

La pratique de la modélisation dans le cadre du web sémantique

*Aurélien Arena
23 mai 2013*

- 1. Présentation générale du web sémantique**
- 2. Aspects théoriques**
- 3. Aspects applicatifs**

1. Présentation générale du web sémantique

1. Web sémantique: les principes

- **Initiateur:** Tim-Berners Lee, 2001 (inventeur du web, HTML...)
- **Objectif:** Rendre le web plus « intelligent » de sorte qu'il soit capable de juger de la pertinence d'une réponse en analysant la sémantique de la question correspondante.
- **Contrainte:** Pas par refonte mais par extension de l'existant

1. Web sémantique: contexte socio-scientifique

- **Evolue sur plusieurs communautés**
 - Web
 - Représentation des connaissances
 - Logique et algorithmique
 - Base de données
 - Communautés respectives des objets traités (traitement du langage, image...)

1. Web sémantique: les notions de sémantique

- **Sémantique en linguistique:**

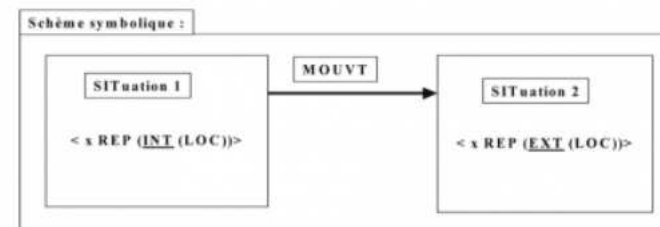
- Décomposition en niveaux

Niveaux	Unités
Sémantique	Sème
Syntaxique	Syntagme
Lexical	Lexème
Morphologique	Morphème
Phonologique	Phonème

- Plusieurs approches de la sémantique

Sèmes →	dossier	bras	individuel	collectif
canapé	x	x		x
chaise	x		x	
tabouret			x	

sortir(x,y)



1. Web sémantique: les notions de sémantique

Mais aussi:

- **Sémantique en mathématique (règles d'interprétation)**

- Conditions de vérité des énoncés

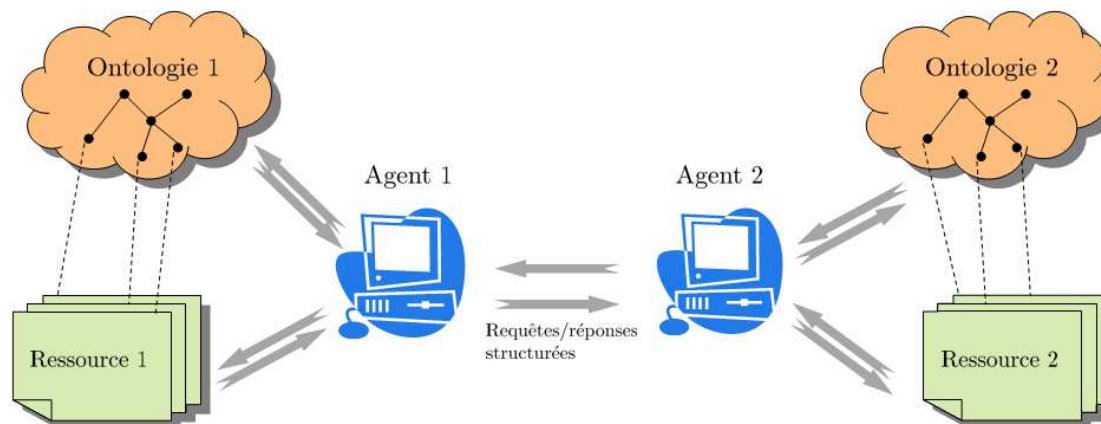
$V(\phi)$	$V(\psi)$	$V((\phi \wedge \psi))$	$V((\phi \vee \psi))$
1	1	1	1
1	0	0	1
0	1	0	1
0	0	0	0

- **Sémantique en compilation**

- Analyse lexicale, syntaxique et sémantique

1. Web sémantique: les principes

- **Une ressource:** tout ce qui est référencable par une URL (qu'elle ait une dénotation réelle ou pas)
 - « http://fr.wikipedia.org/wiki/Web_semantique »
 - « <http://www.unsite.com/ddd#45211> »
- **Méta-données:**
 - Les ressources sont classifiées à l'aide de méta-données
 - Les méta-données sont structurées (ontologie)



1. Web sémantique / enrichissement

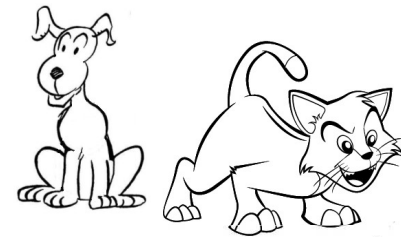
- **Classification:** opération de regroupement de ressources partageant les mêmes propriétés
- **Classe/catégorie (méta-donnée):**

$$estChien : D \rightarrow \{\top; \perp\}$$

- *On affirme qu'un objet (une ressource) possède les propriétés impliquées par la classe à laquelle il appartient.*

$$estChien(Rantanplan) = \top$$

→ ENRICHISSEMENT SEMANTIQUE

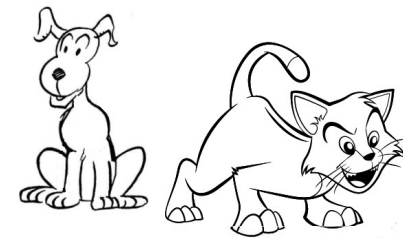
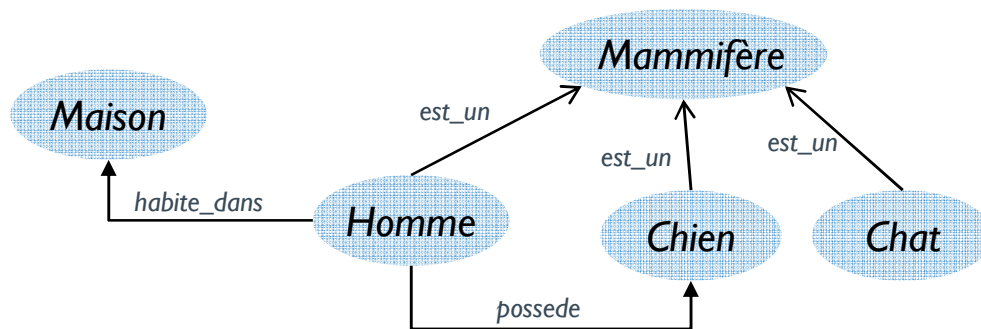


- Exemple d'enrichissements textuels: DBPedia

<div> <div></div> <div>About: Bauhaus</div> </div> <div> An Entity of Type : place, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org </div>	
<div> Das Staatliche Bauhaus wurde 1919 von Walter Gropius in Weimar als Kunstschule gegründet. Nach Art und Konzeption war es damals etwas völlig Neues. Das historische Bauhaus stellt heute die einflussreichste Bildungsstätte im Bereich der Architektur, der Kunst und des Designs dar. Das Bauhaus bestand von 1919 bis 1933 und gilt heute weltweit als Heimstätte der Avantgarde der Klassischen Moderne auf allen Gebieten der freien und angewandten Kunst. </div>	
<div>Property</div> <div>dbpedia-owl:abstract</div>	<div>Value</div> <div> <ul style="list-style-type: none"> Das Staatliche Bauhaus wurde 1919 von Walter Gropius in Weimar als Kunstschule gegründet. Nach Art und Konzeption war es damals etwas völlig Neues. Das historische Bauhaus stellt heute die einflussreichste Bildungsstätte im Bereich der Architektur, der Kunst und des Designs dar. Das Bauhaus bestand von 1919 bis 1933 und gilt heute weltweit als Heimstätte der Avantgarde der Klassischen Moderne auf allen Gebieten der freien und angewandten Kunst. Die Resonanz des Bauhauses hält bis heute an, und prägt wesentlich das Bild deutscher Entwürfe im Ausland. Staatliches Bauhaus, commonly known simply as Bauhaus, was a school in Germany that combined crafts and the fine arts, and was famous for the approach to design that it publicized and taught. It operated from 1919 to 1933. The term Bauhaus is German for "House of Building" or "Building School". The Bauhaus school was founded by Walter Gropius in Weimar. In spite of its name, and the fact that its founder was an architect, the Bauhaus did not have an architecture department during the first years of its existence. Nonetheless it was founded with the </div>
<div>dc:terms:subject</div>	<div> <ul style="list-style-type: none"> category:World_Heritage_Sites_in_Germany category:Modernist_architecture category:1926_architecture category:German_loanwords category:Industrial_design category:Bauhaus category:Art_schools_in_Germany </div>
<div>rdf:type</div>	<div> <ul style="list-style-type: none"> owl:Thing dbpedia-owl:Place yago:ArtMovements dbpedia-owl:WorldHeritageSite yago:ArchitecturalStyles yago:ArchitectureSchools </div>
<div>rdfs:label</div>	<div> <ul style="list-style-type: none"> Bauhaus Bauhaus Escuela de la Bauhaus Bauhaus Bauhaus パウハウス </div>
<div>is dbpedia-owl:wikiPageRedirects of</div>	<div> <ul style="list-style-type: none"> dbpedia:The_Bauhaus dbpedia:Bahaus dbpedia:Bauhaus_art_school dbpedia:Bauhaus_manifesto dbpedia:Bauhaus_Dessau dbpedia:Bauhaus_movement </div>
<div>is dbpedia-owl:movement of</div>	<div> <ul style="list-style-type: none"> dbpedia:Paul_Klee dbpedia:Oskar_Schlemmer dbpedia:Walter_Honegger-Lavater </div>

1. Web sémantique: ontologie

- **Ontologie:** un ensemble structuré de catégories



- Introduit un vocabulaire de description
- Exprime formellement des connaissances

$$\forall x, [estHomme(x) \rightarrow estMammifere(x)]$$
$$\{x \mid estMammifere(x) \wedge \exists y, [possede(x, y) \wedge estChien(y)]\}$$



1. Web sémantique: ontologie

- Autre exemple : Google knowledge graph

Marie Curie



Marie Skłodowska-Curie was a French-Polish physicist and chemist famous for her pioneering research on radioactivity. She was the first person honored with two Nobel Prizes—in physics and chemistry. Wikipedia

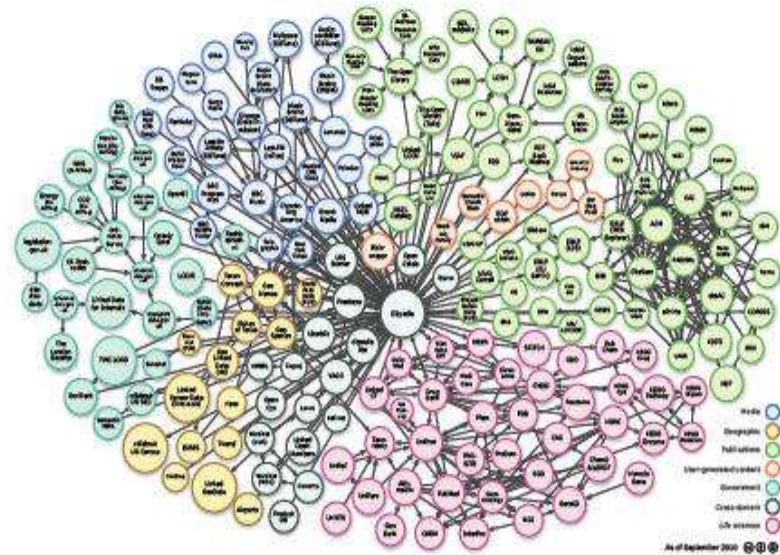
Born: November 7, 1867, **Warsaw**
Died: July 4, 1934, **Sancellemoz**
Spouse: **Pierre Curie** (m. 1895–1906)
Children: **Irène Joliot-Curie**, **Eve Curie**
Discovered: **Radium**, **Polonium**
Education: **École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris**, **University of Paris**

People also search for



Albert Einstein Pierre Curie Ernest Rutherford Louis Pasteur John Dalton

Ontologie (graphe RDF)



1. Web sémantique: deux axes

- *Web Sémantique : deux axes de recherche*

- ***Ontologies « légères »***

- *Communauté web « linked-data »*
- *Effort de standardisation*
DublinCore : titre, auteur, format...
- *Simplicité des descriptions*
- *Données distribuées*
- *W3C: HTML + RDFa*

- ***Ontologies « riches »***

- *Communauté de la représentation des connaissances (+ BD)*
- *Capacités d'inférences / Expressivité des langages de description*
- *Données locales*



2. *Aspects théoriques*

2.1. Quelques notions générales pour la modélisation

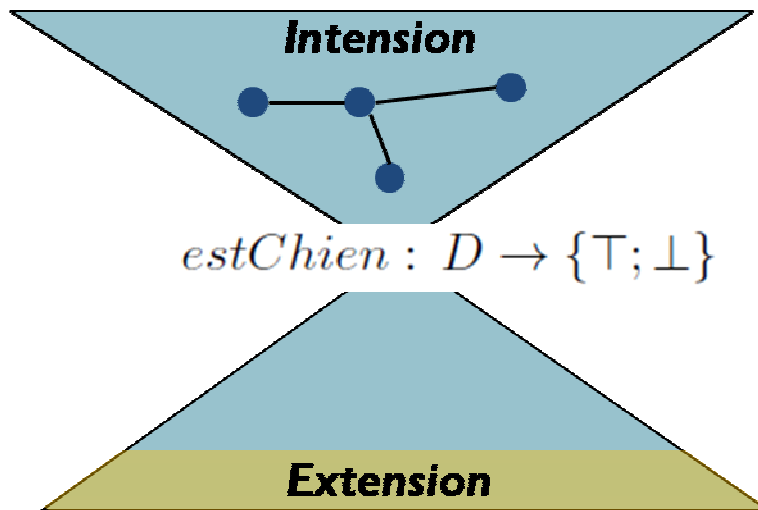
2.2. Cadre formel du web sémantique: les logiques de description

2.1 Quelques notions générales pour la modélisation

- La notion de **concept**

- Abstraction de propriétés partagées par un ensemble d'entités individuelles
- Fonction propositionnelle¹: $estChien : D \rightarrow \{\top; \perp\}$

- Les notions d'**extension** / **intension**



$$int(estChien) = \{p \mid estChien \rightarrow p\}$$

avec “ $f \rightarrow p$ ” se lisant “le concept p est une propriété du concept f ”.

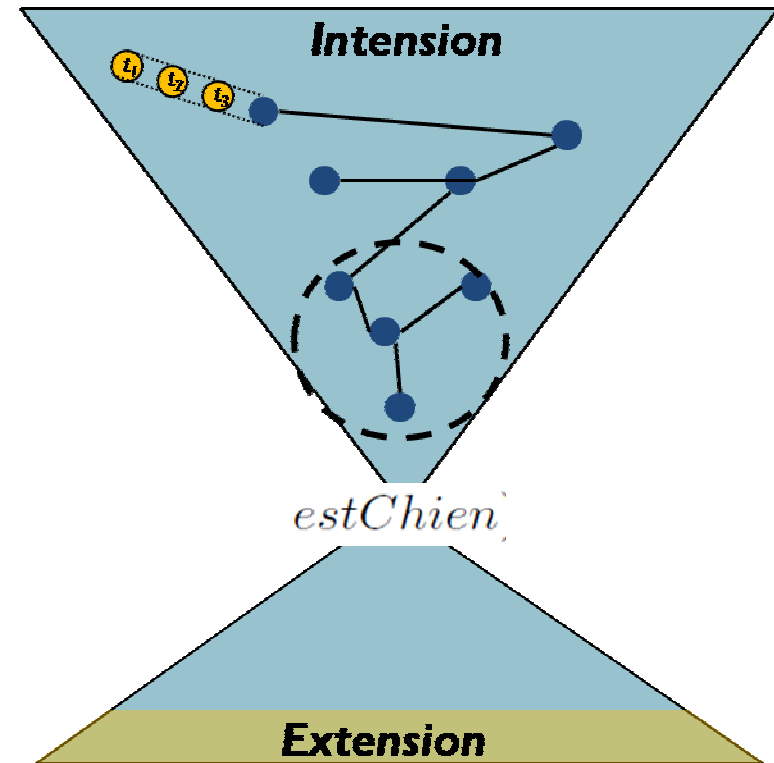
$$ext(estChien) = \{x \mid estChien(x) = \top\}$$

¹: ou prédicat de type

2.1 Quelques notions générales pour la modélisation

- **Différents types de propriétés**

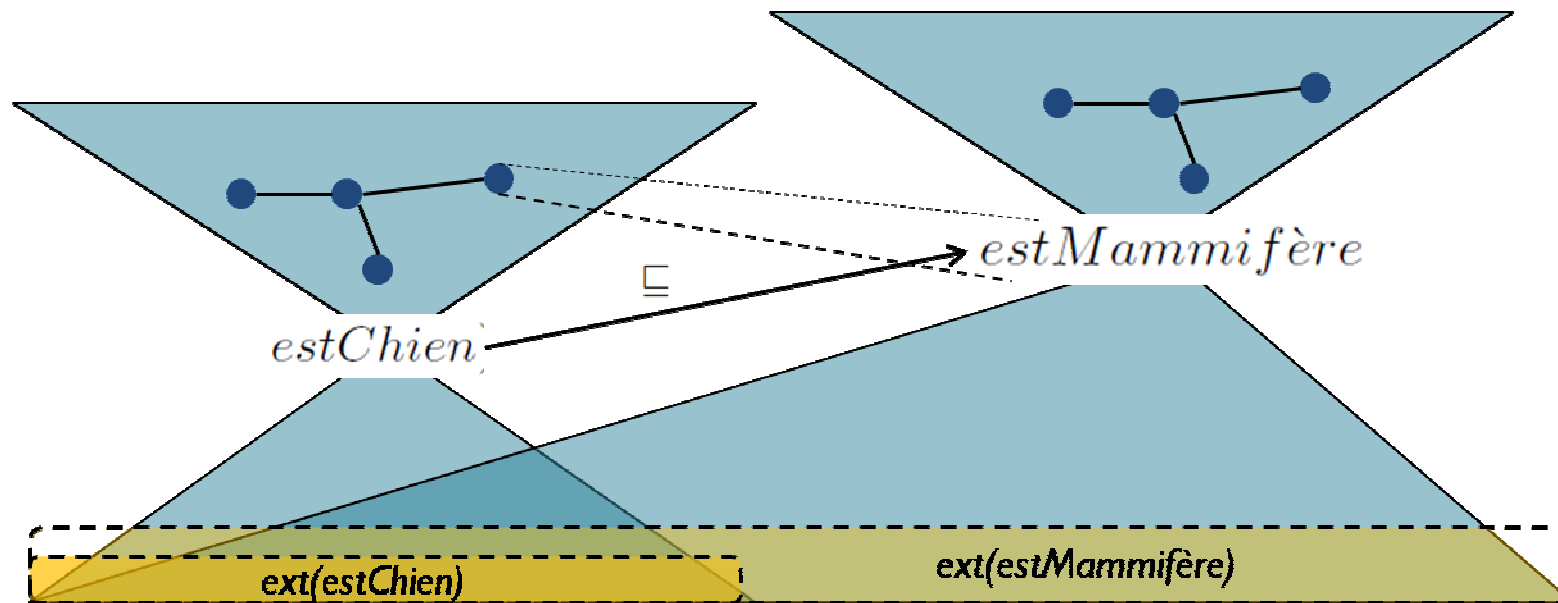
- Catégorisant:
 - ex: estMammifère()
- Non catégorisant:
 - ex: estPoilu()
- Relationnelle:
 - Ex: mangeDesOs()
- Distributif:
 - Ex: estBleu() \neq estSolidaire()
- Propriétés essentielles/contingentes
 - Ex: êtreMammifère(), êtreDebout()
 - Cf. E. Rosch sur la typicalité
 - Indexation temporelle, $P_t(x)$



2.1 Quelques notions générales pour la modélisation

- Relations de subsumption (hiérarchie) entre concepts

$$\begin{aligned} estChien \sqsubseteq estMammifère &\stackrel{Def}{\equiv} ext(estChien) \subseteq ext(estMammifère) \\ estChien \sqsubseteq estMammifère &\stackrel{Def}{\equiv} int(estMammifère) \subseteq int(estChien) \end{aligned}$$

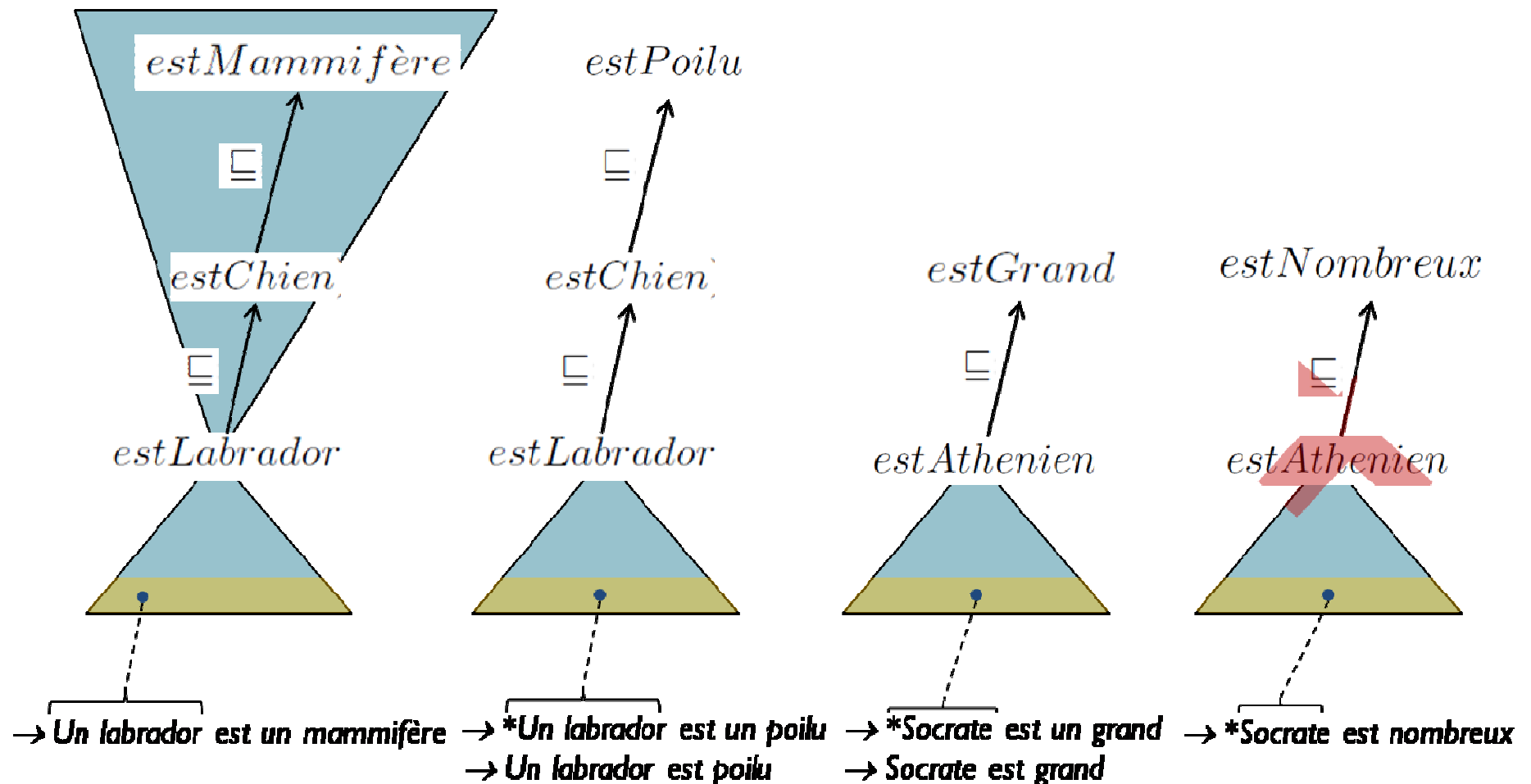


! : exemple d'approche intensionnelle: héritage des attributs dans les langages objets.

\sqsubseteq : relation d'ordre

2.1 Quelques notions générales pour la modélisation

- Exemples de schémas d'inférence



2.2 Cadre formel du web sémantique: les logiques de description

- **2 composantes:**
 - 1) Un langage définissant l'ensemble des *EBF*
 - 2) Une *interprétation sémantique* $I=(\Delta^I, \cdot^I)$ dans laquelle sont interprétées les EBF

2.2 *Cadre formel du web sémantique: les logiques de description*

- **1) Syntaxe des expressions de concepts**
 - ‘A’ est un concept *atomique*
 - ‘R’ est une relation binaire *atomique*
 - Et les concepts *complexes* se forment par:

$$C ::= A \mid \top \mid \perp \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \forall R.C \mid \exists R.C \mid \geq n R \mid \leq n R$$

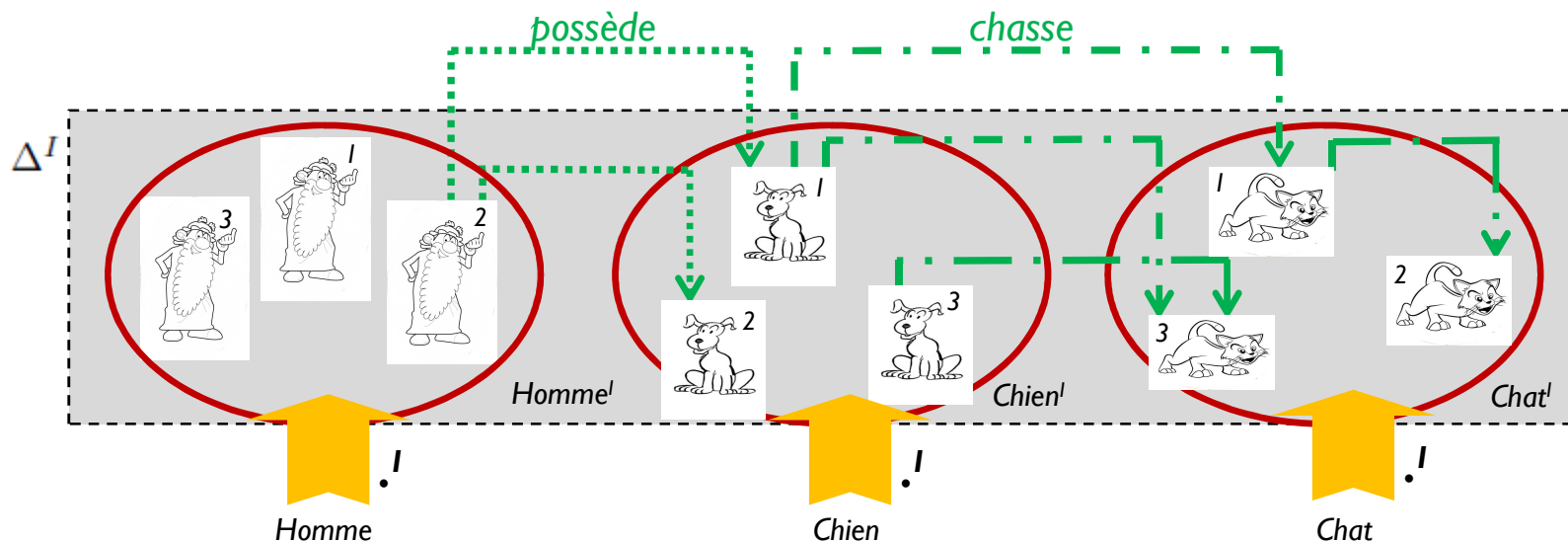
2.2 Cadre formel du web sémantique: les logiques de description

• 2) Interpretation sémantique $I=(\Delta^I, \cdot^I)$

- Δ^I correspond à l'univers (ensemble des éléments du modèle)
- La fonction d'interprétation \cdot^I assigne
 - un ensemble $A^I \subseteq \Delta^I$ à chaque concept atomique A ,
 - et un ensemble $R^I \subseteq \Delta^I \times \Delta^I$ à chaque relation atomique R .

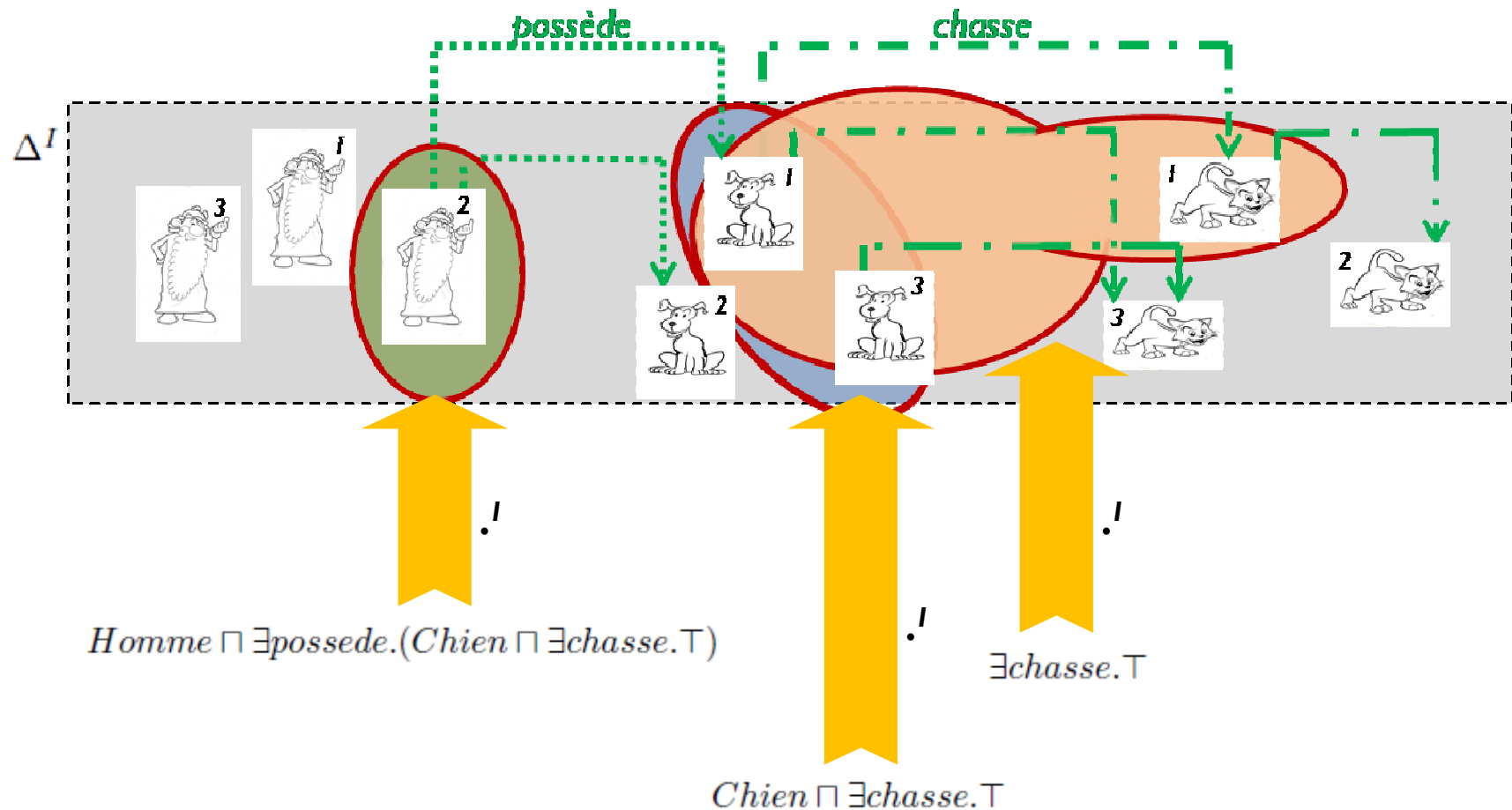
- La fonction d'interprétation est étendue aux concepts complexes par les définitions suivantes

- $\top^I = \Delta^I$, $\perp^I = \emptyset$
- $(\neg C)^I = \Delta^I / C^I$
- $(C \sqcap D)^I = C^I \cap D^I$, $(C \sqcup D)^I = C^I \cup D^I$
- $(\forall R.C)^I = \{a \in \Delta^I \mid \forall b, (a, b) \in R^I \rightarrow b \in C^I\}$
- $(\exists R.C)^I = \{a \in \Delta^I \mid \exists b, (a, b) \in R^I \wedge b \in C^I\}$
- $(\geq n R)^I = \{a \in \Delta^I \mid \text{Card}\{b \mid (a, b) \in R^I\} \geq n\}$
- $(\leq n R)^I = \{a \in \Delta^I \mid \text{Card}\{b \mid (a, b) \in R^I\} \leq n\}$



2.2 Cadre formel du web sémantique: les logiques de description

- Quelques exemples d'expressions de concepts complexes



2.2 Cadre formel du web sémantique: les logiques de description

- Une **base de connaissance** consiste en 2 types de formules:

- 1) générales, Tbox (intension) (axiomes de modélisation)

$$B \sqsubseteq C \text{ si } B^I \subseteq C^I$$

Exemple de TBox:

$Chien \sqsubseteq Mammifere, Chien \sqsubseteq \neg Homme$

$Chat \sqsubseteq Mammifere, Chat \sqsubseteq \neg Chien$

$Homme \sqsubseteq Mammifere$

$Chasseur \equiv \exists chase. \top$

$Maitre \equiv Homme \sqcap \exists possede. (Chien \sqcup Chat)$

- 2) individuelles, Abox (extension)

$$C(a) \text{ si } a^I \in C^I$$

$$R(a, b) \text{ si } \langle a^I, b^I \rangle \in R^I$$

Exemple de ABox:

$Chien(Rantanplan), Chat(Felix)$

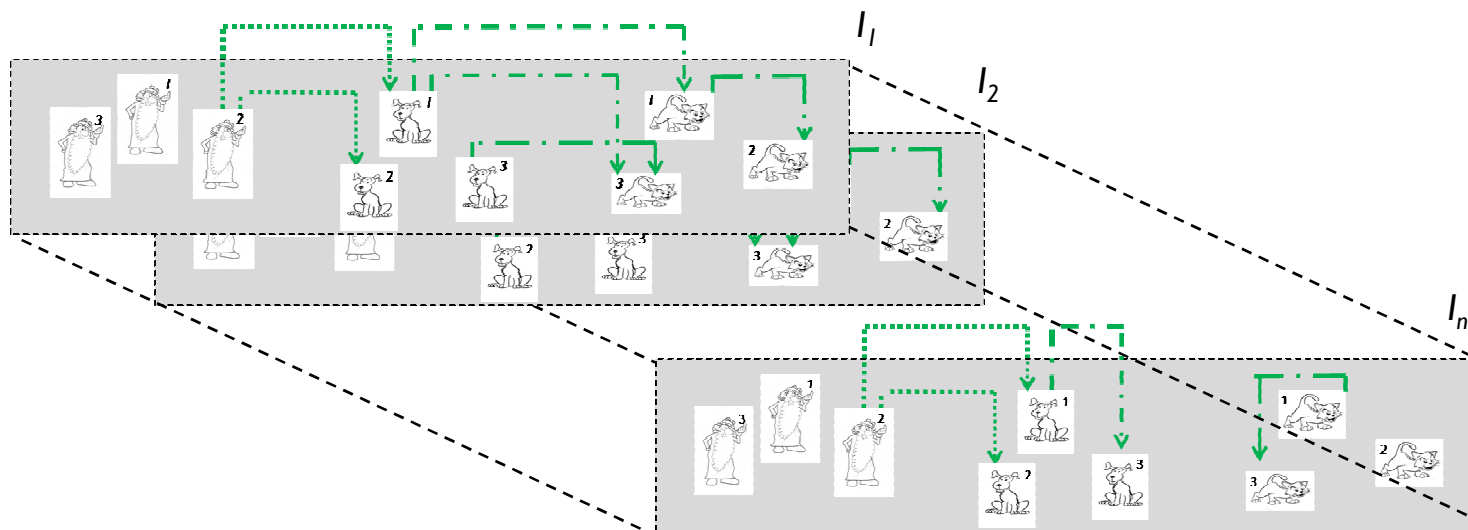
$Maitre(LuckyLuc)$

- Relation de **satisfaction**: une interprétation satisfait une formule

$$I \models Maitre(LuckyLuc), I \not\models Maitre(Rantanplan), I \models Chat \sqsubseteq \neg Chien$$

2.2 Cadre formel du web sémantique: les logiques de description

- Définition de la connaissance sur un **domaine** par l'introduction d'**axiomes** dans la TBox.
- Les objets « métiers » sont décrits par des propriétés
 - Ex: $Maitre \equiv Homme \sqcap \exists possede. (Chien \sqcup Chat)$
- Si φ est un axiome, alors $I_i \models \varphi$ pour tout i .



2.2 Cadre formel du web sémantique: les logiques de description

- Intérêt du cadre **logique** :

- Sémantique claire (pas forcément le cas avec d'autres formalismes, E/R...)
- Récupération de l'appareillage de raisonnement

- Tâches de **raisonnement**

- Classification automatique
- Connaissances explicites/implicites
- Vérification de la consistance
- Test de subsumption

$Maitre \equiv Homme \sqcap \exists possede.(Chien \sqcup Chat)$

$Labrador \sqsubseteq Chien \sqsubseteq Mammifere$

$Chat \sqsubseteq \neg Chien, Chien \sqsubseteq Chat$

$Chien \sqcap \exists chasse.T \sqsubseteq \exists chasse.T$

Démonstration de l'atelier de modélisation **Protégé**¹

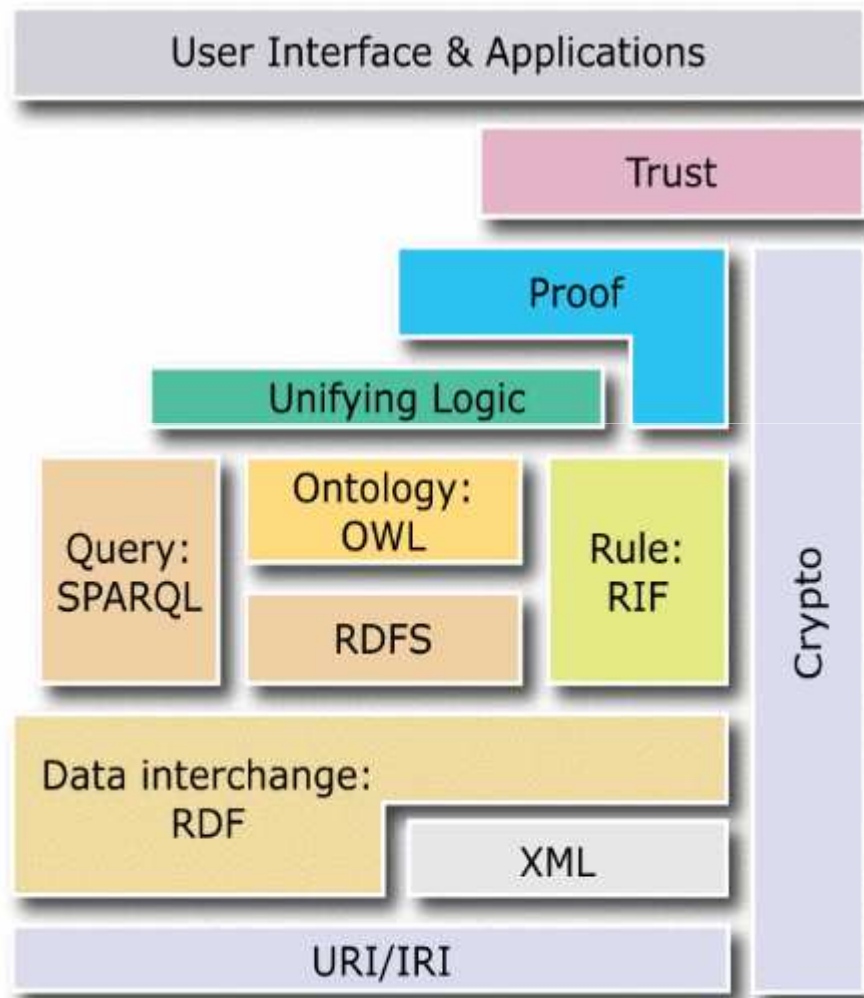
¹:<http://protege.stanford.edu/>

3. Aspects applicatifs

3. Serialisation XML

- Les logiques de description sont sérialisables selon une syntaxe **XML**
 - OWL2 (SROIQ)
 - OWL-QL (DL-Lite)
- Ces langages XML sont **spécifiés** et **normalisés** par le W3C
- Différents **profils** existent (rapport expressivité / inférence)

3. Pile de langages WS



- **Basic Technologies**

- *URI*
 - Uniform resource identifier
- *RDF*
 - Resource description framework
- *RDFS*
 - RDF Schema
- *OWL*
 - Web ontology language

3. Outils

- Bibliothèques pour **charger/manipuler/raisonner** sur du RDF/OWL
 - OWL Api
 - Jena (HP)
 - Raisonneurs (Fact++, Hermit, Pellet...)
- Langages d'interrogations
 - SPARQL

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX khi: <http://www.khiplus.fr#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
SELECT ?individu
WHERE {
    { ?individu rdf:type ?Chien }
    UNION
    { ?individu rdf:type ?Chat .
      ?individu khi:a_pour_age ?age }
}
FILTER ( ?age > "5"^^xsd:integer )

orderby ?individu
```

FIN

Screenshot: Protégé

