

Objets et représentation : un couple (encore) en devenir ?

Essai de nouvel essai sur les objets et la représentation ...

de la bande des cinq ...

Bernard Carré (LIFL Lille)

Roland Ducournau (LIRMM Montpellier),

Jérôme Euzenat (INRIA Rhône-Alpes, Grenoble)

Amedeo Napoli (LORIA Nancy)

François Rechenmann (INRIA Rhône-Alpes, Grenoble)

Histoires et points d'eau ...

- L'aube, quelque part ...
- Les mardis objets (1987), et les langages à objets (le livre, 1989).
- RPO-1992, puis LMO-1994, LMO-1995 (à Nancy), et cela dure encore ...
- En parallèle, un numéro spécial TSI sur les systèmes à objets (1996).
- Une école d'été à Nice en 1996, qui commence par une chanson ... et un livre LMO (encore !) qui en découle en 1998.
- Les années 2000 avec LMO qui traverse l'Atlantique, XML et sa bande qui pointent le bout de leur nez ...
- Un numéro spécial de L'Objet : objectif XML ou XML et les objets (2003).
- LMO-2004, dix ans déjà, et ensuite on verra bien ... (*failed* ou *not failed* ...)

Les langages de frames

- Le triplet *frame-slot-value* et le frame `Atom`, pas de distinction entre classes et instances (*prototype*).
- L'héritage multiple (et ses multiples parcours ..), le masquage, les valeurs par défauts, et les exceptions (raisonnement non monotone).
- Les réflexes, l'attachement procédural, la programmation dirigée par les accès, et le maintien de la cohérence.
- Les méta-connaissances.
- Les systèmes : KRL, FRL, et Shirka, seul dans sa catégorie (à faire en particulier la distinction schéma-instance).

Les langages hybrides

- La combinaison des réflexes et des méthodes.
- Les langages « tout en un » (comme les vestes ...).
- Les objets composites, les objets temporels, les versions, ...
- Les systèmes : LOOPS, KEE, Yafool, unique en son genre (« trois en un »), Objlog et les Prolog objets, ...

Mais aussi pendant ce temps là ...

- Les logiques terminologiques, qui deviennent " de descriptions " ...
- Le raisonnement par classification, les premiers pas des systèmes de représentation de connaissances par objets (RCO), dits aussi classificatoires.
- Mais aussi : les points de vue, TROPES d'abord, TROPES ensuite, la famille ROME (CROME, FROME, ...).
- Et un peu plus tard, Le Web sémantique.

Les bases des systèmes de connaissances

- L'objectif est de résoudre des problèmes dans un certain domaine, en exploitant des connaissances propres au domaine et des règles d'inférences.
- Les ingrédients d'un formalisme de représentation sont :
 - syntaxe + sémantique + règles d'inférences.
 - Les propriétés (hautement) souhaitables sont :
 - la *correction* et la *complétude* qui expriment la bonne adéquation entre la syntaxe et la sémantique.
 - Un modèle de référence possible est celui des *systèmes formels* en logique.

Vers la notion de système de RCO

Un système de RCO est à considérer comme une « logique » avec :

- une syntaxe, une sémantique et un système de déduction,
- une hiérarchie de concepts, ou classes, décrites en *intension* par un ensemble de *propriétés* et munies d'une interprétation *extensionnelle*,
- un ensemble d'assertions formant une description (incomplète) du monde,
- un ensemble d'opérations où figurent

l'*héritage* pour connaître les attributs d'un objet.

une *classification des espèces* (les classes), et une *classification des*

individus,

le *filtrage* pour déterminer les objets qui peuvent s'exprimer par l'intension

d'une classe,

la *complétion d'instances* pour donner une valeur aux attributs d'un objet.

Les concepts : vue intérieure

- Une *concept* C représente un ensemble d'individus (du monde réel) ayant des *propriétés* communes : il possède une *intension* (ensemble de ses propriétés) et une *extension* (ensemble de ses « instances », ou individus recouverts par le concept), et éventuellement une *identité*, un nom par l'intermédiaire duquel il est référencé.
- Une propriété exprime soit une relation avec un autre concept, son *co-domaine*, ou bien est à valeur dans un type de données (nombre, chaîne, ...).
- Un concept C *doit* posséder une sémantique qui lui est donnée (généralement) par une *fonction d'interprétation* $(C \rightarrow C^I)$ et un *domaine d'interprétation* ΔI .
- Un concept C n'a véritablement un sens que s'il peut posséder effectivement des instances.

Les concepts : vue extérieure

- Les concepts peuvent être organisés en une hiérarchie \mathcal{H}_C par une relation de *subsumption* (\sqsubseteq , ordre partiel).
- Un concept peut être considéré et introduit dans \mathcal{H}_C comme *primitif* : ses propriétés expriment des *conditions nécessaires* d'appartenance d'un individu à l'extension du concept, et l'introduction est *déclarée*.
- Un concept peut être considéré et introduit comme *défini* : ses propriétés expriment des *conditions nécessaires et suffisantes* d'appartenance d'un individu à l'extension du concept, et son insertion dans la hiérarchie \mathcal{H}_C est *calculée*.
- La mécanique de la classification repose sur le caractère nécessaire et suffisante de l'ensemble des propriétés d'un concept défini.

Correction et complétude de l'opération de classification

- Le processus de classification permet de placer un concept X dans la hiérarchie des concepts \mathcal{H} , en recherchant ses subsumants les plus spécifiques (SPS) et ses subsumés les plus généraux (SPG).
- Si deux concepts C et D vérifient $C \sqsubseteq D$, alors :
 - $\text{intension}(D) \subseteq \text{intension}(C)$,
 - $\text{extension}(C) \subseteq \text{extension}(D)$.
- Les deux propriétés duales suivantes *doivent* être vraies :
 - la *correction* : si $C \sqsubseteq D$ alors $\text{extension}(D) \subseteq \text{extension}(C)$ (l'extension correspond à l'interprétation).
 - la *complétude* : si $C \sqsubseteq D$ est vrai, alors le système de déduction doit le détecter.

La composition et la relation partie-de

- Une classe *composite* possède un ensemble de composants qui sont eux-mêmes des classes, éventuellement composites.
- La relation partie-de entre composants et composite est représentée par un *graphe de composition*.
- La gestion du partage de propriétés doit être contrôlée par des connaissances du domaine : la relation partie-de s'interprète de plusieurs façons différentes et licites, et la transitivité de la relation n'est pas systématique.

Les points de vue

- La relation de subsomption est un ordre partiel, elle est donc antisymétrique : deux concepts ayant la même extension sont nécessairement *identiques*.
- Si la relation n'est plus considérée comme antisymétrique, alors deux concepts de même extension peuvent posséder une description différente : ce qui donne naissance à *deux points de vue disjoints* et différents sur la même extension.
- L'exemple du système TROEPS : système générique pour la représentation de connaissances multi-points de vue, avec multi-instanciation, et raisonnement par classification multi-points de vue.

Caractère monothétique et polythétique

- Une classe C est dite *monothétique* si elle est caractérisée par une *conjonction de propriétés* qui définit une *condition nécessaire et suffisante* d'appartenance d'un individu à l'extension de C .
- Une classe est *polythétique* pour l'ensemble d'attributs $A = \{a_1, a_2, \dots, a_p\}$ si :
 - (i) chaque individu recouvert par la classe possède un « grand nombre » des attributs de A ,
 - (ii) chaque attribut de A est possédé par un « grand nombre » des individus *recouverts* par la classe,
 - (iii) il n'existe pas d'attribut (dans le cadre booléen) qui soit possédé par tous les individus *recouverts* par la classe.
- Vers le traitement de classes et d'objets *semi-structures*, qui incluent des disjonctions, avec une classification par similarité (ou une inertie intra-classe et une inertie inter-classes).

Il reste encore beaucoup d'autres choses ...

- objets temporels, ...
- méta-connaissances, réflexivité,
- ...
- et tout ce qui a été oublié ...
- et comme le temps passe, on passe ...

Juste en passant une histoire de musiciens ...

Une personne dont tous les enfants sont musiciens, et dont tous les enfants musiciens sont intermittents, doit nécessairement être une personne dont tous les enfants sont intermittents :

```
(and Personne
  (all enfant Musicien)
  (all (restrict enfant Musicien) Intermittent)
   $\bar{\square}$ 
  (and Personne
    (all enfant Intermittent))
```

Ce qui peut s'interpréter comme un *modus ponens* :

```
{musicien, musicien}  $\rightarrow$  {intermittent}  $\vdash$  intermittent
```

Le cadre des logiques de descriptions

- Les éléments du monde réel sont représentés par des *concepts*, des *rôles* et des *individus*, ou *instances*.
- Adéquation syntaxe – sémantique :
 - les concepts et les rôles possèdent une *description structurée* à laquelle est associée une *sémantique*,
 - les manipulations opérées sur les concepts et les rôles sont réalisées en accord avec la sémantique.
- Deux types de connaissances sont prises en compte : les concepts avec leurs composants (*TBox*), et les *faits* ou *assertions*, où interviennent les concepts et les instances (*ABox*).
- La relation de subsomption organise concepts et rôles en hiérarchies.
- Le raisonnement terminologique s'appuie sur la *classification* (concepts) et l'*instanciation* (ou *reconnaissance individus*).

Une histoire d'équipe ...

```
PERSONNE :< TOP
ENSEMBLE :< TOP
HOMME :< PERSONNE
FEMME :< (and PERSONNE (not HOMME))
membre :< toprole
chef :< membre
EQUIPE := (and ENSEMBLE
            (all membre PERSONNE)
            (atleast 2 membre))
PETITE-EQUIPE := (and EQUIPE
                  (atmost 5 membre))
EQUIPE-MODERNE := (and EQUIPE
                   (atmost 4 membre)
                   (atleast 1 chef)
                   (all chef FEMME))
```

La famille de logiques \mathcal{AL}

$$\mathcal{AL} = \{T, \perp, \neg, \forall, \exists, \sqcap, \sqcup, \exists r, \forall r\}$$

- $\mathcal{ALC} = \mathcal{AL} \cup \{\neg C\}$ (négation de concepts primitifs ou définis).
- $\mathcal{ALNI} = \mathcal{AL} \cup \{C \sqcup D\}$ (disjonction de concepts).
- $\mathcal{ALCE} = \mathcal{AL} \cup \{\exists r.C\}$ (quantification existentielle typée).
- $\mathcal{ALCN} = \mathcal{AL} \cup \{\geq n r, \leq n r\}$ (cardinalité des rôles).
- $\mathcal{ALR} = \mathcal{AL} \cup \{r_1 \sqcap r_2\}$ (conjonction de rôles).
- ...

La sémantique des concepts et des rôles pour *ALCNR*

Une *interprétation* $\mathcal{I} = (\Delta_{\mathcal{I}}, \cdot_{\mathcal{I}})$ est la donnée d'un ensemble $\Delta_{\mathcal{I}}$ appelé *domaine*

d'interprétation et d'une fonction d'interprétation $\cdot_{\mathcal{I}}$ qui fait correspondre à un concept un sous-ensemble de $\Delta_{\mathcal{I}}$ et à un rôle un sous-ensemble de $\Delta_{\mathcal{I}} \times \Delta_{\mathcal{I}}$, de telle sorte que les

équations suivantes soient satisfaites :

$$\begin{aligned} \top_{\Delta_{\mathcal{I}}} &= \Delta_{\mathcal{I}} \\ \top &= \top_{\mathcal{I}} \\ \emptyset &= \top_{\mathcal{I}} \\ (\mathcal{C} \sqcap \mathcal{D})_{\mathcal{I}} &= \mathcal{C}_{\mathcal{I}} \cap \mathcal{D}_{\mathcal{I}} \\ (\mathcal{C} \sqcup \mathcal{D})_{\mathcal{I}} &= \mathcal{C}_{\mathcal{I}} \cup \mathcal{D}_{\mathcal{I}} \\ (\mathcal{C} \setminus \mathcal{D})_{\mathcal{I}} &= \mathcal{C}_{\mathcal{I}} \setminus \mathcal{D}_{\mathcal{I}} \\ (\text{Ar}.\mathcal{C})_{\mathcal{I}} &= \{r \in \Delta_{\mathcal{I}} \mid \exists p \in \Delta_{\mathcal{I}} : pA/r \in \mathcal{C}_{\mathcal{I}}\} \\ (\text{Er}.\mathcal{C})_{\mathcal{I}} &= \{r \in \Delta_{\mathcal{I}} \mid \exists p \in \Delta_{\mathcal{I}} : pE/r \in \mathcal{C}_{\mathcal{I}}\} \\ (\geq n r)_{\mathcal{I}} &= \{r \in \Delta_{\mathcal{I}} \mid |\{p \in \Delta_{\mathcal{I}} \mid pA/r \in \mathcal{C}_{\mathcal{I}}\}| \geq n\} \\ (\leq n r)_{\mathcal{I}} &= \{r \in \Delta_{\mathcal{I}} \mid |\{p \in \Delta_{\mathcal{I}} \mid pA/r \in \mathcal{C}_{\mathcal{I}}\}| \leq n\} \\ (\mathcal{I} r_1 \sqcup \dots \sqcup \mathcal{I} r_n) &= \mathcal{I} r_1 \cup \dots \cup \mathcal{I} r_n \end{aligned}$$

Satisfiabilité, équivalence, incompatibilité, et subsomption

- Un concept C est *satisfiable* si et seulement s'il existe une interprétation \mathcal{I} telle que $C^{\mathcal{I}} \neq \emptyset$; C est *non satisfiable* sinon.
- Un concept C est *subsumé* par un concept D ($C \sqsubseteq D$), si et seulement si, pour toute interprétation \mathcal{I} , $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.
- Deux concepts C et D sont *équivalents* ($C \equiv D$) ssi $C^{\mathcal{I}} = D^{\mathcal{I}}$ pour toute interprétation \mathcal{I} ($C \equiv D$ ssi $D \sqsubseteq C$ et $C \sqsubseteq D$; $D \equiv C \sqcap D$ ssi $D \sqsubseteq C$).
- Deux concepts C et D sont *incompatibles* ssi $C^{\mathcal{I}} \cup D^{\mathcal{I}} = \emptyset$ pour toute interprétation \mathcal{I} ($D \sqcap C \sqsubseteq \perp$).

Base de connaissances et satisfiabilité

- L'interprétation $\mathcal{I} = (\Delta_{\mathcal{I}}, \cdot_{\mathcal{I}})$ est prolongée aux individus : l'assertion $C(a)$ est *satisfiable* pour \mathcal{I} ssi $a_{\mathcal{I}} \in C_{\mathcal{I}}$; l'assertion $r(a,b)$ est *satisfiable* par \mathcal{I} ssi $(a_{\mathcal{I}}, b_{\mathcal{I}}) \in r_{\mathcal{I}}$.

- Une *base de connaissances* $\Sigma = (T, \mathcal{A})$ se compose d'un ensemble T de déclarations de concepts (primitifs et définis) et d'un ensemble \mathcal{A}

d'assertions ($C(a)$ ou $r(a,b)$).

- Une interprétation \mathcal{I} est un *modèle* pour une base de connaissances ssi toute déclaration de concepts et toute assertion sont satisfiables pour \mathcal{I} : si \mathcal{I} existe alors la base est *satisfiable*.

et les objets où sont-ils ?

et tout cela pour quoi faire ? ...

- Les objets sont bien là mais pas nécessairement en pleine lumière ...
- Et l'on voit réapparaître un vieil ami, Lisp (en couleur ?), et les frames des débuts ...

Le cadre du Web sémantique

- *Un Web où les machines parlent aux machines et sont au service des personnes.*
- Demain, le Web sera exploité — en priorité — par des machines, qui traiteront les problèmes posés par des personnes, et qui délivreront les résultats obtenus à ces personnes.
- Le Web va devenir un *espace partagé, déclaratif et navigable*, où les machines exploitent les ressources disponibles pour résoudre des problèmes.
- Pour cela, il faut que les machines soient capables de « comprendre » les documents qu'elles manipulent : il faut pouvoir attacher une sémantique aux documents et y avoir accès, mais aussi doter les machines d'une certaine « intelligence ».

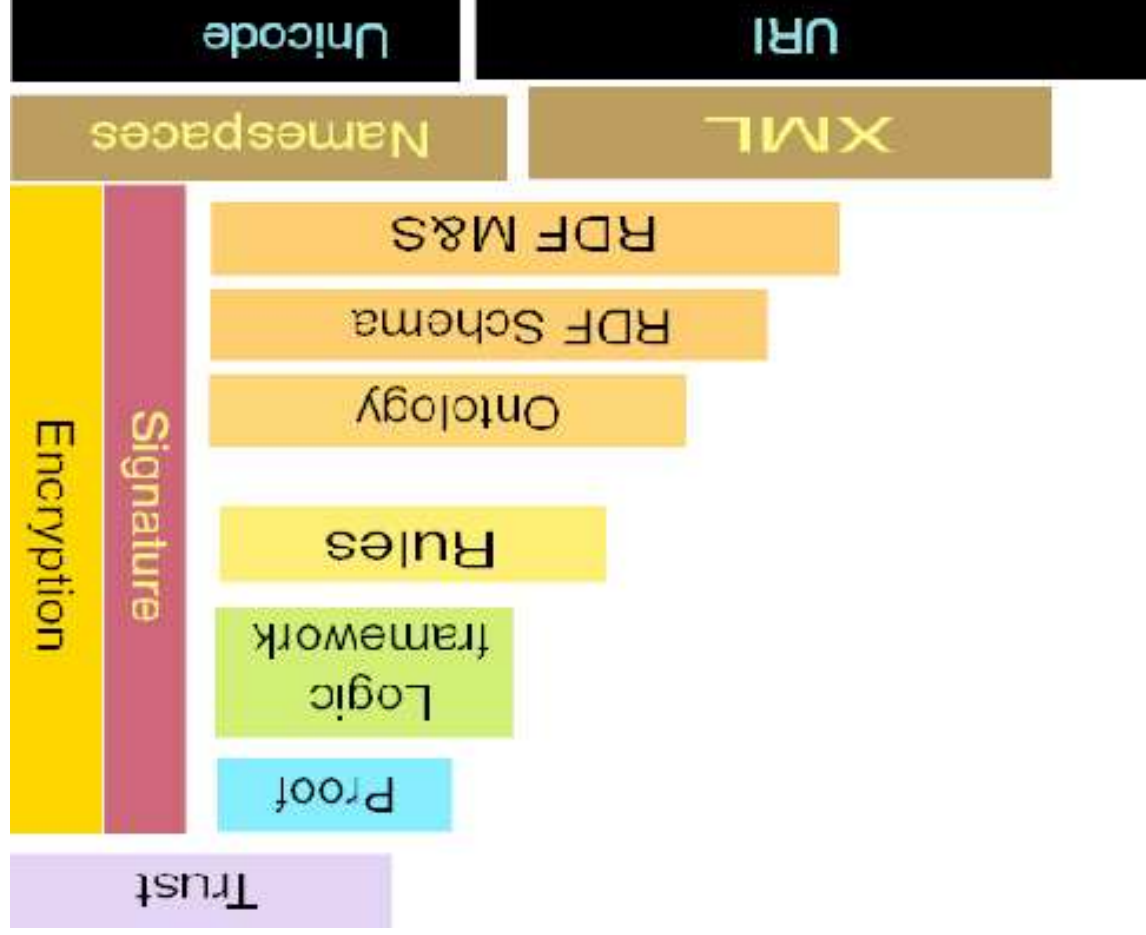
La manipulation de documents pour le Web sémantique

- Manipuler des documents en fonction de leur contenu et des connaissances disponibles sur le domaine.
- Donner une *sémantique* aux documents par l'intermédiaire d'*annotations* et de *méta-données* (*Dublin core*).
- document en contexte = description du (contenu du) document + annotations + méta-données + connaissances (du domaine).

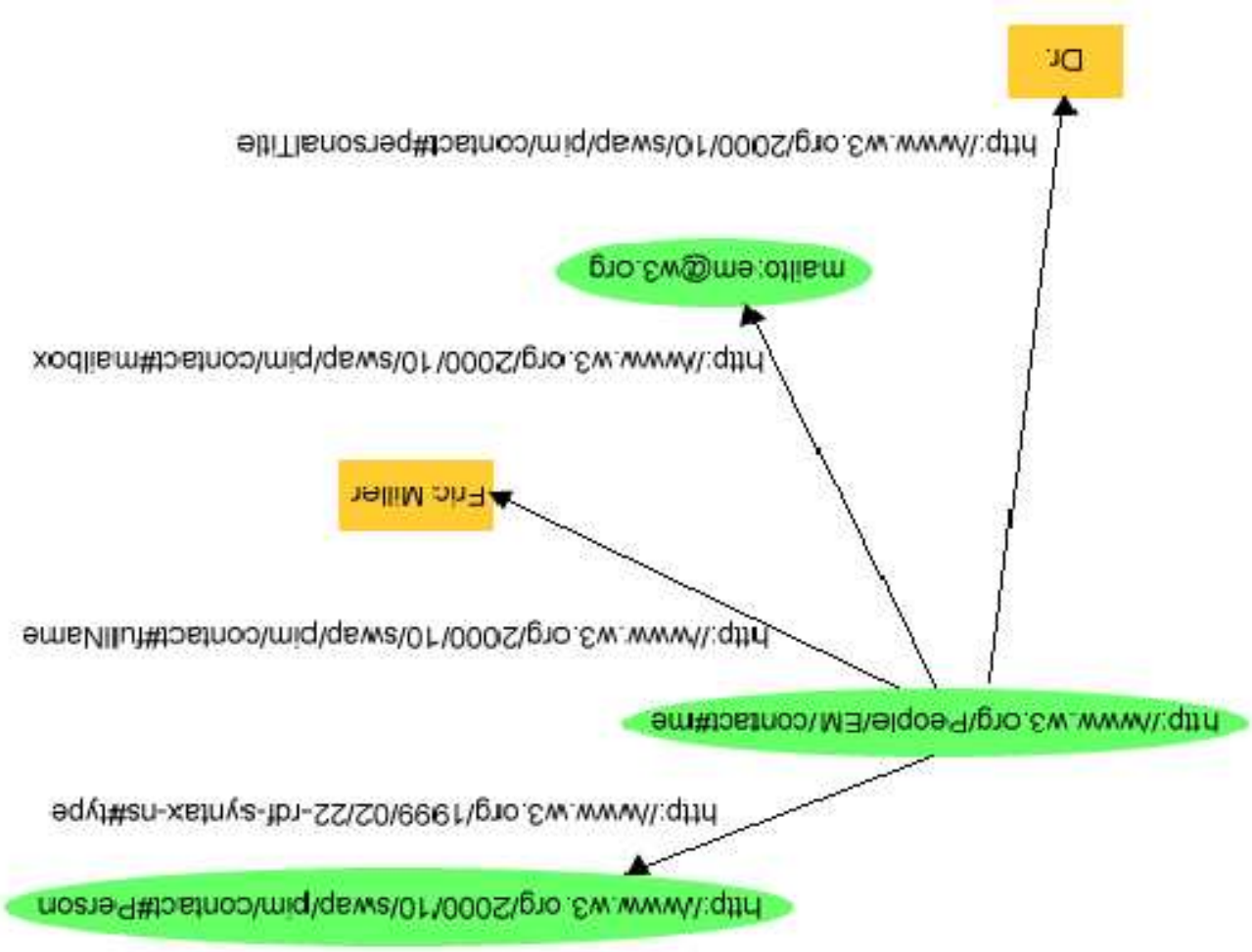
Quelques technologies à mettre en œuvre

- La description des documents et des ressources : URI, XML et RDF(S).
- La représentation des connaissances et le raisonnement : ontologies et langages de représentation (OIL, DAML+OIL, OWL).
- La gestion de (grandes) bases de données : stockage, requêtes, recherche d'information, intégration, données semi-structurées.
- L'extraction d'information dans les documents et la fouille de données (extraction de concepts à partir du contenu et classification).
- La mise en œuvre et le traitement des services (e-commerce, e-n'importe quoi...).

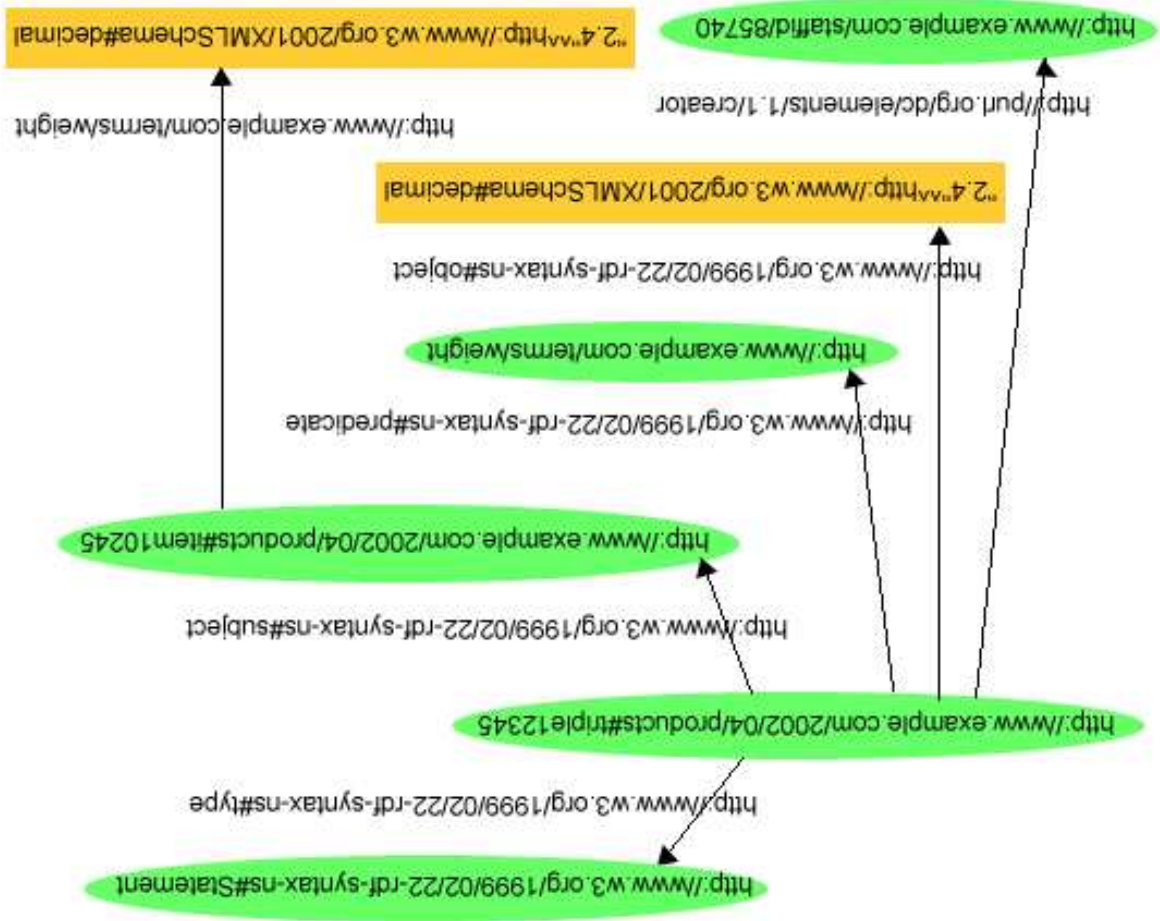
Le Web sémantique : du gâteau ...



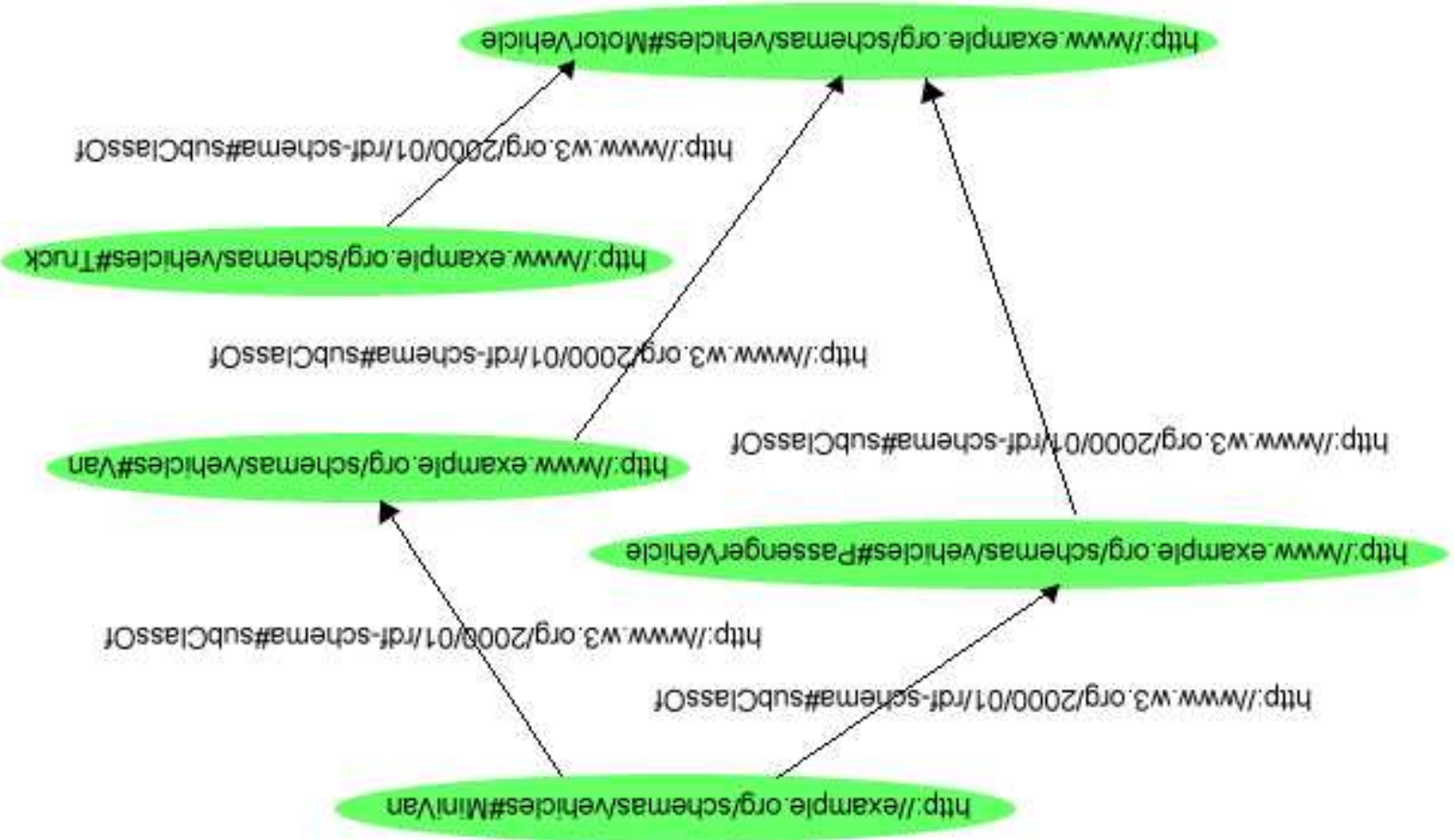
Un exemple de description en RDF



Un exemple de réification en RDF



Un exemple de hiérarchie en RDFS



La notion d'ontologie

- La *sémantique* des documents sur le Web doit être accessible aux machines.
- Un élément majeur de cette sémantique est constitué par un *modèle explicite* du domaine des données.
- Un tel modèle décrit le *vocabulaire* et la *structure* des informations relatives au domaine d'intérêt, qui doit être *communément admis et partageable* : c'est là l'essence même de la notion d'*ontologie*.

Logiques de descriptions :

le retour et c'est d'un $SHIQ$...

- S pour $ALC = \{T, \perp, C \sqcap D, C \sqcup D, \neg C, \text{Ar.C}, \text{Er.C}\}$.
- \tilde{Q} = pour $\{\geq n \text{ r.C}, \leq n \text{ r.C}\}$.
- \mathcal{H} = pour l'existence d'une hiérarchie de rôles,
 \mathcal{I} = pour les rôles inverses,
- on peut aussi avoir un $SHOQ$... avec $\mathcal{O} =$ pour les types énumérés,
et $SHOQ(\mathcal{D})$ où \mathcal{D} fait référence à un type de données concret (ou
domaine concret).
- $SHIQ$ est la logique de descriptions sur laquelle s'appuie le langage de
représentation d'ontologies OWL (Lite et DL).

Un exemple en OWL (ou les frames, le retour ...)

Student = Person $\sqcap \geq 1$ enrolledIn

```
<owl:Class rdf:ID="Student">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdfs:about="Person" />
    <owl:Restriction>
      <owl:onProperty rdf:resource="enrolledIn" />
      <owl:minCardinality rdf:type="xsd:Integer">
        1
      </owl:minCardinality>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```


Un début d'ontologie ...

```
<owl:Class rdf:about="#Opera">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasLibrettist" />
      <owl:minCardinality rdf:type="xsd:nonNegativeInteger">
        1
      </owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <owl:Restriction>
    <owl:Restriction>
      <rdfs:subClassOf rdf:resource="#Opera">
        <owl:Class rdf:ID="TraditionalItalianOpera">
          <rdfs:subClassOf rdf:resource="#Opera" />
        <rdfs:subClassOf>
          <owl:Restriction>
            <owl:onProperty rdf:resource="#hasOperaType" />
            <owl:Restriction>
              <owl:Restriction>
                <owl:Restriction>
                  <owl:Restriction>
                    <owl:someValuesFrom>
                      <owl:Class>
                        <owl:oneOf rdf:parseType="collection">
                          <owl:Thing rdf:about="#Operaseria" />
                          <owl:Thing rdf:about="#Operabuffa" />
                        </owl:oneOf>
                      </owl:Class>
                    </owl:Restriction>
                  </owl:Restriction>
                </owl:Restriction>
              </owl:Restriction>
            </owl:Restriction>
          </owl:Restriction>
        </owl:Restriction>
      </owl:Restriction>
    </owl:Restriction>
  </owl:Restriction>
  <owl:someValuesFrom>
    <owl:Class>
      <owl:oneOf rdf:parseType="collection">
        <owl:Thing rdf:about="#Operaseria" />
        <owl:Thing rdf:about="#Operabuffa" />
      </owl:oneOf>
    </owl:Class>
  </owl:someValuesFrom>
  </owl:Restriction>
</owl:Class>
```

Conclusion

- C'est fini pour aujourd'hui, mais il reste encore beaucoup à faire, de tous les côtés ...
- Une meilleure adéquation entre logiques (de descriptions) et représentation par objets.
- La prise en compte des problèmes de sémantique pour la représentation (des connaissances par objets) et pour le raisonnement (par classification).
- La prise en compte des comportements (par l'intermédiaire de niveaux) dans la représentation.
- La prise en compte des méta-connaissances, de la combinaison avec les bases de données, de la classification « polythétique » ...
- ...
- Et merci à tous ...