « moto ». Les concepts moto et voiture sont des sous-classes de véhicule car une voiture et une moto sont des types de véhicules.

Un **thésaurus** est une taxonomie avec plus d'informations à propos de chaque concept incluant les termes alternatifs et préférés (« animals » en anglais ou « animaux » en français). De plus, un thésaurus peut contenir des liens vers des concepts connexes. Par exemple, les concepts « animaux » et « forêt » ont plusieurs types de liens. La plupart des sites de commerce utilisent des taxonomies et des thésaurus pour aider les utilisateurs de leurs recherches.

Thesaurus

Un lexique (ensemble de mots d'une langue)
descripteurs et non descripteurs

Un **sous-ensemble de définitions** et de notes d'application

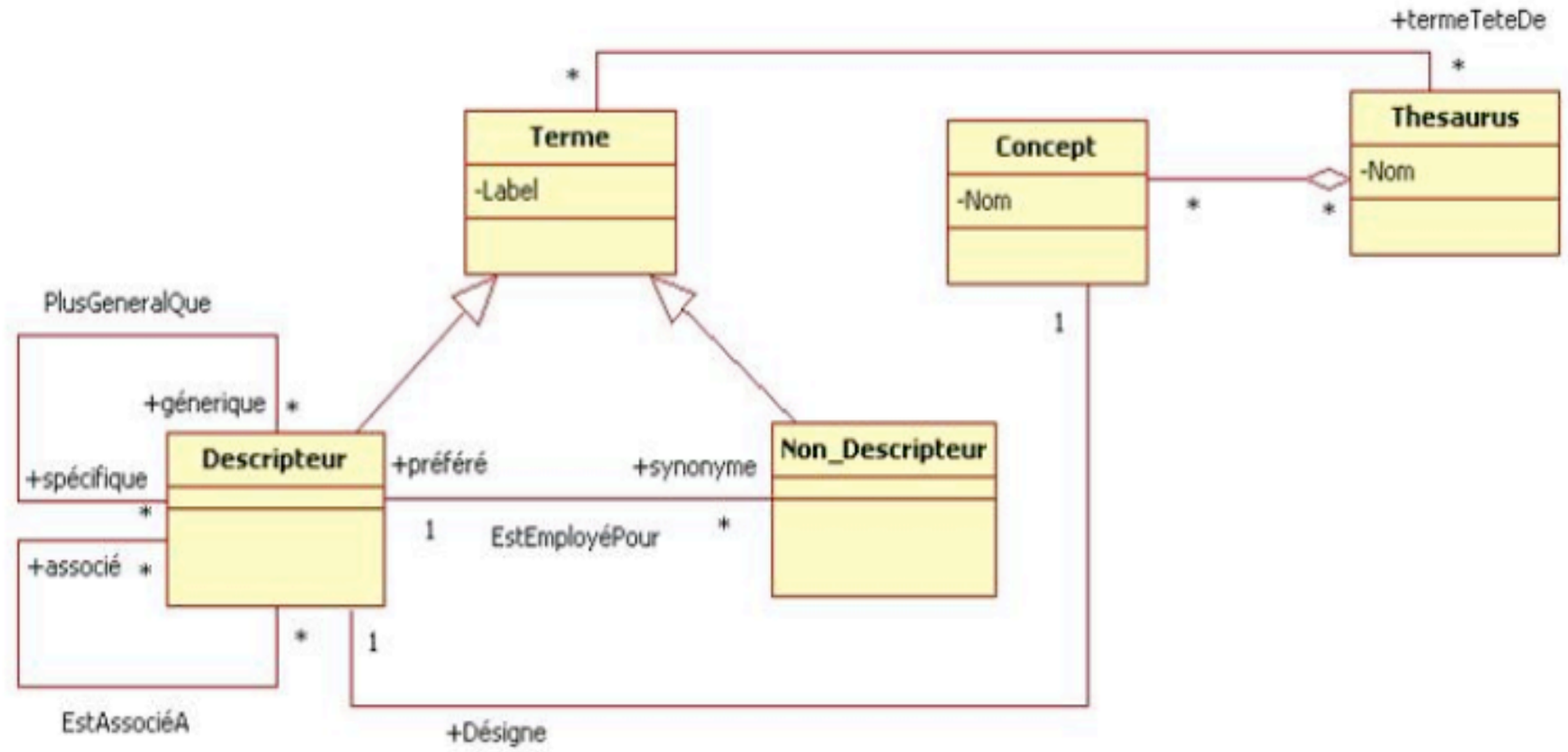
Une **structure classificatoire** exprimée par des relations sémantiques entre les termes du lexique, et parfois traduits par une notation symboliques

relation hiérarchique
relation d'association
relation d'équivalence intra linguistique
relation d'équivalence inter linguistique

•

Modèle SKOS

Spécification en langage RDF développé par le W3C, pour la publication et l'utilisation des thésaurus dans le cadre du Web sémantique.



Ontologie (définitions)

[Gruber, 93]

« An ontology is an explicit specification of a conceptualisation»

« Une ontologie est une spécification explicite d'une conceptualisation»

In the context of knowledge sharing, I use the term ontology to mean a specification of a conceptualization. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. What is important is what an ontology is for.

[Uschold, Gruninger, 96]

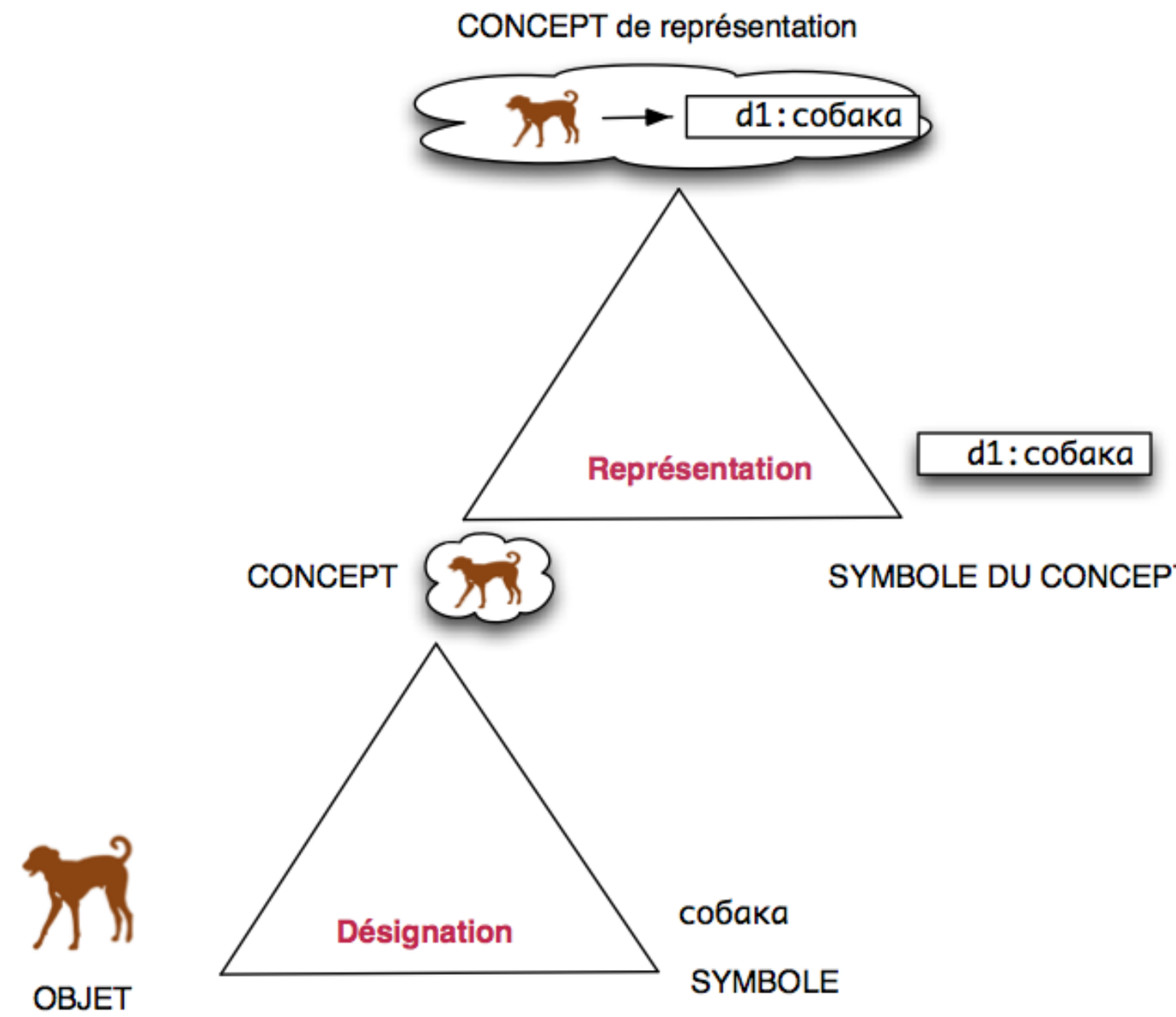
« An ontology is a shared understanding of some domain of interest»

«Une ontologie est une compréhension partagée d'un domaine d'intérêt»

[Guarino]

« en IA, une ontologie représente un artefact d'ingénierie, constitué par un vocabulaire spécifique utilisé pour décrire une certaine réalité, accompagné d'un ensemble d'hypothèses implicites concernant la signification des mots de ce vocabulaire »

Ontologies


CONCEPT de représentation
d1 : собака
Représentation
d1 : собака
CONCEPT
SYMBOLE DU CONCEPT
Désignation
собака
SYMBOLE
OBJET

Démarche

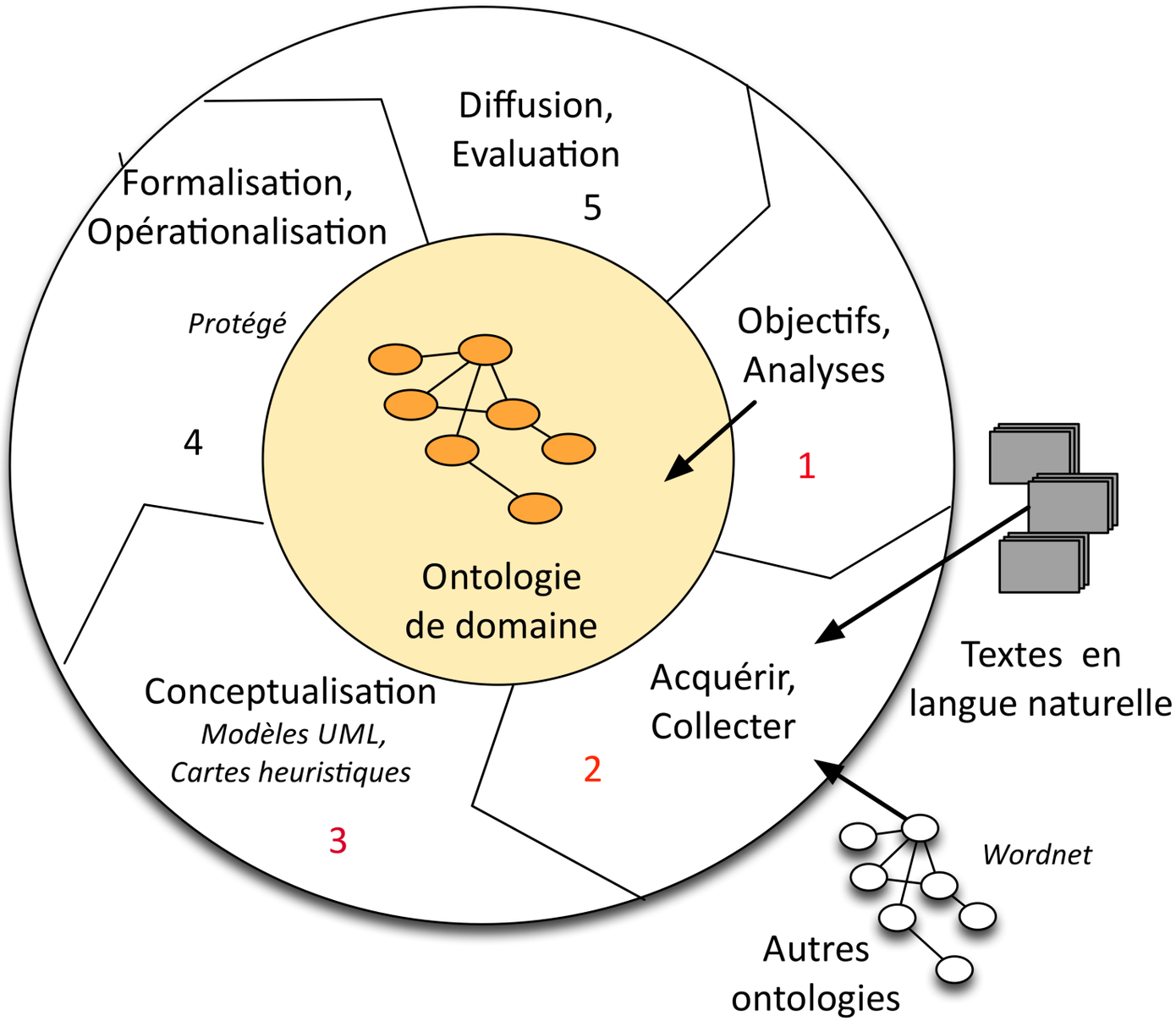
Les premières ontologies ont été développées de façon complètement artisanale, sans suivre de méthode prédéfinie.

Actuellement, variété de méthodes analogues à celles

- Génie Logiciel
  - METHONTOLOGY Fernandez, Gomez Perez - Madrid http://ir.ii.uam.es/s5t/
  - TOVE (Gruninger et Fox, Toronto)

- Représentation et gestion des connaissances
  - Common Kads
  - Approche FIPA (Foundation for Intelligent Physical Agent)
  - TERMINAE (Biébow & Szulman, 99) ..

Démarche


Diffusion,
Evaluation
5
Formalisation,
Opérationalisation
Protégé
4
Objectifs,
Analyses
1
Ontologie
de domaine
Conceptualisation
Modèles UML,
Cartes heuristiques
3
Acquérir,
Collecter
2
Textes en
langue naturelle
Wordnet
Autres
ontologies

Démarche

Phase préalable : appréhender l’étendue et le but

Phase de conceptualisation

Phase d’ontologisation

Phase opérationalisation

Etape 1

Phase préalable : appréhender l'étendue et le but

- Quel domaine va couvrir l'ontologie ?
- Dans quel but utiliserons nous l'ontologie ?
- A quels types de questions l'ontologie devra t-elle fournir des réponses ?
- Qui va utiliser et maintenir l'ontologie

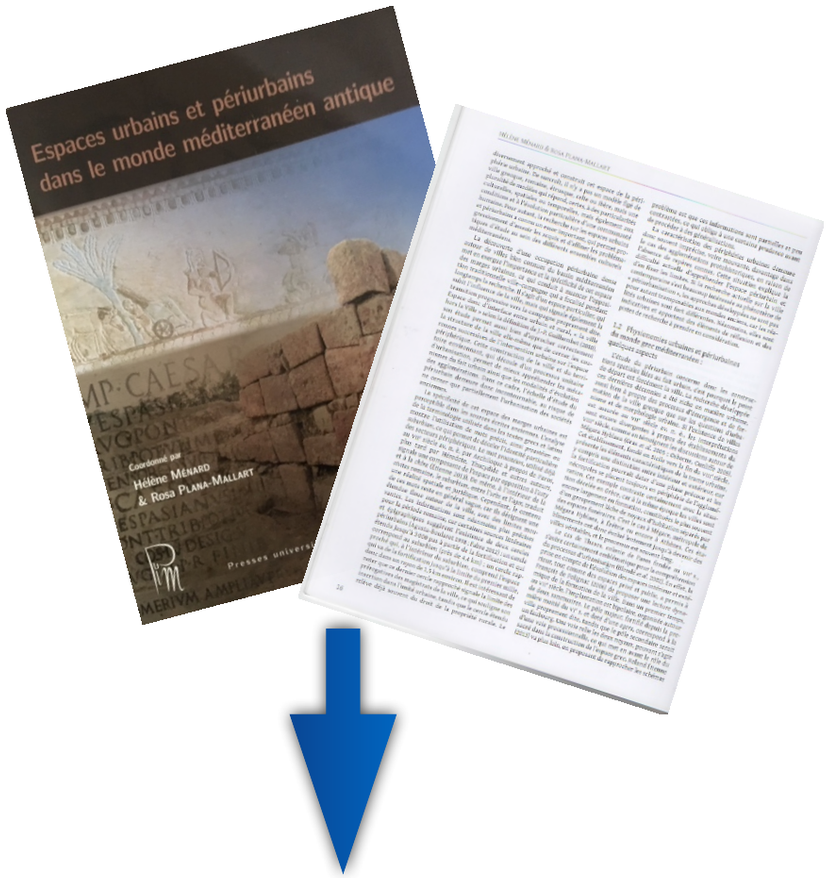
Etape 2

Collecte et analyse des textes divers

Confrontation des connaissances ainsi explicitées avec celles déjà présentes et analysées au sein de diverses ontologies

Etape 3

Conceptualisation



Analyse de texte

Extraction
substantifs, verbes, qualificatifs

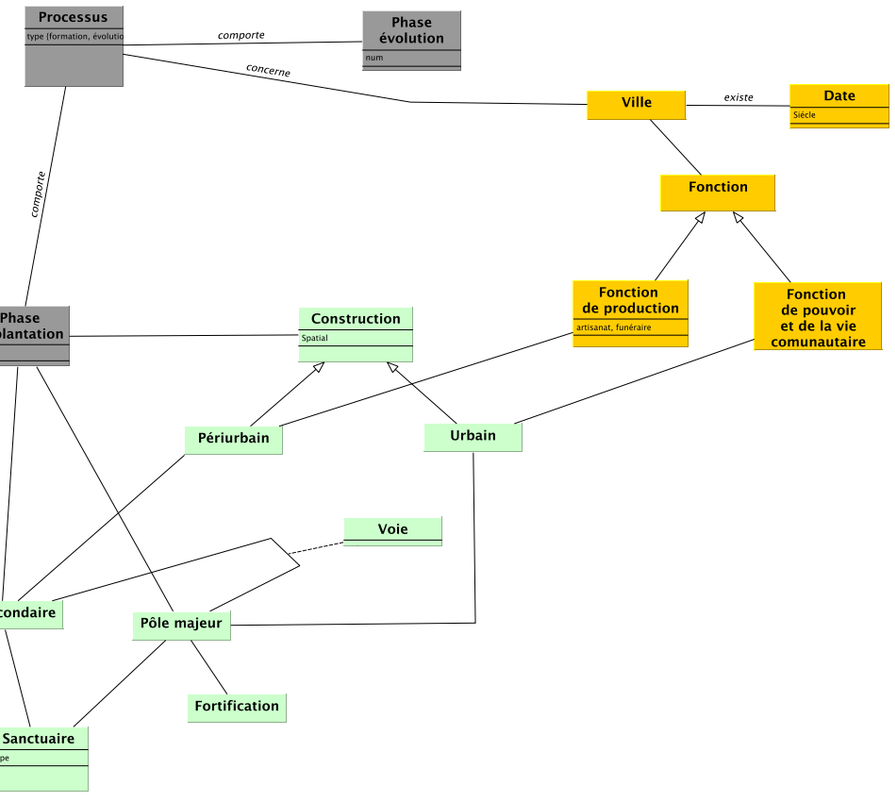
	Analyse	
Nom	**Verbe**	**qualificatif**
périurbain	concerne	
constructions spatiales	etre liée	
urbain		
ville		
processus		émergence/ formation/ urbanisation
existent	assurée	
VIIémé siècle		
établissement	fondé	
sicile	présente	
éléments	caractérise	

Etape 3

Conceptualisation

- Modélisation UML

	Analyse	
Nom	**Verbe**	**qualificatif**
périurbain	concerne	
constructions spatiales	etre liée	
urbain		
ville		
processus		émergence/ formation/ urbanisation
existent	assurée	
VIIémé siècle		
établissement	fondé	
sicile	présente	
éléments	caractérise	



Etape 4

Phase de formalisation

- Outil Protégé
- S’assurer des possibilités de réutilisation d’ontologies existantes

Etape 4

La définition de Gruber :
« Une ontologie est la spécification d'une conceptualisation. [...] Une conceptualisation est une vue abstraite et simplifiée du monde que l'on veut représenter »

s'appuie sur deux dimensions :

une ontologie est la *conceptualisation* d'un domaine, c'est-à-dire un *choix* quant à la manière de décrire un domaine ;

c'est par ailleurs la *spécification* de cette conceptualisation, c'est-à-dire sa *description formelle*.

Formalisation (Ontologie abstraite)

Une ontologie abstraite est une structure
$O=(C, \leq c, R, F, \leq r, IR)$

C et R sont deux ensembles disjoints
appelés Concepts et Relations

$\leq c$ un ordre partiel sur C (hiérarchie de concepts)

$\leq r$ un ordre partiel sur R (hiérarchie de relations)
(nécessité de signature fonction de C sur C)
IR un ensemble de régles d'inférences exprimées dans un langage logique

Les formalismes

Réseaux sémantiques (Quillian 1968) graphes étiquetés

Frames Language (Minsky, 1975)

Graphes conceptuels (Sowa, 1991)

Logiques de description (Brachman, Levesque 1984)

Modélisation objet

Web sémantique : XML, RDF, RDFS, OWL

Ontologie formelle

Logiques de description

But représenter des connaissances, raisonner à partir des connaissances

Le niveau terminologique ou TBox

décrit les connaissances générales d'un domaine sous forme intentionnelle
définit les concepts (classes) et les rôles (associations) et leur taxonomie
Une sémantique est associée à chaque description de concept et de rôle par l'intermédiaire d'une *interprétation*

Le niveau factuel ou ABox

décrit les individus en le nommant et en spécifiant leur classes et attributs (concepts et rôles)
spécifie des assertions portant sur des individus nommés

Les r**aisonnements** interviennent aux deux niveaux et s'appuient sur deux opérations de base :

La classification (place des concepts ds la hiérarchie)
L'instanciation (concept dont l'individu est susceptible d'être instance)

Ontologie formelle

Logiques de description

La base de connaissance = TBox ∪ Abox

La TBox contient
- concepts et rôles atomiques (4 primitives Top, Bottom)
- concepts et rôles composés par l'intermédiaire de constructeurs (rque : l'expressivité de la LD dépend des constructeurs)
- deux axiomes
- subsomption C2 ≤C 1
  - hiérarchie de concepts
- équivalence C2 ≡ C1

Personne ≤ Top
Homme ≤ Personne
Femme ≤ (Personne and (not Homme))

Ontologie formelle

Web sémantique

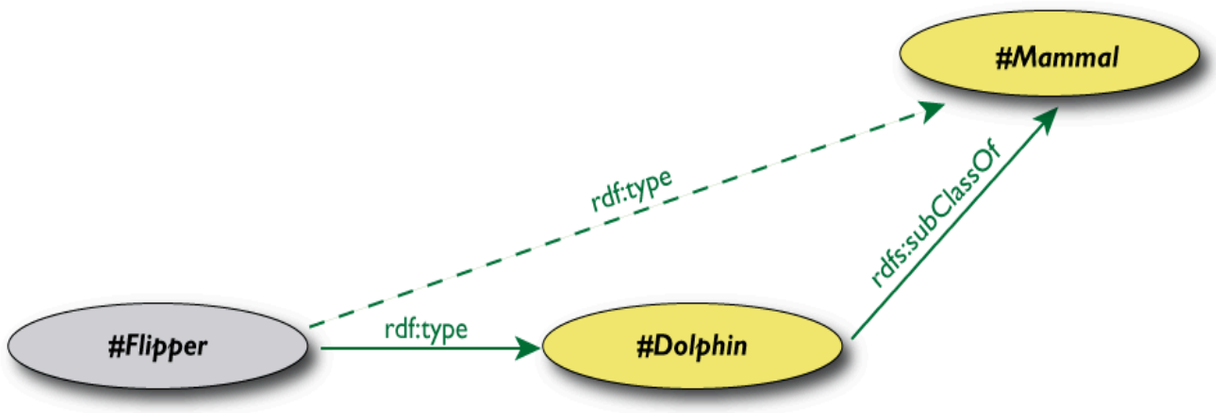
Standards du W3C

URI Uniform Ressource Identifier

RDF Resource Description Framework
Langage standard destiné à l'annotation sémantique pour la description de ressources disponibles sur le Web

Notion de triplet (Prédicat, Sujet, Objet)

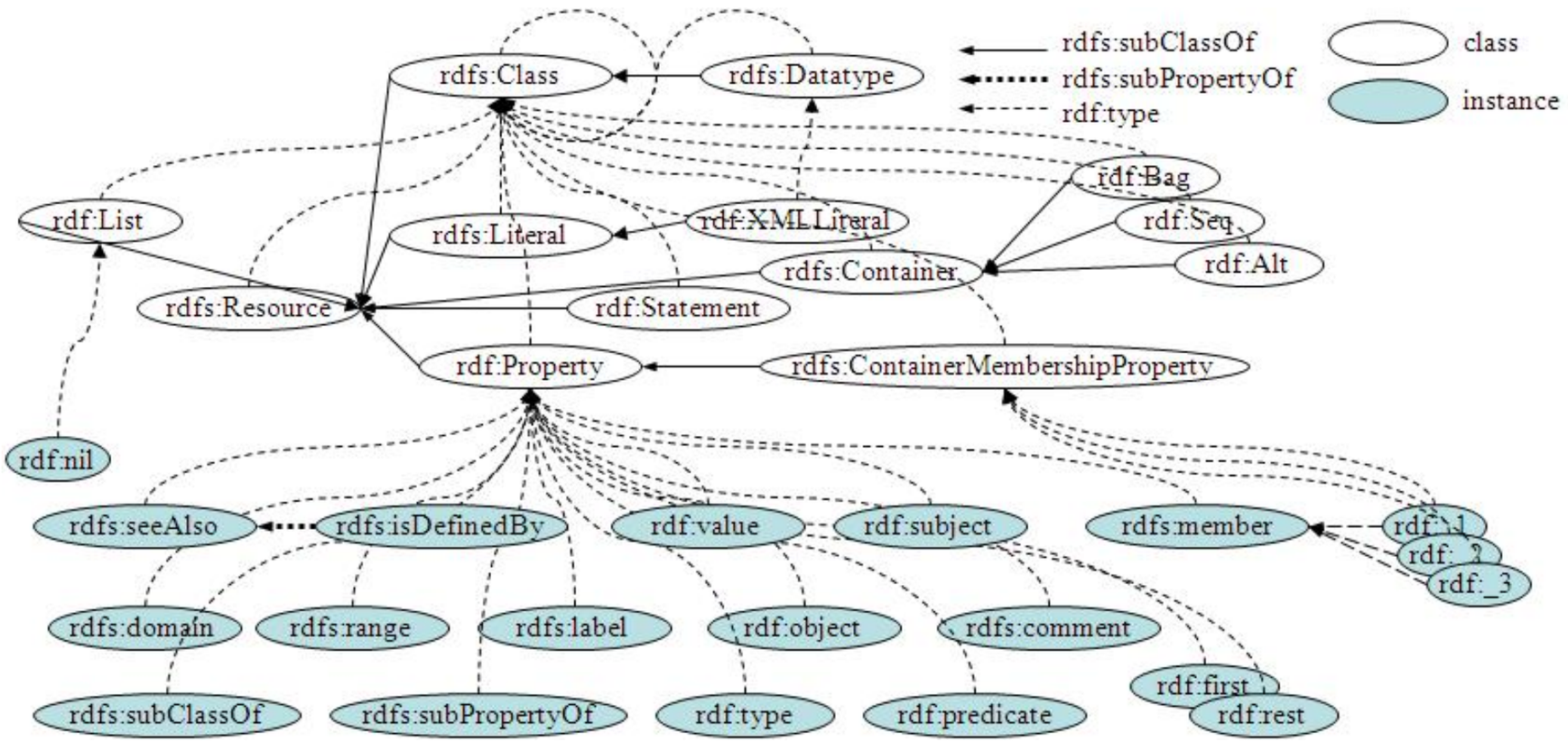
Exemple



Ontologie formelle

Web sémantique

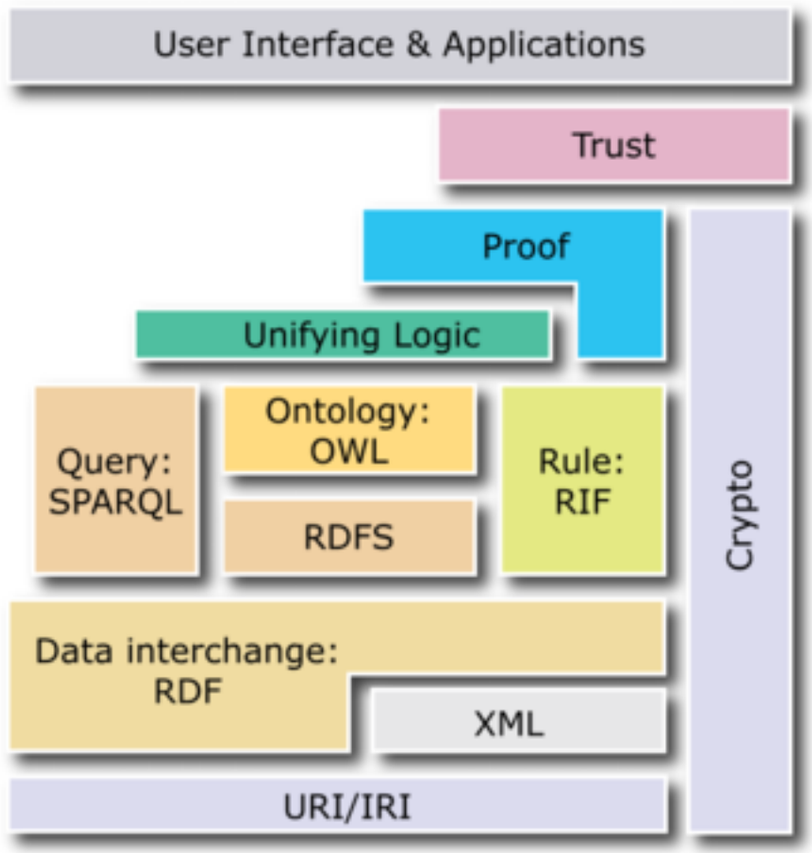
RDFS : RDF Schema
Enrichi RDF avec notion de classes et de hiérarchies



Ontologie formelle

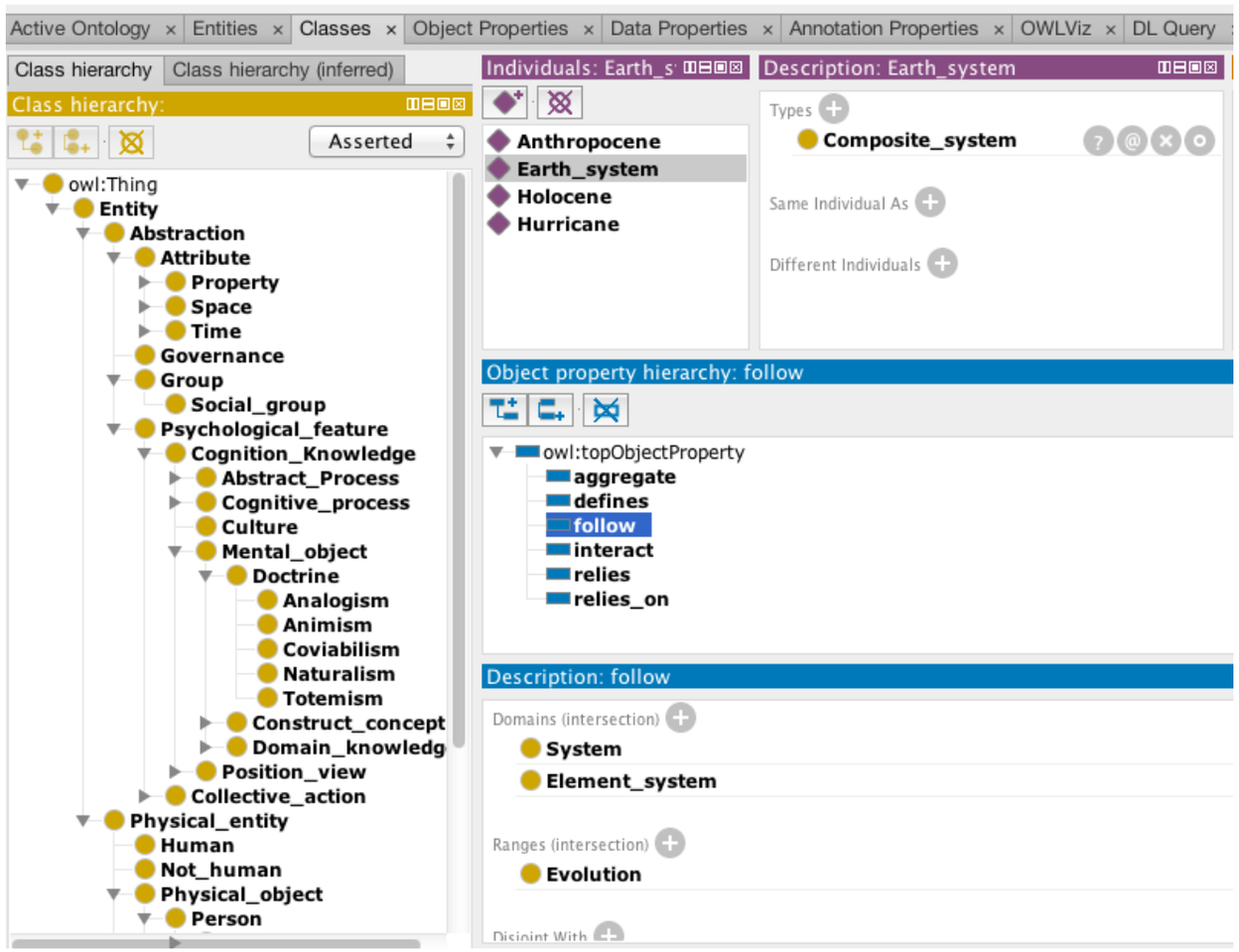
Web sémantique

OWL The Web Ontology Language



Etape 4

Ontologisation : Obtention de modèles sous le logiciel PROTEGE



Ce logiciel a été conçu pour le Département d'informatique Médicale de l'Université de Stanford, notamment pour construire des ontologies.

Plus généralement il permet de concevoir des SBCs par réutilisation de modèles du domaine et de méthodes de résolution de problèmes.

Une nouvelle version de l'outil - Protégé2000 - développée en JAVA et autorisant la construction coopérative d'ontologies, est accessible à l'adresse : http://Protege.Stanford.EDU.

Conclusion

Quelques réflexions

- *Respecter les standards*
- *Modularité*
- *Ne pas réinventer la roue*
  - Vérifier l'existant
  - Fusion
  - Extraction