MDA — 29 avril 2004

Quelques mots sur les ontologies

Amedeo Napoli

Équipe Orpailleur

LORIA – UMR 7503

BP 239, 54506 Vandœuvre-lès-Nancy Cedex

(Email : `Amedeo.Napoli@loria.fr`)

`http://www.loria.fr/ napoli/`

`http://www.loria.fr/LORIA/EXT/equipes/ORPAILLEUR/`

Données – informations – connaissances

- *Donnée* : un signe ou un signal non interprété, par exemple `...---...` en alphabet Morse, `a`, `o` ou `eau` en alphabet latin.
- *Information* : une donnée et un sens associé, par exemple `(...---...,SOS)`, `(feu-rouge,stop)`, ...

  Pour un daltonien, la donnée “feu rouge” est la même que pour nous, mais l’information est (peut être) différente.
- *Connaissance* : l’information avec son mode d’emploi (compétence) pour atteindre un certain but, ce qui implique un contexte et un processus de génération (de production de nouvelles connaissances).

  Par exemple : si le feu est rouge je dois impérativement m’arrêter ;

  si l’information `(...---...,SOS)` est émise, alors il faut déclencher les secours ...

Pour info ...

- La commune n'est pas morte !

  Ça va péter – Chorale "les sans noms"

- V'là les dates de concert :

  1er mai, 12h30 : CNT Pépinière

  1er mai, 17h30 : CGT Parc Charles 3

  4 juin - Grand Sauvoy

  6 juin - Zénith Paris (hé hé hé)

- 21 juin, Fête de la Zique / Le refuge

La notion d’ontologie

- De *ontos – être* et *logos – science* : connaissance de ce qui est, de l’être en soi : l’ontologie se confond avec la métaphysique.
- Intelligence artificielle : une ontologie est la “spécification d’une conceptualisation”.

  Elle peut être utilisée pour fournir une spécification concrète des termes d’un domaine avec une sémantique associée.
- Exemple : un dictionnaire avec des couples entrées – définition.

  Le problème ensuite est de représenter ces termes et leur sémantique : c’est exactement là le propos de la représentation des connaissances.

Les logiques de descriptions

Les grands principes

- Les éléments du monde réel sont représentés par des *concepts*, des *rôles* et des *individus*.
- Adéquation syntaxe – sémantique :

  - les concepts et 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, et les *faits* ou *assertions*, où interviennent les concepts et les instances de concepts.
- La relation de subsomption organise concepts et rôles en hiérarchies.
- Raisonnement : *classification* et *instanciation*.

Un exemple simple

Un **Homme** est une **Personne**.

Une **Femme** est une **Personne**.

Aucune **Femme** n’est un **Homme** et vice-versa.

Une **Équipe** est (définie comme) un **Ensemble** ayant au moins 2 **membres** qui sont tous des **Personnes**.

Une **Petite-équipe** est (définie comme) une **Équipe** ayant au plus 5 **membres**.

Une **Équipe-moderne** est (définie comme) une **Équipe** ayant au plus 4 **membres**, ayant au moins 1 **chef**, et dont tous les **chefs** sont des **Femmes**.

$\mathcal{ALN}$: description des des concepts et des rôles

Deux types de concepts :

$\longrightarrow$ Les *concepts primitifs* où les rôles déterminent des conditions nécessaires (d'appartenance à l'extension du concept).

$\longrightarrow$ Les *concepts définis* où les rôles déterminent des conditions nécessaires et suffisantes (d'appartenance à l'extension du concept).

concept $\longrightarrow$ identificateur $|$
(**and** concept$^+$) ou (concept $\sqcap$ concept) $|$
(**not** identificateur) ou $\neg$ identificateur $|$
(**all** role concept) ou $\forall \texttt{role.concept}$ $|$
($($**atleast** entier-positif role) ou $(\geq \texttt{entier-positif role})$ $|$
(**atmost** entier-positif role) ou $(\leq \texttt{entier-positif role})$

role primitif $\longrightarrow$ identificateur

La représentation en $\mathcal{ALN}$

```
PERSONNE ⊑ TOP (primitif)
ENSEMBLE ⊑ TOP (primitif)
HOMME ⊑ PERSONNE (primitif)
FEMME ⊑ PERSONNE (primitif)
(and FEMME HOMME) ⊑ ⊥

ÉQUIPE := (and ENSEMBLE
              (all membre PERSONNE)
              (atleast 2 membre)) (défini)
PETITE-ÉQUIPE := (and ÉQUIPE
                     (atmost 5 membre)) (défini)
ÉQUIPE-MODERNE := (and ÉQUIPE
                      (atmost 4 membre)
                      (atleast 1 chef)
                      (all chef FEMME)) (défini)
```

Problèmes (entres autres) :

- `HOMME` et `FEMME` sont ils des concepts primitifs ou définis ?

- Comment représenter la relation entre `membre` et `chef` ?

$$\texttt{membre} \sqsubseteq \texttt{toprole}$$
$$\texttt{chef} \sqsubseteq \texttt{membre}$$

La famille des logiques $\mathcal{AL}$

$$\mathcal{AL} = \{\top,\ \bot,\ \neg\mathtt{A},\ \mathtt{C} \sqcap \mathtt{D},\ \forall\mathtt{r}.\mathtt{C},\ \exists\,\mathtt{r}\}$$

- $\mathcal{ALC} = \mathcal{AL} \cup \{\neg\mathtt{C}\}$ (négation de concepts primitifs ou définis).
- $\mathcal{ALU} = \mathcal{AL} \cup \{\mathtt{C} \sqcup \mathtt{D}\}$ (disjonction de concepts).
- $\mathcal{ALE} = \mathcal{AL} \cup \{\exists\mathtt{r}.\mathtt{C}\}$ (quantification existentielle typée).
- $\mathcal{ALN} = \mathcal{AL} \cup \{\geq\mathtt{n}\,\mathtt{r},\ \leq\mathtt{n}\,\mathtt{r}\}$ (cardinalité des rôles, en particulier, $(\exists\,\mathtt{r}) \equiv (\geq\mathtt{1}\,\mathtt{r})$).
- $\mathcal{ALR} = \mathcal{AL} \cup \{\mathtt{r}_1 \sqcap \mathtt{r}_2\}$ (conjonction de rôles).
- Et encore les rôles inverses ($\mathcal{I}$), une hiérarchie de rôles ($\mathcal{H}$), les constructeurs `domain` et `range` pour les rôles, ou restriction sur les domaines et co-domaines ($\mathcal{Q}$) ...

La sémantique des concepts et des rôles en $\mathcal{ALCN}$

Une *interprétation* $\mathcal{I} = (\Delta_{\mathcal{I}}, .^{\mathcal{I}})$ est la donnée d'un ensemble $\Delta_{\mathcal{I}}$ appelé *domaine d'interprétation* et d'une fonction d'interprétation $.^{\mathcal{I}}$ 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 :

$$
\begin{aligned}
&\top^{\mathcal{I}} = \Delta_{\mathcal{I}} \\
&\bot^{\mathcal{I}} = \emptyset \\
&(\mathtt{C} \sqcap \mathtt{D})^{\mathcal{I}} = \mathtt{C}^{\mathcal{I}} \cap \mathtt{D}^{\mathcal{I}} \\
&(\mathtt{C} \sqcup \mathtt{D})^{\mathcal{I}} = \mathtt{C}^{\mathcal{I}} \cup \mathtt{D}^{\mathcal{I}} \\
&(\neg \mathtt{C})^{\mathcal{I}} = \Delta_{\mathcal{I}} \setminus \mathtt{C}^{\mathcal{I}} \\
&(\forall \mathtt{r}.\mathtt{C})^{\mathcal{I}} = \{d \in \Delta_{\mathcal{I}} / \forall d' : (d,d') \in \mathtt{r}^{\mathcal{I}} \rightarrow d' \in \mathtt{C}^{\mathcal{I}}\} \\
&(\exists \mathtt{r}.\mathtt{C})^{\mathcal{I}} = \{d \in \Delta_{\mathcal{I}} / \exists d' : (d,d') \in \mathtt{r}^{\mathcal{I}} \wedge d' \in \mathtt{C}^{\mathcal{I}}\} \\
&(\geq \mathtt{n}\, \mathtt{r})^{\mathcal{I}} = \{d \in \Delta_{\mathcal{I}} / |\{d' \in \Delta_{\mathcal{I}} / (d,d') \in \mathtt{r}^{\mathcal{I}}\}| \geq n\} \\
&(\leq \mathtt{n}\, \mathtt{r})^{\mathcal{I}} = \{d \in \Delta_{\mathcal{I}} / |\{d' \in \Delta_{\mathcal{I}} / (d,d') \in \mathtt{r}^{\mathcal{I}}\}| \leq n\}
\end{aligned}
$$

Remarque : $(\mathtt{r}\,|\mathtt{C})^{\mathcal{I}} = \{(d,d') \in \Delta \times \Delta / \ (d,d') \in \mathtt{r}^{\mathcal{I}} \wedge d' \in \mathtt{C}^{\mathcal{I}}\}$.

La satisfiabilité et la subsomption

- Un concept $\mathtt{C}$ est *satisfiable* si et seulement s'il existe une interprétation $\mathcal{I}$ telle que $\mathtt{C}^{\mathcal{I}} \neq \emptyset$; $\mathtt{C}$ est *non satisfiable* sinon.
- Un concept $\mathtt{C}$ est *subsumé* par un concept $\mathtt{D}$, ou bien $\mathtt{D}$ *subsume* $\mathtt{C}$, ce qui se note $\mathtt{C} \sqsubseteq \mathtt{D}$, si et seulement si, pour toute interprétation $\mathcal{I}$, $\mathtt{C}^{\mathcal{I}} \subseteq \mathtt{D}^{\mathcal{I}}$.
- $\mathtt{D}$ est appelé le concept *subsumant* et $\mathtt{C}$ le concept *subsumé*.

  Exemple : $\mathtt{ÉQUIPE{-}MODERNE} \sqsubseteq \mathtt{PETITE{-}ÉQUIPE}$.

Assertions et instanciation

- Instanciation de concept : $\mathtt{C}(\mathtt{i})$
- Instanciation de rôle : $\mathtt{R}(\mathtt{i},\mathtt{j})$

Exemple:

```
EQUIPE-MODERNE(Trio-54)
HOMME(Paul)
membre(Trio-54,Paul)
membre(Trio-54,Jean)
chef(Trio-54,Julia)
(atmost 3 membre)(Trio-54)
```

Base terminologique et satisfiabilité

- Si $\mathcal{I} = (\Delta_{\mathcal{I}}, .^{\mathcal{I}})$ est une interprétation, l'assertion $\mathtt{C}(\mathtt{a})$ est *satisfaite* par $\mathcal{I}$, ou *satisfiable*, si et seulement si $\mathtt{a}^{\mathcal{I}} \in \mathtt{C}^{\mathcal{I}}$; l'assertion $\mathtt{r}(\mathtt{a},\mathtt{b})$ est *satisfaite* par $\mathcal{I}$, ou *satisfiable*, si et seulement si $(\mathtt{a}^{\mathcal{I}},\mathtt{b}^{\mathcal{I}}) \in \mathtt{r}^{\mathcal{I}}$.
- Une *base (de connaissances) terminologique* Σ pour $\mathcal{ALCNR}$ est un ensemble d'assertions de la forme $\mathtt{C}(\mathtt{a})$ ou $\mathtt{r}(\mathtt{a},\mathtt{b})$.
- Une interprétation $\mathcal{I}$ est un *modèle* pour Σ si et seulement si toute assertion α est satisfaite par $\mathcal{I}$.
- Une base terminologique est *satisfiable* si elle admet un modèle.
- L'assertion α est *logiquement impliquée* par la base Σ, ce qui s'écrit $\Sigma \models \alpha$, si et seulement si α est satisfaite par tout modèle de Σ.

Les opérations à la base du raisonnement terminologique

- Le *test de subsomption* permet de vérifier qu'un concept $\mathtt{C}$ subsume un concept $\mathtt{D}$ (base de l'opération de *classification* et de la recherche de propriété pour un concept).
- Le *test de satisfiabilité d'un concept* $\mathtt{C}$ permet de vérifier qu'un concept $\mathtt{C}$ admet des instances.
- Le *test de satisfiabilité d'une base terminologique* Σ permet de vérifier que Σ admet un modèle $\mathcal{I}$.
- Le *test d'instanciation* permet de vérifier qu'un individu $\mathtt{a}$ est instance d'un concept $\mathtt{C}$ pour une base terminologique Σ.

Deux méthodes pour tester la subsomption

- Les *algorithmes de type normalisation–comparaison* (algorithmes NC) : un processus de normalisation produit les *formes normales* des concepts définis qui sont ensuite effectivement comparées (à l'aide de règles de comparaison).
- La *méthode des tableaux sémantiques* : la question *est-ce que* $\mathtt{D}$ *subsume* $\mathtt{C}$ est transformée en *est-ce que* $\mathtt{C} \sqcap \neg\mathtt{D}$ *est non satisfiable*.

  Note : le langage de description des concepts doit être muni de la négation des concepts définis.

La normalisation dans $\mathcal{ALN}$

- Factorisation des conjonctions de concepts :

  $(\mathtt{C}_1 \sqcap \mathtt{C}_2) \sqcap \mathtt{C}_3 \rightarrow \mathtt{C}_1 \sqcap \mathtt{C}_2 \sqcap \mathtt{C}_3$

- Factorisation des co-domaines :

  $\forall \mathtt{r}.(\mathtt{C}_1 \sqcap \mathtt{C}_2) \sqcap \forall \mathtt{r}.(\mathtt{C}_3 \sqcap \mathtt{C}_4) \rightarrow \forall \mathtt{r}.(\mathtt{C}_1 \sqcap \mathtt{C}_2 \sqcap \mathtt{C}_3 \sqcap \mathtt{C}_4)$

- Traitements des expressions insatisfiables :

  $\mathtt{A} \sqcap \neg \mathtt{A} \rightarrow \perp$

  si $\mathtt{r}_2 \sqsubseteq \mathtt{r}_1$ et si $\mathtt{n}_1 < \mathtt{n}_2$, alors $(\leq \mathtt{n}_1\, \mathtt{r}_1) \sqcap (\geq \mathtt{n}_2\, \mathtt{r}_2) \rightarrow \perp$

  avec $(\forall \mathtt{r}.\perp) \rightarrow (\leq 0\, \mathtt{r})$

Quelques règles de comparaison pour $\mathcal{ALNH}$

[ALL]

Si $\mathtt{s} \sqsubseteq \mathtt{r}$ et $\mathtt{C} \sqsubseteq \mathtt{D}$ alors $\forall \mathtt{r}.\mathtt{C} \sqsubseteq \forall \mathtt{s}.\mathtt{D}$

$\forall \mathtt{enfant}.\mathtt{DOCTEUR} \sqsubseteq \forall \mathtt{fils}.\mathtt{BACHELIER}$

[ATLEAST]

Si $\mathtt{n} \leq \mathtt{m}$ et $\mathtt{s} \sqsubseteq \mathtt{r}$ alors $(\geq \mathtt{m}\ \mathtt{s}) \sqsubseteq (\geq \mathtt{n}\ \mathtt{r})$

$(\geq 2\ \mathtt{fils}) \sqsubseteq (\geq 1\ \mathtt{enfant})$

[ATMOST]

Si $\mathtt{m} \leq \mathtt{n}$ et $\mathtt{r} \sqsubseteq \mathtt{s}$ alors $(\leq \mathtt{m}\ \mathtt{s}) \sqsubseteq (\leq \mathtt{n}\ \mathtt{r})$

$(\leq 1\ \mathtt{enfant}) \sqsubseteq (\leq 2\ \mathtt{fils})$

Complexité de la subsomption

$\mathtt{C} := (\geq 1\, \mathtt{fille}) \sqcap (\geq 1\, \mathtt{fils})$

$\mathtt{D} := (\geq 2\, \mathtt{enfant})$

Question : $\mathtt{C} \sqsubseteq \mathtt{D}$?

Correction : les inférences produites sont en accord avec la sémantique associée, autrement dit, ce qui est vrai sur le plan syntaxique l'est sur le plan sémantique.

Complétude : toutes les formules valides — vraies sur le plan sémantique — peuvent être démontrées sur le plan syntaxique.

Le calcul du test de subsomption est un problème NP-difficile ou NP-complet, selon la richesse de l'ensemble des constructeurs associés à la logique utilisée.