

Représentation des Connaissances : Introduction aux Réseaux Sémantiques

Bernard ESPINASSE
Professeur à l'Université d'Aix-Marseille

2008

Plan

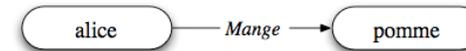
- Définition et concepts des Réseaux Sémantiques
- Héritage, partition, quantification dans les Réseaux Sémantiques
- Logique & Réseaux Sémantiques
- Interprétation des connaissances dans les Réseaux Sémantiques
- Forces et faiblesses des Réseaux Sémantiques

Les Réseaux Sémantiques : définition

- conçus à l'origine en **linguistique** pour devenir ensuite un **langage pour la représentation de concepts** très divers, une **structure informatique** utilisée en IA (QUILLIAN / COLLINS 1966)
- **un réseau sémantique est un graphe composé** :
 - d'un ensemble de **noeuds** étiquetés : *représentant généralement des objets*,
 - d'un ensemble de **liens** orientés et étiquetés entre ces noeuds : *représentant généralement des relations entre des objets*,
 - d'un ensemble **d'opérations** d'exploitation de ce graphe : *constituants les mécanismes de raisonnement*

représentation graphique:

facilite la lecture, ne correspond généralement pas au formalisme d'implémentation,



représentation non-graphique:

(alice, manger, pomme)

Concepts de base des Réseaux Sémantiques (RS)

les NOEUDS

- **atomiques** : entités élémentaires (valeurs, individus,...)
- **complexes** : entités complexes (propositions, phrases,...)
- ils doivent être **typés** : concept, individu, action, proposition, etc...

les LIENS

- **structuraux**: indépendants de la sémantique du domaine,
- **spécifiques**: dépendants de la sémantique du domaine,

il faut essayer d'augmenter la proportion des liens structuraux par rapport aux liens spécifiques

les OPERATIONS

- souvent représentées par le **programme**,
- doivent être **définies clairement**,

Noeuds concepts

"les canaris / sont des / oiseaux"

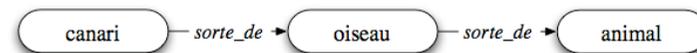
canaris et oiseaux =

concepts (nom communs) --> classe

sont des =

relation --> inclusion de classes

lien « sorte_de »



- lien **structurel** indépendant du domaine
- représente une **inclusion**
 - de **propriétés** (pt de vue intentionnel, cas général)
 - d'**individus** (pt de vue extensionnel)

Noeuds individus

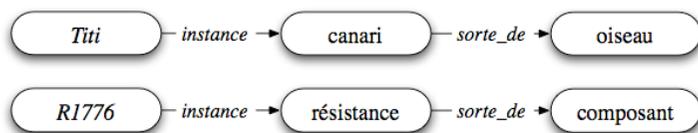
"Titi / est un / canari"

canari = concepts

Titi = individu (nom propre) --> élément d'un ensemble

est un = relation --> appartenance d'un élément à une classe

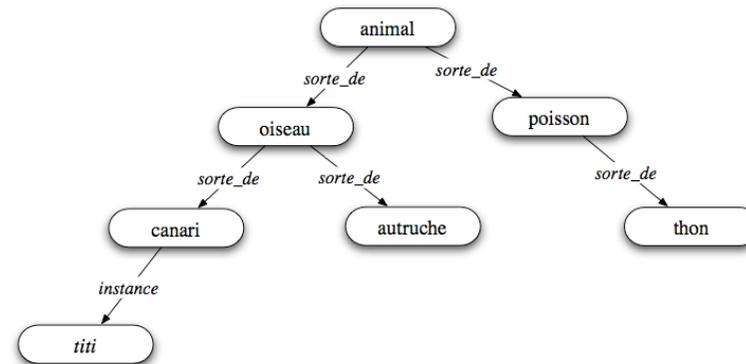
lien « instance »



lien « instance » = lien **structurel**

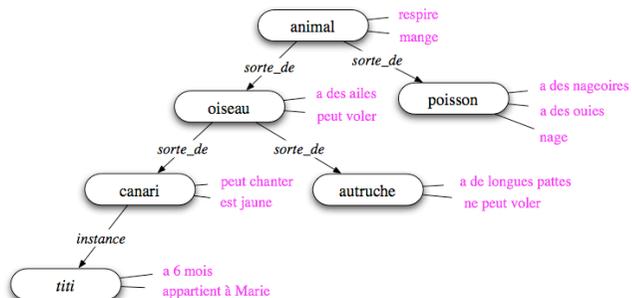
Les réseaux sémantiques : taxonomies

• liens **sorte_de** et noeuds **concepts** + liens **d'instance** et noeuds **individu** :



Propriétés

Les **propriétés** sont des informations rattachées à chaque nœud du RS :



• simples

• elles ne permettent pas de répondre à des questions comme :

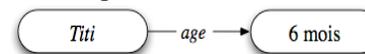
"quel est l'âge de Titi ?" "quelle est la couleur des canaris ?"

-> **notion d'attribut**

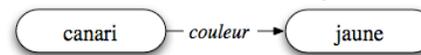
Attributs (1)

• **attribut** = relation qui relie un **noeud concept** ou un **noeud individu** à une **valeur** ou **propriété**

« l'âge de Titi est de 6 mois »

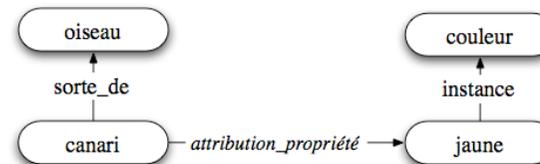


« la couleur des canaris est le jaune »



• **lien spécifique** dont le sens dépend du domaine d'application -> interprétation ad-hoc,

• on peut le rendre plus **structurel** en créant un **noeud-attribut**:

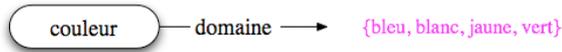


Attributs (2)

Notion **d'attribut** :

- une classe **sémantique** de nœud dont les instances sont des **propriétés**

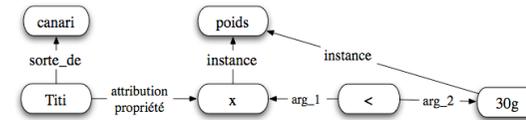
- un **attribut** peut lui-même être **caractérisé** :



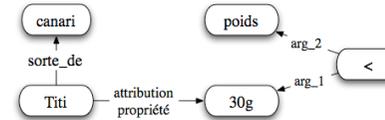
- **domaine** = relation structurelle permettant de vérifier des contraintes d'intégrités.

Rapports attribut / valeur

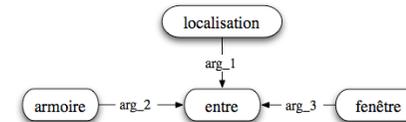
- un **noeud-attribut** peut être relié à une ou plusieurs valeurs par l'intermédiaire d'un opérateur:



soit en simplifiant:



- cet opérateur peut être n-aire:



Héritage dans les RS

L'héritage dans les RS [Collins & Quillan] repose sur des liens de type « **est_un** » ou « **sorte_de** » reliant un concept à un autre concept plus élevé :

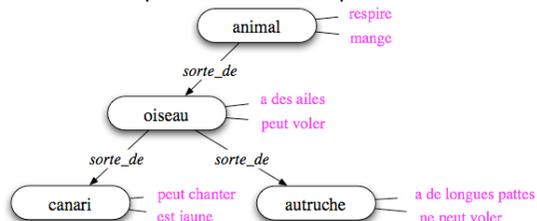
exemple: "canari" est une sorte de "oiseau"

- **héritage des propriétés** rattachées au concept père au concept fils :

Ainsi, on pourra dire que « le canari a des ailes et une peau » en remontant les liens « sorte_de »

- le principe d'héritage permet :

- de **nombreuses déductions** automatiques
- de définir la notion de **distance sémantique** entre 2 concepts = nombre de liens devant être traversés pour aller d'un concept à l'autre.

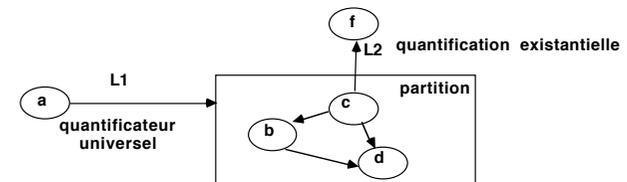


Partition dans les RS

- **Partition** [Hendrix] = regroupement de noeuds et d'arcs du réseau dans des espaces spécifiant la portée de relations

- **intérêts des partitions** :

- définition de **contextes**
- permet la **quantification**



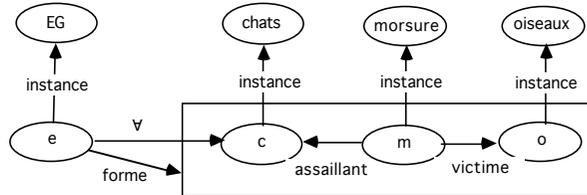
- **cadres** : définissent l'étendue des identificateurs universels
- **lien L1** : quantification universelle, quelque soit a, pointe sur un cadre représentant l'étendue de la variable quantifiée universellement.
- **lien L2** : quantificateur existentiel explicite sur le noeud f par rapport au noeud c

(certains systèmes experts (Prospector) l'utilise)

RS et quantification

- **quantification** traitée par la notion de **partition**

- Soit le fait à représenter suivant : « **tout chat a mordu un oiseau** »
 - représentation logique : $\forall x \text{ chat}(x) \rightarrow (\exists y \text{ oiseau}(y) \wedge \text{mordre}(x,y))$
 - **encodage de la variable quantifiée universellement x** en utilisant une partition (cadre rectangulaire) :

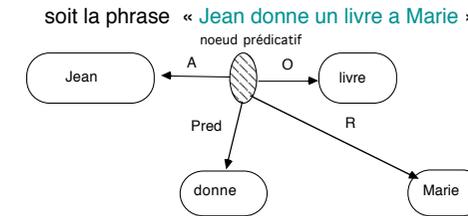


- les nœuds c, m, o sont des instances de chats, morsure, oiseaux,
- le cadre introduit dans le réseau définit l'étendue de l'identificateur universel,
- le nœud e représente l'assertion à représenter, instance de l'ensemble des énoncés généraux EG sur le monde,
- chaque élément de EG possède :
 - une connexion « **forme** » pointant vers le cadre de la partition et énonce l'affirmation,
 - une ou plusieurs connexions « **∀** » pointant vers chaque variable quantifiée universellement, ici variable c (les variables m et o sont ici quantifiées existentiellement).

Logique & RS

Travaux de [Schubert & Cerone]

- introduction de la **logique des prédicats du premier ordre**
- introduction d'un "**nœud prédicatif**" instancié en lui associant :
 - un pointeur vers le prédicat
 - un pointeur vers chaque argument du prédicat
- Ex :

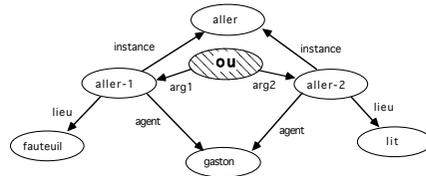


soit : **donne (A, O, R)**

avec : A (agent) = Jean; O (objet) = livre; R (receveur) = Marie

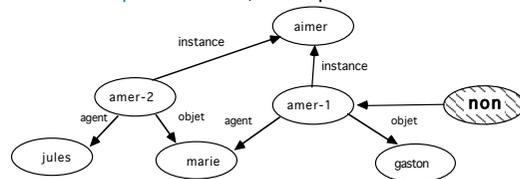
Connecteurs logiques ET,OU

- Soit la phrase suivante: « **Gaston ira sur le lit ou sur le fauteuil** », une représentation :



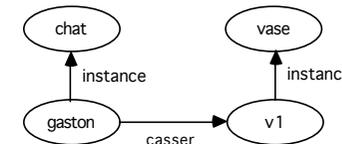
Représentation de la négation

- Soit la phrase: « **Marie n'aime pas Gaston** », une représentation :

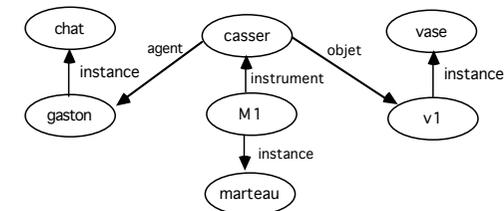


Représentation d'évènements ou d'actions

- Représenter l'évènement : « **Gaston casse le vase** » :



le lien « **casser** » est **spécifique**. On peut s'en séparer en le traduisant par des liens plus structurels : agent, objet, instrument, temps, lieu, :



Interprétation des connaissances dans les RS

- l'accès aux données stockées dans un RS n'est **pas assuré par le réseau lui-même**
- appel à un **interpréteur** (transformer les données du réseau en connaissances opératoires)

• on doit disposer alors de:

- soit d'un **langage élaboré de navigation et d'inférence** dans le réseau,
- soit d'un **langage limité à l'accès** dans le réseau + autre **programme assurant les inférences**

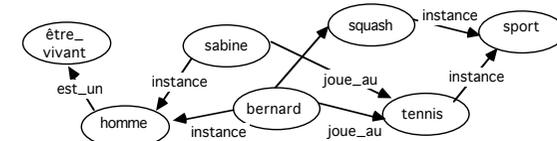
Exemple interprétation d'un RS par règles de production (Snark) :

- moteur d'inférences à règles de production à variables
- logique d'ordre 0,1,2
- faits = triplets (O1, R, O2)

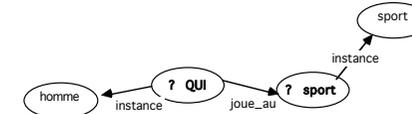


Interprétation des connaissances dans les RS

On dispose d'une base de connaissances organisée en réseau sémantique :



- la question « **quelqu'un fait-il du sport ?** » représentée par le fragment de réseau suivant :



- les réponses seront :

sabine joue_au tennis
bernard joue_au tennis
bernard joue_au squash

si réseau important → problèmes combinatoires

Intérêts des réseaux sémantiques (1)

Axes organisationnels qu'ils offrent pour structurer une base de connaissances :

• La classification, ensembles/sur-ensembles

Un objet peut être associé avec son ou ses types génériques,

Ex : *Titi peut être associé à oiseau, animal.*

- conduit à la distinction fondamentale de **type** (canari) et d'**occurrence** (Titi).
- peut être **récursive** -> définir des **méta-types** ayant pour instance d'autres types.

• Agrégation

- rattacher à un objet des propriétés ou d'autres objets y intervenant comme parties.

Ex : *Titi, vu comme objet physique possède des ailes, une tête et une queue, considéré dans son environnement, il possède un nid, un territoire, un chant, une nourriture.*

- peut être appliqué **récursivement**: un composant peut être à son tour composé d'autres composants

Intérêts des réseaux sémantiques (2)

• La généralisation, la spécialisation

relie un type à un autre type plus générique,

Ex : *oiseau à animal*

- La généralisation (lien « sorte_de ») = un ordre partiel organisant 2 types dans une généralisation ou une hiérarchie.
- économie de place en mémoire (propriétés associées à des types généraux hérités par d'autres types plus spécialisés).
- généralisation plus facile de grandes bases de connaissances (bases de données)

• La partition

regroupe des objets et éléments de relations dans des partitions qui sont organisées de façon hiérarchiques;

Ex : *si une partition P1 est au-dessous d'une autre P2, toute chose visible ou présente dans P2 l'est aussi dans P1, sans pour autant l'y avoir été spécifiée.*

- principal intérêt = **quantification** [Hendrix], la représentation du **temps** et de **l'hypothétique**. [Cohen 78].

Conclusion sur les Réseaux Sémantiques

Forces des RS :

- les objectifs d'extraction de connaissance dans la base de connaissances s'expriment en **chemins de traversée sur la structure même de la base**
- possèdent des **principes d'organisation relativement puissants** (généralisation, partition, agrégation) permettant de **structurer** la base de connaissances
- **formalisme graphique** : bonne **compréhension**, intéressant à un **premier stade de formalisation de la connaissance**
- formalisation **déclarative** : finesse et cohérence de représentation des concepts

Faiblesses des RS :

- **manque de sémantique formelle** et de **terminologie standard**
- **interprétation difficile** des connaissances : toujours un compromis à faire entre *complexité d'une structure de données* et *complexité de l'interpréteur*
- **critique** si taille du réseau importante (nb de nœuds et liens) → **explosion combinatoire**