

INGÉNIERIE DES CONNAISSANCES

INTRODUCTION (pages choisies par A Mille)

Nathalie Aussenac-Gilles (IRIT)
Jean Charlet (DSI/AP-HP)

Références :

Second Generation Expert Systems.

Éd. : J.-M. David, J.-P. Krivine et R. Simmons. Springer Verlag. 1993

KADS, a principled approach to knowledge based system development

Schreiber G. et al. Academic Press. 1993

Introduction to knowledge systems

M. Stefik. Morgan Kaufman. 1995

Acquisition et ingénierie des connaissances : tendances actuelles

Éd. : N. Aussenac-Gilles, P. Laublet et C. Reynaud. Cépaduès Éditions, Toulouse, 1996

Ingénierie des connaissances : évolutions récentes et nouveaux défis

Éd. : J. Charlet, M. Zacklad, G. Kassel et D. Bourigault. Eyrolles, Paris, 2000

INGÉNIERIE DES CONNAISSANCES

- Finalité applicative
 - aider un opérateur dans sa tâche (ou un groupe)
 - transmettre, enregistrer des savoir-faire, gérer des connaissances
 - améliorer les performances, l'adaptabilité ou les résultats de systèmes artificiels
- Interdisciplinarité
 - informatique - systémique
 - psychologie cognitive et ergonomie
 - linguistique
- Approche symbolique

Plan du cours

- 1 - L'ingénierie des connaissances dans l'IA
 - Les projets de l'IA (rappels)
 - Place de l'Ingénierie des connaissances
- 2 - À l'origine, les systèmes experts
- 3 - Deux exemples d'applications :
 - SADE : aide à la résolution de problème
 - MENELAS : reconnaissance du langage naturel
- 4 - L'Ingénierie des connaissances aujourd'hui
 - enjeux de la modélisation
 - tentative de définition
- 5 - Approfondissements

Arguments fondamentaux de l'IA

- Produire des modèles de l'action cérébrale
 - Reproduire capacités et comportements du cerveau sans support physique
 - Définir ce qu'est l'intelligence
 - Approche formelle** : Via la logique, les mathématiques, des métriques ... le comportement des machines et non leurs propriétés physiques
 - Approche cognitive et sociale** : Via l'étude des comportements intelligents individuels et collectifs
- « Calculer », c'est manipuler et traiter des symboles physiques
- ligne de partage entre
- informatique : traitement de l'information numérique
 - et IA : traitement logique de symboles

Évolution des recherches en IA

- **Produire des techniques de programmation efficaces**
Réseaux de neurones, programmation et algorithmes génétiques, Apprentissage numérique et symbolique
Formalisation logique de différents types de raisonnements (modal, causal, incertain, spatial, temporel, ...)
- **Augmenter l'intelligence du couple homme-ordinateur**
Résolution de contraintes, déduction, simulation qualitative
Multi-agents, émergence sociale de l'intelligence
Représentation des connaissances, compréhension du langage naturel, Ingénierie des connaissances, ...

Statut des expressions symboliques :

- manipulées par l'informatique en fonction de leur syntaxe
- interprétées par l'humain en tant que connaissances
 - ↳ convergence Informatique / IA

Des applications de l'IA

- **Résolution de problèmes**
 - difficiles et spécifiés, sans solution algorithmique
 - ↳ problèmes NP-complets, jeux, etc
 - mal spécifiés ou dont les connaissances ne sont pas accessibles
- **Types d'applications**
reconnaissance des formes, de la parole, synthèse de la parole et d'images, systèmes experts et résolution automatique de problèmes, réalité virtuelle, interprétation de textes, compréhension et génération du langage naturel, ...
- **Types de problèmes résolus**
Diagnostic, simulation, ordonnancement, planification, fusion de données, aide à la décision, aide à la conception, à la maintenance, etc.

Qu'est-ce que l'Ingénierie des connaissances ?

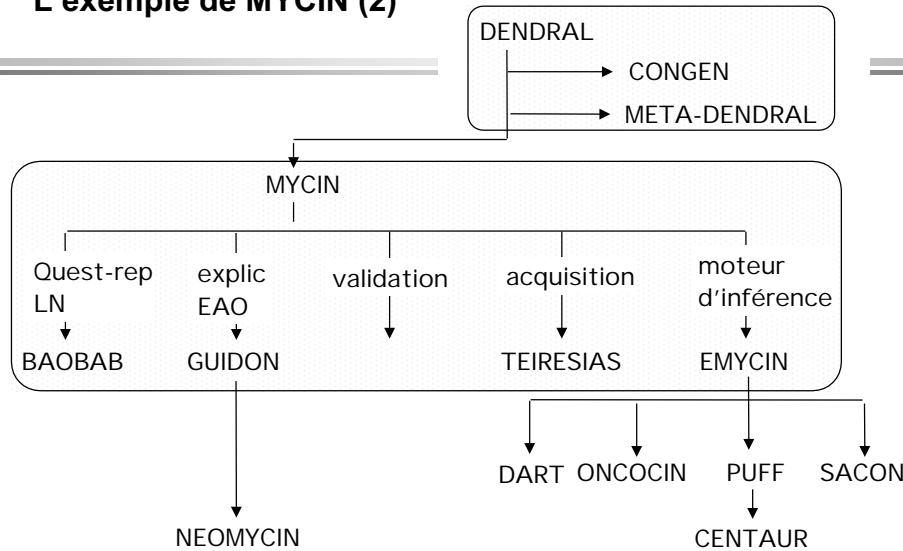
- **L'Ingénierie des connaissances intervient pour**
 - Définir une aide à l'utilisateur (méthodes, outils logiciels ou non, organisation du travail)
 - Modéliser des connaissances, individuelles ou collectives, explicites ou implicites, stabilisées ou évolutives, expertes ou techniques ...
 - Rendre ces connaissances accessibles sous une forme définie en fonction du contexte, opérationnelle ou non
- **La recherche en Ingénierie des connaissances produit**
 - Des méthodes et des techniques de recueil, d'analyse et de structuration des connaissances
 - Des plates-formes de modélisation
 - Des représentations des connaissances opérationnelles ou non

L'Ingénierie des connaissances se veut une ingénierie

L'exemple de MYCIN (1)

- **La tâche : aide à la détermination de la meilleure thérapie lors d'une infection bactérienne - 4 étapes :**
 - déterminer l'importance de l'infection
 - déterminer l'organisme responsable
 - Identifier des médicaments potentiellement utilisables
 - choisir le meilleur traitement pour le cas considéré
- **Le domaine : aucun système informatique n'a été développé**
Mycin ne contient que de l'expertise, pas d'algorithme
- **Taille : 200 règles sur les maladies infectieuses du sang**
300 règles sur la méningite
Formalisme : règles de production avec coefficients de certitude
2 modules : base de règles et interprète des règles
 - ↳ notion de moteur : EMYCIN - Chaînage arrière

L'exemple de MYCIN (2)



Principes généraux

- **Déclarativité : Séparation**
 - connaissances / mécanismes d'exploitation (contrôle)
 - BC (faits + règles) / moteur
 - connaissances déclaratives / connaissances procédurales
- **Hautes capacités de performance, du niveau de l'expert**
- **Raisonnement symbolique**
- **Souplesse et modularité :**
 - ↳ Conception aisée
 - ↳ Adaptation à l'utilisateur - Evolutivité
- **Compréhensibilité :**
 - ↳ Capacité à expliquer le raisonnement produit (via des traces) pourquoi ? Comment ? pourquoi pas ? justification de résultats intermédiaires
 - ↳ Représentation lisible par une personne du domaine

Les règles de production

SI condition **ALORS** conclusion (ou action)

SI le spectre possède 2 pics aux masses m_1 et m_2 tels que

- $m_1 + m_2 = m + 28$
- et $m_1 - 28$ est un pic élevé
- et $m_2 - 28$ est un pic élevé
- et au moins m_1 ou m_2 est élevé

ALORS la molécule contient un groupe Céton

- Quand les conclusions sont des faits, système dit analytique
- Quand les conclusions sont des actions, système dit synthétique

Les performances de MYCIN

- **Pas d'utilisation opérationnelle**
- **Comment valider un SBC ? Quel critère ?**
 - par rapport à l'expert
 - savoir traiter 90 % des pbs ou aussi les pbs graves
 - taux de réussite absolu → résultats
 - évaluation du raisonnement suivi et des résultats
- **MYCIN : technique utilisée pour valider une nouvelle thérapie**
 - 10 cas - 8 experts + Mycin + solution du cas appliquée = 10 prescripteurs
 - 1) Chaque expert évalue toutes les prescriptions (note /80)
 - 2) Le collège évalue toutes les prescriptions (note /8)
 - 3) Le collège signale les solutions inadmissibles
- **MYCIN : meilleurs résultats sur les 10 prescripteurs**

Les limites : comment décrire MYCIN

- Description trop générale, ne dit pas comment le système résout réellement
MYCIN : diagnostic des maladies du sang
- en terme d'implémentation, granularité trop fine
MYCIN : règles, chaînage arrière, coefficients de vraisemblance...
- mêmes primitives, ne fait apparaître ni les caractéristiques des tâches, ni comment elles sont réalisées
- mélange plusieurs sortes de connaissances : celles de contrôle et celles du domaine

Une règle de MYCIN

SI
la coloration de l'organisme est GRAM+
et si la morphologie de l'organisme est Cocci
et si le mode de développement de l'organisme est en colonies
ALORS
il existe une évidence (0.7) que l'identité de l'organisme soit Staphylococcus.

- Pour écrire une telle règle, le concepteur devra savoir comment MYCIN fonctionne. Ici, il faut savoir que cette règle, parce qu'elle conclut sur l'identité de Staphylococcus et parce que le moteur fonctionne en chaînage arrière, ne sera envisagée que lorsque le système se préoccupera de déterminer l'identité d'un organisme. Il faut de plus savoir qu'à chaque instant, le moteur s'intéresse à un « context » particulier, c'est à dire un organisme à retrouver. Un ordre des prémisses dans une autre règle fera en sorte que MYCIN s'intéressera d'abord à l'organisme courant avant de rechercher des organismes présents antérieurement. À chaque type d'organisme est associé un ensemble de caractéristiques (par exemple, identité de l'organisme pour l'organisme courant). À chaque caractéristique est associé un attribut « mis-à-jour-par » qui contient la liste des règles qui permettent d'établir la caractéristique en question (l'identité de l'organisme, par exemple). La règle R1 ci-dessus fait partie de cette liste de règles...

Dépasser les limites

- Une stratégie de résolution implicite
- Des capacités auto-explicatives assez pauvres
- De grandes difficultés pour l'acquisition des connaissances
- Une maintenance de la BC assez difficile
 - ↳ Mieux modéliser le raisonnement
 - ↳ « Knowledge Level »
- Une dégradation brusque des compétences aux bornes du domaine
- Des connaissances de surface
- Absence de généralité des systèmes construits
- Mieux modéliser le domaine
 - ↳ Raisonnement basé sur un modèle

Mieux modéliser le raisonnement (1)

- Exprimer les connaissances stratégiques
- MYCIN → Neo MYCIN : méta-règles explicitant la stratégie
ex : SI une observation doit toujours être expliquée
ET elle n'est pas expliquée pour le « différentiel »
ALORS Evoquer des hypothèses qui peuvent l'expliquer
- Chaque méta-règle décrit une sous-tâche :
Etablir le diagnostic →
 - 1) identifier le pb (→ info. initiales et plaintes somatiques)
 - 2) établir espace des hyp (→ établir différentiel, grouper et différencier, explorer et raffiner)
 - 3) effectuer les analyses

Mieux modéliser le raisonnement (2)

- **Connaissances stratégiques** : expriment comment une tâche va être effectuée. Utilisent des connaissances du domaine
- **Connaissances du domaine** : concepts manipulés et leurs relations. « Ce qui est vrai » dans le domaine pour la tâche considérée
- **Modèle conceptuel** : modèle explicitant la manière dont le système résout ainsi que les connaissances qu'il utilise

MC = Connaissances stratégiques + connaissances du domaine

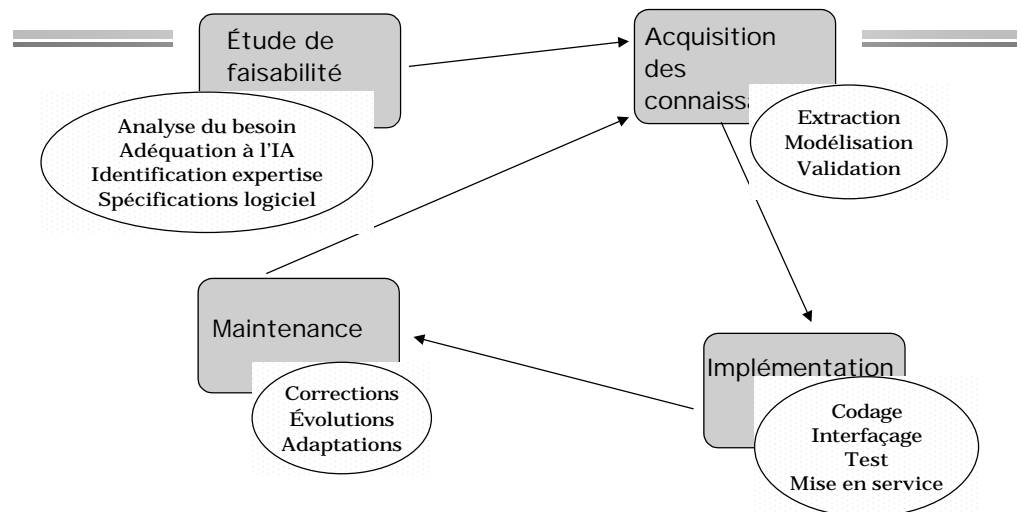
Le « *knowledge level* » de Newell - 1982 -

- **Un niveau au dessus du niveau symbolique**
But = décrire un système informatique
différencier connaissances et représentations
description du *comportement observé* du système à l'aide de connaissances, indépendamment de leur formalisation
- **Agent rationnel**
dispose de *connaissances*, doit atteindre des *buts*, sait effectuer des *actions*, et est *rationnel* : choisit (avec ses connaissances) l'action suivante qui va le mener le plus directement au but
- **Application à la construction de SBC**
système = agent rationnel interagissant avec son environnement
descriptions plus structurées que ce que propose Newell.

Le raisonnement basé sur les modèles

- **Description de systèmes physiques**
Modèles fonctionnels : description des fonctionnalités (bon vs mauvais fonctionnement)
Modèles topologiques : description des composants
Modèles causaux : relations entre états ou événements
- **Physique qualitative**
calcul sur des variables discrètes
s'inspire de la physique quantitative
- **Modélisation du raisonnement**
méthode de parcours du modèle
explicite les stratégies

Le cycle de vie



Qu'est-ce qu'un modèle ?

● Rôles d'un modèle conceptuel :

- ↳ *conceptualiser entités du domaine et interactions*
- ↳ *intermédiaire entre connaissances et code : spécification ?*
 - Langage partagé par l'expert, le cognicien
 - Langage d'expression des connaissances
 - Langage « compréhensible » par l'artefact

Un modèle est une abstraction qui permet de réduire la complexité en se focalisant sur certains aspects, en fonction de certains buts

MAIS un modèle devrait permettre plus : manipuler les objets et interpréter les résultats de la manipulation

Le modèle d'expertise : Qu'est ce qui est réellement modélisé ?

Jusque vers 1990 : KADS I, ...

modèle du monde réel en tant que modèle du processus de résolution de problème tel qu'il est *observé* à l'aide des experts du domaine ou d'autres sources

Après 1990 : KADS II, ...

spécification fonctionnelle de la partie résolution de problème de l'*artefact* à *construire*

On passe de l'observation à l'interprétation

Concevoir des SBC ?

Trois difficultés :

- **Élicitation** : déterminer quelles expressions linguistiques permettent d'expliciter le problème à traiter ;
 - **Modélisation « au niveau des connaissances »** : modèle conceptuel compréhensible et interprétable par un spécialiste ;
 - **Opérationnalisation au niveau des programmes** : modèle opérationnel interprétable par la machine.
- ↳ Revenir aux problèmes fondamentaux

La connaissance (1)

1. Connaissance dans un environnement technique

La connaissance n'est possible que dans un environnement *technique* qui est le nôtre, qu'il apparaisse fortement technique (industrie, monde professionnel), ou non (monde quotidien,...)

2. Connaissance=interprétation humaine

Il y a connaissance et représentation des connaissances quand les *manipulations symboliques* effectuées par la machine via des programmes, prennent un sens et une *justification pour les utilisateurs* interagissant avec ces programmes

3. Connaissance = f (support=outil informatique)

La pensée repose sur la *médiation externe du signe*. La technique (ici l'informatique), par ses outils et ses capacités de mémorisation, permet alors à l'homme de se constituer des connaissances qui évoluent et s'accumulent et n'existeraient pas sans cela

La connaissance (2)

4. Connaissances → action

Il y a présomption de connaissances, si la faculté d'utiliser l'information à bon escient est attestée. La connaissance est ainsi liée à l'*action*. Cette action est ici une réécriture (ontologie dépendant de la tâche, savoir-lire ⇒ savoir-écrire)

5. Connaissance=f (contexte)

Il y a connaissance quand il y a *contexte* d'utilisation de l'information

6. Connaissance=f (tâche)

La connaissance *a priori* n'existe pas : elle est construite à partir d'un projet propre au modélisateur. Étudier cette construction permet d'essayer d'en dégager des invariants méthodologiques

SBC (1) : Manipulation formelle et interprétation sémiotique

Deux constats :

Les SBC ne manipulent pas des *représentations* de connaissances, mais des *expressions symboliques* ou *inscriptions*

Les règles formelles de manipulation ne sont pas les règles de l'interprétation : le calcul formel n'est pas une explicitation du sens (interprétation) mais une réécriture symbolique

SBC (2) : Ce que sont les SBC

Un principe : la manipulation symbolique automatique construit des inscriptions dont l'interprétation permet à l'utilisateur d'effectuer une tâche donnée (2)

Définition 1 :

systèmes sémiotiques de manipulation d'inscriptions symboliques dont le fonctionnement informatique doit permettre à l'utilisateur d'interpréter et de comprendre le système dans le cadre de son activité, en utilisant les termes du domaine

Le point de vue des modèles formels vs des usages

Constat : les SBC sont des *systèmes techniques plongés dans des systèmes d'usages* (1-3)

Définition 2 :

L'Ingénierie des connaissances correspond à l'étude de modèles symboliques formels plongés dans des systèmes d'usage : c'est l'ingénierie informatique et logique des modèles en fonction des usages qu'ils rendent possibles et des appropriations qu'ils permettent

En guise de programme

Mobiliser les points de vue précédents pour se donner un programme de travail

L'objet de la recherche sont les systèmes dont le fonctionnement est interprétable par l'humain en termes de connaissances (2)

Le but correspond alors à l'intégration de ces réflexions dans des artefact informatiques à modéliser, construire, faire évoluer, ... (1)

En n'oubliant pas que l'artefact informatique modifie les usages singuliers comme collectifs, jusqu'aux organisations (3)

En guise de méthode

Mobiliser les points de vue précédents pour se donner des méthodologies de travail

La connaissance étant contextuelle, se donner les moyens de représenter ce contexte dans une démarche de représentation des connaissances (5)

La connaissance étant liée à l'action, veiller à l'usage qui est fait des représentations ; en particulier qu'elles sont bien mémorisées dans un but, pour un usage prévu (4)

La connaissance, contextuelle, ne pouvant être indépendante de la tâche à réaliser, se donner les moyens d'assumer explicitement cette dépendance (réutilisation limitée) (6)

En guise de champ d'action

L'ingénierie des connaissances propose des concepts, méthodes et techniques permettant de modéliser et/ou d'acquérir les connaissances dans des domaines se formalisant, a priori, peu ou pas (concepts d'usage versus concepts techniques ou formels).

- La finalité reste celle de l'Intelligence Artificielle : construire des systèmes informatiques.
- La caractéristique, « à base de connaissances » perdure.
- La production, « système expert », se diversifie et s'étend à tout système informatique mettant explicitement en jeu (par des modèles) des connaissances.
- Des applications, collaborations, synergies, etc. avec tout domaine où une modélisation explicite des connaissances peut être envisagée

Modélisation (1)

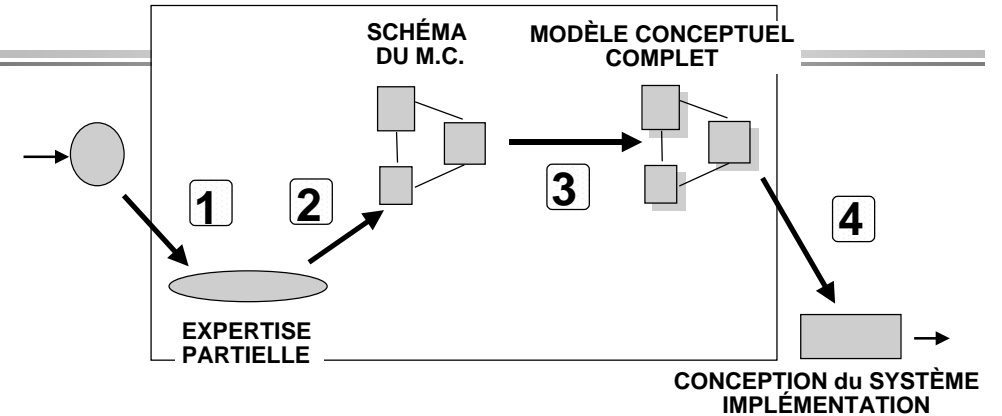
- Acquisition de la connaissance et compréhension des méthodes de résolution à utiliser
Associer recueil et analyse, pour obtenir un *modèle spécifique*
Le modèle d'expertise est utile comme étape intermédiaire dans le développement de SBC
- Modèle indépendant des structures de représentation et des algorithmes dans le SBC
Utilisation d'un langage spécifique non-orienté implémentation, semi-formel (complété ensuite par un langage formel)
- L'acquisition des connaissances (élicitation et analyse) peut être guidée par la réutilisation de modèles ou de composants génériques

Différentes approches

- par extraction d'expertise
- par construction de modèles
- conduite par les modèles

chaque approche enrichit, réutilise et réinterprète la précédente

L'acquisition conduite par les modèles



- 1** Acquisition dirigée par les données **3** Instantiation du schéma du M.C.
2 Construction du schéma du M.C. **4** Opérationnalisation du modèle

Aides à l'extraction des connaissances

- Difficultés :
 - faire expliciter, capturer des connaissances explicites
 - interpréter sans introduire de biais
- Travaux de recherche :
 - en psychologie, ergonomie et IA
 - établir des liens entre types de tâches et techniques d'extraction
- Ce qui existe :
 - des techniques de recueil
 - des outils automatisant des techniques : ETS, AQUINAS, PC_PACK (Univ. de Nottingham) en amont de CommonKADS
 - des méthodes : MACAO, KOD

Principes de recueil des données

Démarche ascendante : Bottom-up knowledge elicitation

Production de « données brutes »

- Méthodes informelles :
 - Résultats bruités, mal structurés... mais méthodes moins contraignantes.
 - Exemples: Interview, « think aloud »
- Méthodes formelles :
 - Moins bruités et mieux structurés. Mais plus spécifiques... Quelle adéquation ?
 - Exemple : Grilles répertoires (ETS)

Techniques de recueil : caractéristiques

- **Place par rapport à l'activité** : préalable / concomitante / consécutive
- **Mode de communication** : direct, téléphone, courrier
- **Analyse des réponses** : qualitative ou quantitative
- **Types de connaissances recueillies** :
 - automatismes / habiletés : observations, tech. indirectes, expérimentations
 - règles : simulations dans le contexte, verbalisations
 - schémas, connaissances déclaratives: entretiens centrés
 - plans et stratégies : entretiens, simulations avec variations d'énoncés
- **Origine des données verbales** : spontanée (activité) / provoquée (questionnaires)
- **Situation du recueil** : conditions habituelles / transformées : simulations, entretiens ; en groupe / individuel
- **Documentation et compte rendu** : notes, retranscriptions
- **Mode d'accès aux connaissances** : direct (observations) / indirect (indicateurs)

Principes de recueil des données

● Analyse des données recueillies

- Enlever le « bruit », ce qui est « inutile »
- Identifier et extraire des fragments (concepts, relations, définitions)

Plus ou moins facile selon ce qui a été fait avant

Exemples : KOD (Vogel) , COGNOSYS (Woodward)

● Choix des techniques

- selon les psychologues : disposer d'une théorie de l'activité étudiée
- au moins : prendre en compte la diversité
 - des formes de connaissances
 - des niveaux d'abstraction et de régulation
 - des individus

Conclusion

- **Utilisation des techniques sans idée a priori du modèle**
 - **Caractériser** en vue de définir un modèle
 - Le modèle comme grille de choix des techniques
- **Problèmes**
 - choix d'une technique adaptée
 - coût, disponibilité de l'expert
 - interprétation
- **Techniques à retenir**
 - Observations, analyse du travail
 - Simulations + verbalisations descriptives + explications consécutives
 - Reformulations, entretiens centrés
 - Travaux à partir de corpus**

Le modèle d'expertise : Qu'est ce qui est réellement modélisé ?

Jusque vers 1990 : KADS I, ...

modèle du monde réel en tant que modèle du processus de résolution de problème tel qu'il est *observé* à l'aide des experts du domaine ou d'autres sources

Après 1990 : KADS II, ...

spécification fonctionnelle de la partie résolution de problème de l'*artefact* à *construire*

On passe de l'observation à l'interprétation

Qu'est-ce qu'un modèle ?

● Rôles d'un modèle conceptuel :

- ☞ *conceptualiser entités du domaine et interactions*
- ☞ *intermédiaire entre connaissances et code : spécification ?*
 - Langage partagé par l'expert, le cognicien
 - Langage d'expression des connaissances
 - Langage « compréhensible » par l'artefact

Un modèle est une abstraction qui permet de réduire la complexité en se focalisant sur certains aspects, en fonction de certains buts

MAIS un modèle devrait permettre plus : manipuler les objets et interpréter les résultats de la manipulation

Le modèle conceptuel

- Connaissances de raisonnement au sens large : **expriment comment une tâche va être effectuée. Utilisent des connaissances du domaine**
- Connaissances du domaine : **concepts manipulés et leurs relations. « Ce qui est vrai » dans le domaine pour la tâche considérée**
- Modèle conceptuel : **modèle explicitant la manière dont le système résout ainsi que les connaissances qu'il utilise**

Modèle conceptuel

=

Connaissances de raisonnement + connaissances du domaine

Le « *knowledge level* » de Newell - 1982 -

● Nécessité d'un niveau au dessus du niveau symbolique

But = décrire un système informatique
différencier connaissances et représentations
description du *comportement observé* du système à l'aide de connaissances, indépendamment de leur formalisation

● Agent rationnel

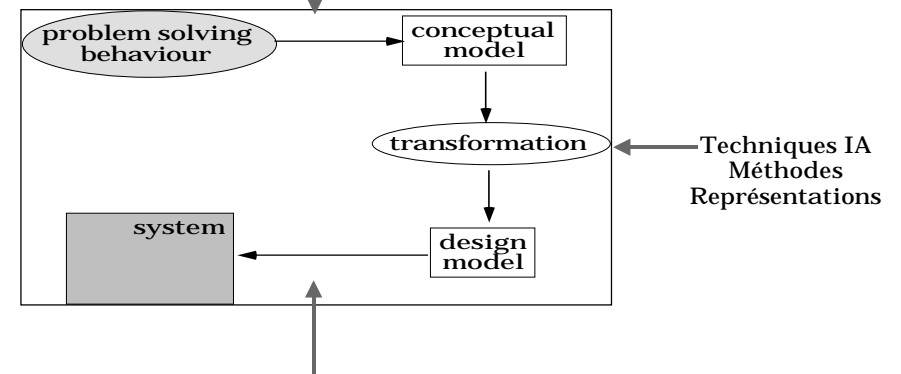
dispose de *connaissances*, doit atteindre des *buts*, sait effectuer des *actions*, et est *rationnel* : choisit (avec ses connaissances) l'action suivante qui va le mener le plus directement au but

● Application à la construction de SBC

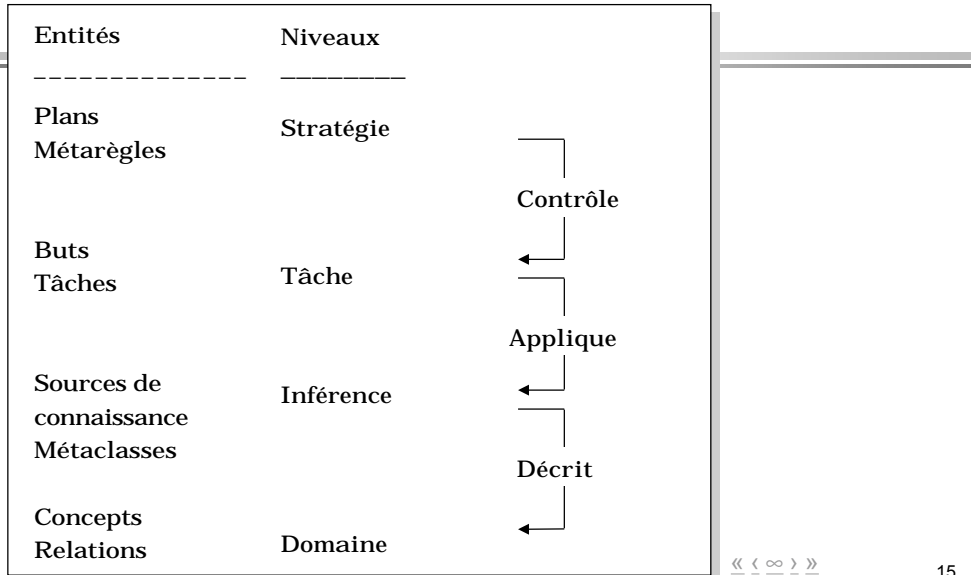
système = agent rationnel interagissant avec son environnement
descriptions plus structurées que ce que propose Newell.

Cadre d'interprétation:
vocabulaire, composants génériques
(modèle d'interprétation)

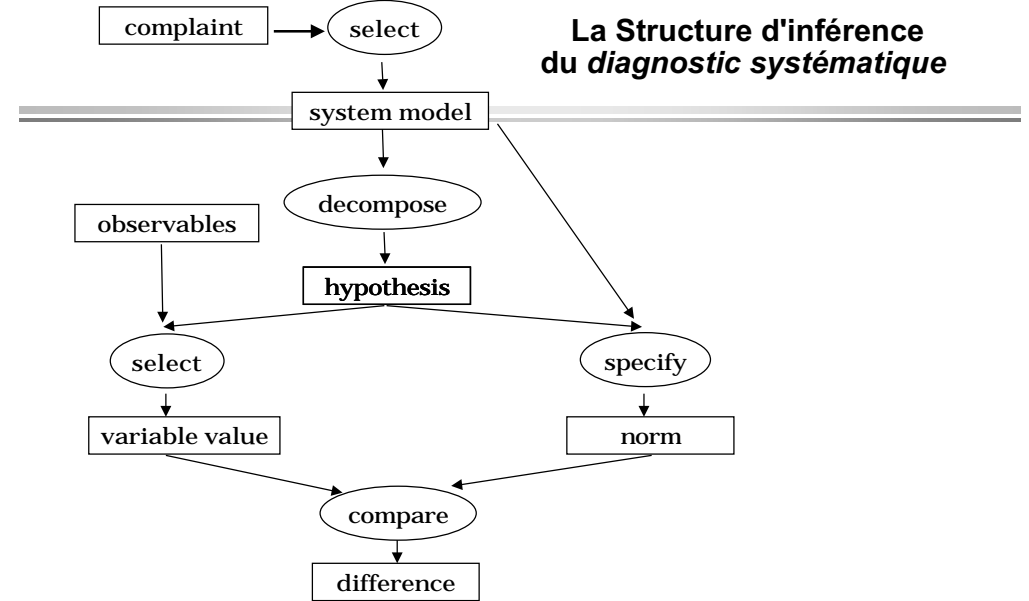
Les modèles de KADS II



KADS : Les quatre couches du modèle d'expertise



La Structure d'inférence du diagnostic systématique



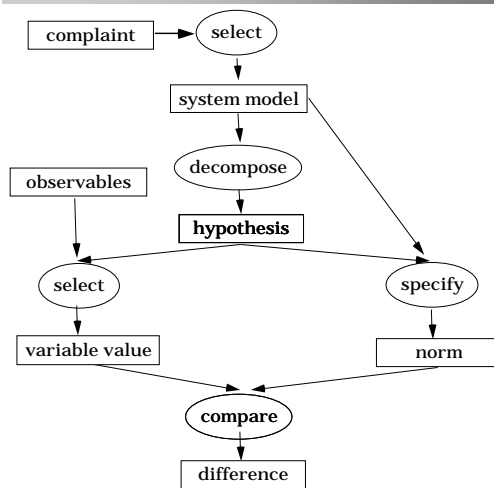
Task Systematic-Diagnosis

Goal
find the smallest component with inconsistent behaviour if one

Control terms
differential = set of currently active hypothesis
inconsistent-sub-system= subpart

Task-structure
systematic-diagnosis
(complaint → inconsistent-sub-system) =
select
(complaint → system-model)
generate-hypotheses
(system-model → differential)
REPEAT
test-hypotheses
(differential → inconsistent-sub-system)
generate-hypotheses
(inconsistent-sub-system → differential)
UNTIL differential is empty

La Structure de tâche du diagnostic systématique (1)



Task Generate-Hypothesis

Goal
generate new set of hypothesis through decomposition

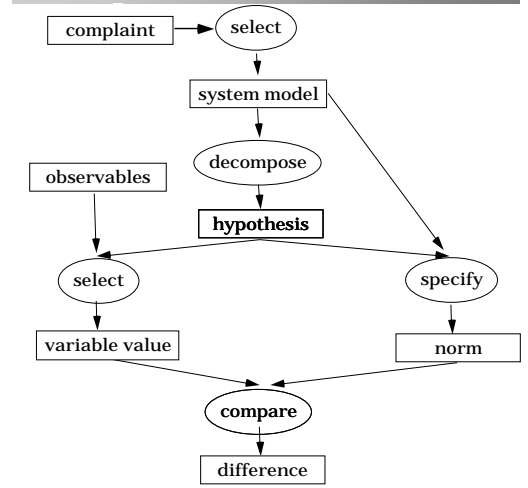
Task-structure
generate (system-model → differential) =
decompose (system-model → differential)

Task Test-Hypothesis

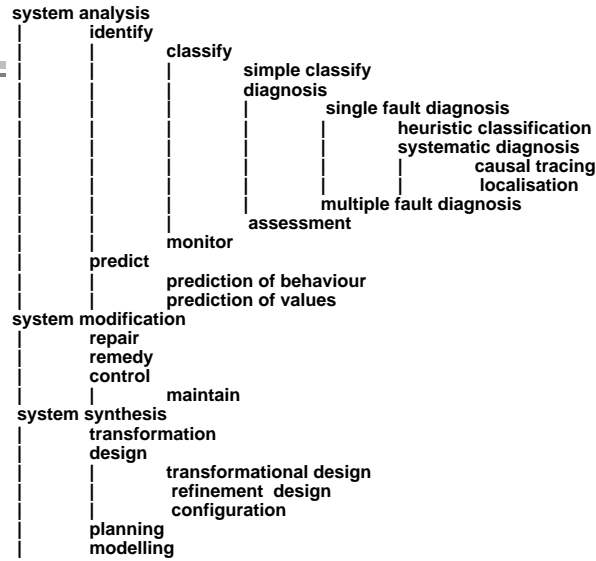
Goal
test whether a hypothesis in the differential behaves inconsistently

Task-structure
test (differential → hypothesis) =
DO FOR EACH hypothesis in differential
specify (hypothesis → norm)
select (hypothesis → observable)
OBTAIN (observable → variable value)
compare (norm + variable value → difference)
UNTIL difference = true

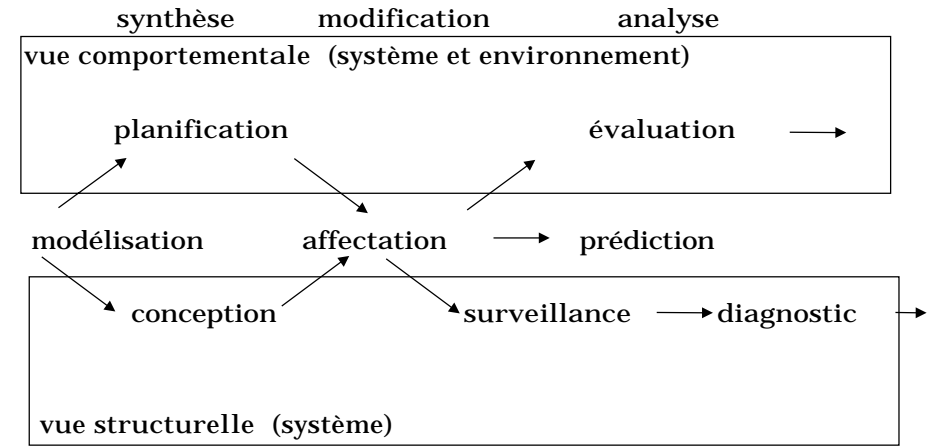
La Structure de tâche du diagnostic systématique (2)



KADS-1 : La bibliothèque des modèles d'interprétation



KADS-2 : imbrication des types de problèmes

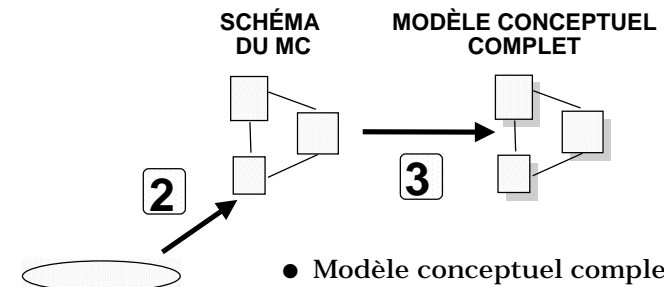


Bibliothèques de composants réutilisables

- **Tâches : modèles et guides de modélisation**
Modèles papier et questions d'aide à la modélisation (livre CommonKADS – Chap. 6)
Modèles CML téléchargeables : WebCOKACE
<<http://www.inria.fr/acacia/Cokace>>
- **Méthodes de résolution de problèmes**
Inventaire de méthodes de résolution, d'opérateurs et de rôles applicables dans une méthode (livre CommonKADS – Chap. 13)
Modèles de méthodes en CML
<<http://www.swi.uva.nl/projects/CommonKADS/home.html>>

L'approche conduite par un modèle

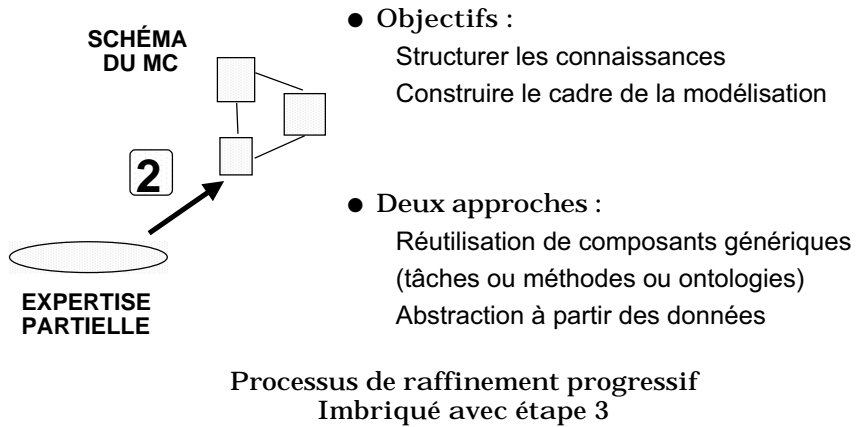
- **Schéma du modèle conceptuel**
Modèle de raisonnement
+ Concepts du domaine



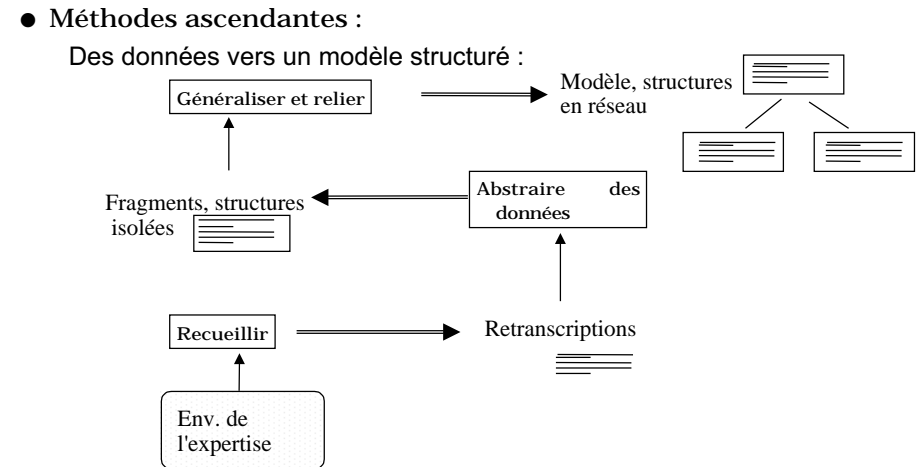
- **Modèle conceptuel complet**

Instantiation du schéma (instances de concepts, inventaire de toutes les règles, etc.)

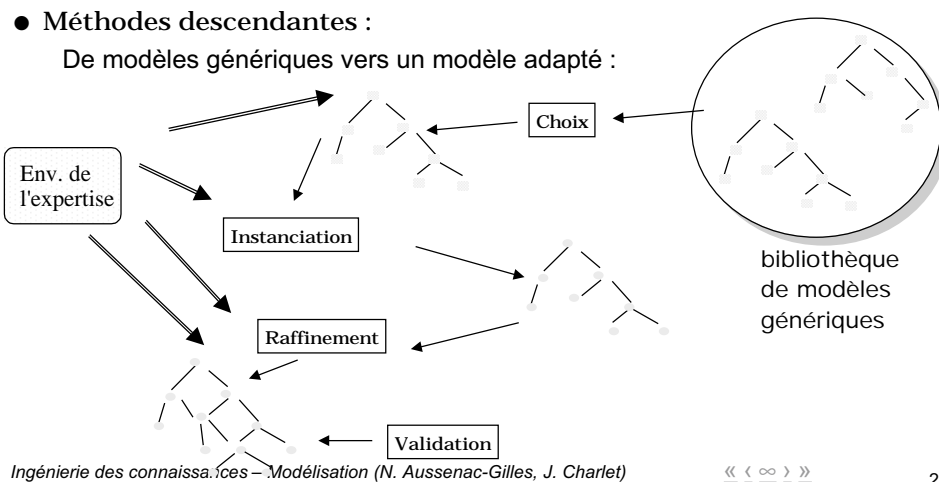
Construction du schéma du MC



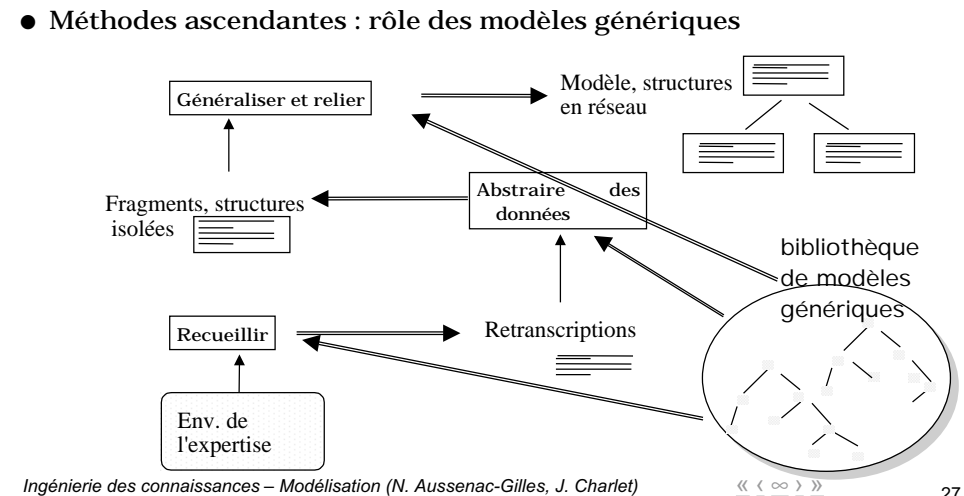
Construction ascendante versus descendante (1)



Construction ascendante versus descendante (2)



Construction ascendante versus descendante (3)



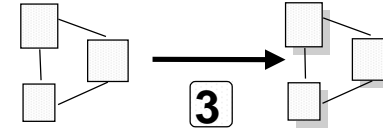
Ascendant versus descendant : deux points de vue, deux cycles de base

Recueillir ... interroger, analyser, retranscrire
Conceptualiser ... abstraire, typer
Formaliser ... représenter
Structurer ... organiser
Valider

Caractériser ... typer, qualifier
 Choisir ... comparer, réutiliser
 Adapter ... modifier, corriger
 Affiner ... préciser, compléter, spécifier
 Instancier ... énumérer, lister
 Valider

Instanciation du modèle conceptuel (1)

SCHÉMA DU MC → **MODÈLE CONCEPTUEL COMPLET**



● **Objectifs**

Instancier le modèle conceptuel
 Affiner le schéma du modèle
 Être exhaustif

● **Méthode**

Identifier tous les paramètres
 Expliciter les contrôles
 Définir les domaines de valeur
 Préciser règles, méthodes et procédures

Processus de raffinement progressif
Imbriqué avec étape 2

Instanciation du modèle conceptuel (2)

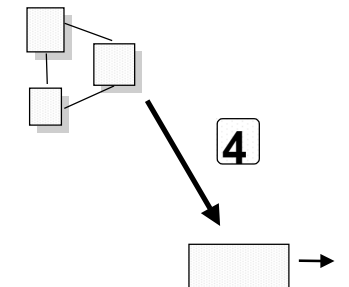
- **On sait ce qu'il faut acquérir.**
 Le modèle conceptuel spécifie précisément ce qu'il faut fournir au système. L'acquisition des connaissances est guidée par le modèle que l'on cherche à remplir. Chaque tâche, chaque méthode de résolution de problème nécessite des connaissances précises, pour un rôle bien spécifié.
- **Définition précise de la fonction des connaissances demandées.**
 Une bonne définition et une bonne compréhension de la fonction (du rôle) des connaissances demandées en facilite l'expression par l'expert.
- **On dispose d'une sorte de métrique.**
 On peut identifier facilement les endroits où il manque des connaissances.
- **Le système peut mieux « savoir » les limites de ce qu'« il sait ».**

Opérationnalisation du Modèle conceptuel (1)

● **Objectifs**

Traduire le MC en un modèle opérationnel
 Valider le modèle par simulation
 Développer la base de connaissances

MODÈLE CONCEPTUEL COMPLET



● **Contraintes**

→ correspondance BC ↔ MC
 → conserver les liens entre structures
 → anticiper les problèmes de maintenance

Opérationnalisation : motivations

- **Motivations : Acquisition/validation et mise à jour**
 - Conception incrémentale d'un modèle conceptuel
 - Conception incrémentale d'une base de connaissances (cycle acquisition, test, validation)
 - Mise à jour de la base de connaissances (les connaissances évoluent)
- « Debugging » et maintenance :
 - Comportement observé dans le système opérationnel
 - Compréhension / interprétation en termes du modèle conceptuel
- **Explications**
 - en terme de connaissances

Opérationnalisation : solution

- **Créer des langages spécifiques pour représenter le modèle conceptuel**
 - Support pour l'acquisition des connaissances (MC, BC, ...)
- **Respecter le principe de correspondance structurelle entre modèle conceptuel et langage d'implémentation**
 - Possibilité de tracer le raisonnement en termes de buts et méthodes
 - Réflexivité permettant la mise en place d'un contrôle de bon niveau (ZOLA, DEF-*) ou à l'utilisateur d'interagir (LISA, ...)
- ...

Opérationnalisation : différentes approches (1)

- **Simuler/valider le modèle conceptuel (MODEL-K, ZOLA, OCML, OMOS,...)**
 - Possibilité de tracer le raisonnement en termes de *buts* et *méthodes*, etc.
 - Conception incrémentale d'un modèle conceptuel
 - Possibilité de validation fonctionnelle de partie du modèle (*tâche*, *méthode* – SACHEM)

Mais

- Représentation que de certains modèles (RLM, KADS – *tâche*, *inférence* [OMOS, MODEL-K] ; *tâche*, objet, classe d'objets, règles [AIDE])
- Perte de propriétés au moment de l'implémentation définitive

Opérationnalisation : différentes approches (2)

- **Formalisation du modèle conceptuel (KARL, TASK, FORKADS, (ML)²...)**
 - Possibilité de vérifications formelles de complétudes, de cohérences...

Versus

- **Modèle conceptuel non formel (MONA, CML)**

Opérationnalisation : différentes approches (3)

- Coder directement le modèle opérationnel (LISA, KARL, DEF-*,)
 - Possibilité de tracer en permanence le raisonnement en termes de *buts* et *méthodes*
 - Réflexivité permettant la mise en place d'un contrôle défini en termes de *buts*, *méthodes* permettant à l'utilisateur d'interagir (LISA, ...) ou la mise en place de primitives de représentation des connaissances manipulées par le langage (DEF-*)
 - Possibilités de validation en introduisant des primitives spécifiques à cette validation (AMD)

Mais

- Pas toujours un langage réellement opérationnel (développement, gestion des erreurs, vitesse, ...)
- Représentation que de certains modèles (*buts*, *méthodes* [LISA], DEF-*)

Opérationnalisation : directions

- Définir un métalangage, plus souple mais demandant plus de modélisation (respect des modèles tâche, méthode mais possibilité de programmation « ad hoc » [ZOLA])
- Développer une version « industrielle » du langage (ZOLA IN JAVA, LISA@RT, DEF-*, ...)
- ...

De toute façon

- Nécessité de faire cohabiter ≠ formes d'implémentation, de la plus déclarative à la plus procédurale

Ateliers logiciels de modélisation

- MOKA
 - Projet européen, avec PSA, Daimler, BAES, Decan-Ingénia (développe et diffuse le logiciel), ... langage de modélisation spécifique, MML, extension d'UML
- PROTEGE 2000
 - Stanford Medical Informatics, Knowledge Modelling Group. Méta-langage de modélisation ; permet de décrire toute structure de modélisation et de générer des interfaces de saisie de connaissances pour ces structures (ex : CML - CommonKADS)
 - <<http://www-smi.stanford.edu/projects/protege/protege-2000/index.html>>

19.10.2000

INGÉNIERIE DES CONNAISSANCES : ONTOLOGIES, TERMINOLOGIES, BCT (3)

Jean Charlet (DSI/AP-HP), Nathalie Aussenac-Gilles (IRIT)
Philippe Laublet (CAMS), Bruno Bachimont (INA)

Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances. B. Bachimont In: *Ingénierie des connaissances, évolutions récentes et nouveaux défis*, J. Charlet, M. Zacklad, G. Kassel, D. Bourigault édés, Eyrolles, 2000.

Pour une terminologie textuelle. D. Bourigault & M. Slodzian. In: *Terminologies nouvelles*, 19, 1999.

Ontologie et réutilisabilité : expérience et discussion. J. Charlet, B. Bachimont, J. Bouaud, P. Zweigenbaum In : *Acquisition et Ingénierie des connaissances, tendances actuelles*, N. Aussenac, P. Laublet & C. Reynaud édés, Cépaduès, 1996.

<http://www.irit.fr/GRACQ/> – <http://www.biomath.jussieu.fr/TIA/>

Historique

- Définitions
- Motivations
- Questions
- Des approches représentatives

Approches pragmatiques pour les SBC

Constats initiaux

- Coût de réalisation de bases de connaissances
- Grande variété des langages de représentation
- Hétérogénéité des ontologies : vocabulaires, catégorisation

Réutilisation

- Créer et conserver des bases de connaissances réutilisables
- Assembler des bases de connaissances à partir de modules réutilisables

Partage de la connaissance et communication

- Interopérabilité entre différents SBC
- Échange de connaissances entre systèmes

Différentes initiatives

- «Knowledge Sharing Effort» (Gruber), ONTOLINGUA, ...

Définitions

Ontologie PHILO. *Partie de la métaphysique qui s'applique à l'être en tant qu'être, indépendamment de ses déterminations particulières* (Le Petit Robert).

Taxinomie 1.DIDACT. *Étude théorique des bases, lois, règles, principes, d'une classification.* 2. *Classification d'éléments* (Le Petit Robert).

➤ Taxinomies en sciences naturelles

Ontologie INGÉNIERIE DES CONNAISSANCES. *Ensemble des objets reconnus comme existant dans le domaine. Construire une ontologie c'est aussi décider de la manière d'être et d'exister des objets.*

«Knowledge Sharing Effort»

Initiative DARPA, NSF, ...

- Neches, Fikes, Gruber, ...

Infrastructure comme support à la réutilisation de connaissance

- INTERLINGUA (KIF) comme langage pivot pour la traduction d'un langage vers un autre
- Common KR System Specifications : ex, Rock de CMU, ...
- Extension de KIF pour la représentation d'ontologies (ONTOLINGUA)
- *Serveur d'ontologies de Stanford*
- Interfaces externes d'agents (KQML : «Knowledge Query and Manipulation Language»)
- API pour les bases de données orientées objets (GFP)
- ...

Différentes ontologies (collection, agents, frame, topologie, quantités et dimensions physiques, unités de mesures, ...)

Évolution des travaux en ingénierie des connaissances

- Comme transfert d'expertise
- ➤ Par construction de modèles : RLM, KADS, ...
- ➤ ➤ Conduite par des modèles
- ☞ Aide à la construction de bases de connaissances par réutilisation de MRP
- ☞ Interaction entre tâche, MRP et le mode de conceptualisation du monde
- ☞ Aide à la construction de bases de connaissances par réutilisation d'ontologies
- ☞ Exemples : Games (médical), Kactus, (...)

Des abstractions partageables ?

- ☞ Lexiques
- ☞ Thesaurus
- ☞ Ontologies
- ☞ Modèles du domaine
- ☞ Modèles du cas
- ☞ Décomposition en tâches récurrentes
- ☞ Méthodes de résolution de problèmes
- ☞ Tâches applicatives abstraites

Des points de vues opposés ?

Première vision

- Une ontologie est universelle *mais différente* de la base de connaissance qui elle serait individuée, relative, finalisée

Deuxième vision

- Nombreuses «ontologies» pour un système du monde
- Impossibilité de délivrer des ontologies universelles incluant tous les points de vues possibles
- Porte la trace de la tâche particulière pour lesquelles elles ont été construites et du raisonnement adopté pour cette tâche

Une classification des ontologies (1)

Ontologies de représentation

- définit un ensemble de primitives de représentation
- exemple : la «frame ontology» d'ONTOLINGUA

Ontologies génériques (Dont le «haut» est réutilisable)

Ontologies du domaine

Ontologies d'une MRP

- le rôle joué par chaque concept dans une méthode particulière est rendu explicite (➤ ontologies de tâche)

Ontologies d'application

- Voir une ontologie d'application comme une double spécialisation : d'une ontologie du domaine et d'une ontologie de méthode

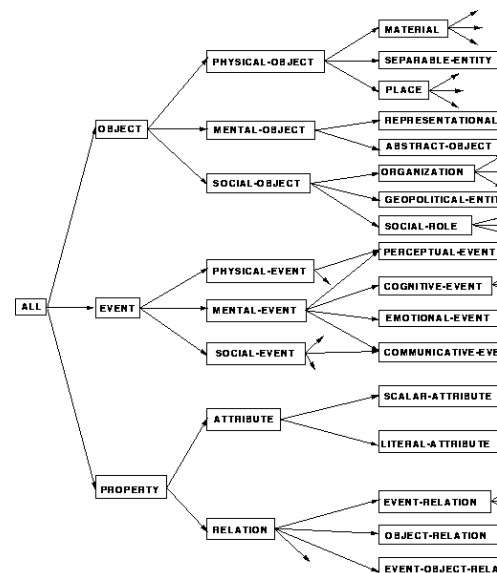
De la représentation des connaissances

La donnée d'un langage formel de représentation des connaissances laisse ouvert la question des symboles fonctionnels et relationnels nécessaires et de la sémantique à y associer.

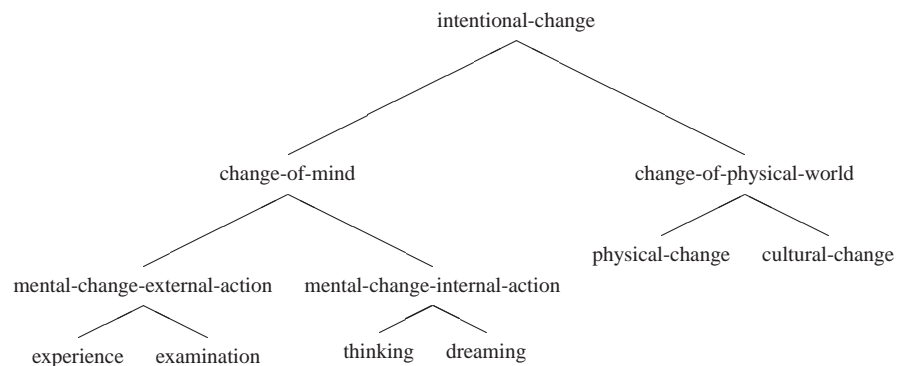
Définir une ontologie pour la représentation des connaissances, c'est définir, pour un domaine et un problème donnés, la signature fonctionnelle et relationnelle d'un langage formel de représentation et la sémantique associée. (Bachimont 2000)

Définir les primitives non logiques d'un langage de représentation et la sémantique associée : déterminer les notions élémentaires à partir desquelles les connaissances du domaine sont construites.

MIKROKOSMOS



Ontologie de MENELAS : «intentional-change»

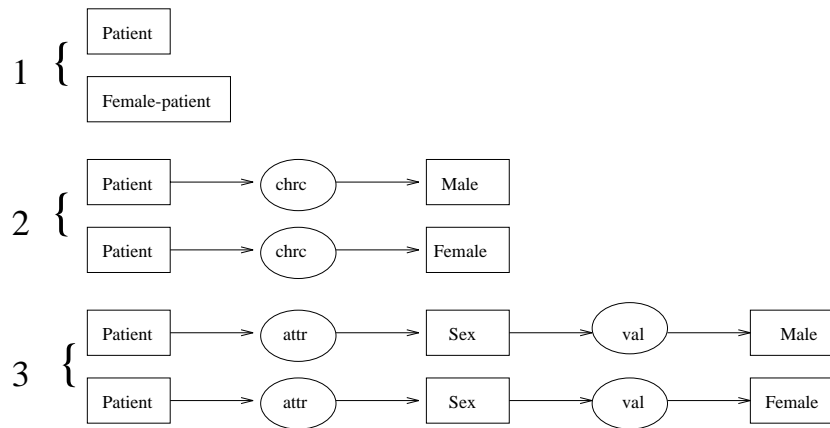


Thinking : «On peut penser que c'est dû à la sténose»

Alternatives

- Normaliser sans retour possible : le sexe est-il un attribut?
- Organiser conceptuellement les notions : des médicaments (MENELAS)
- Normaliser en séparant conceptuellement les notions : ... et de la maladie

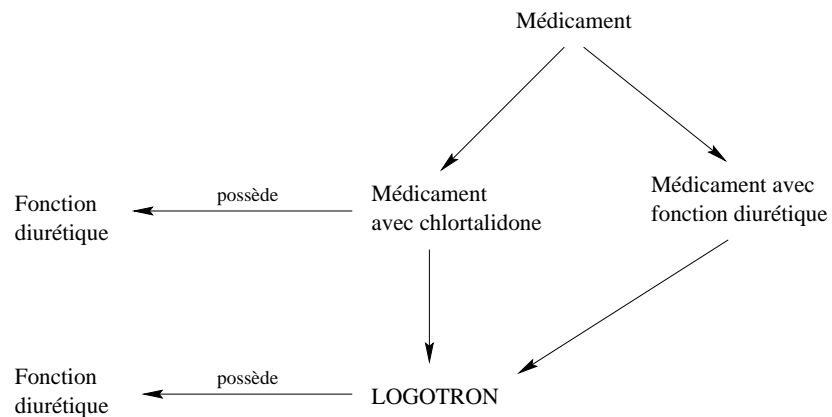
Le sexe est-il un attribut ? (1)



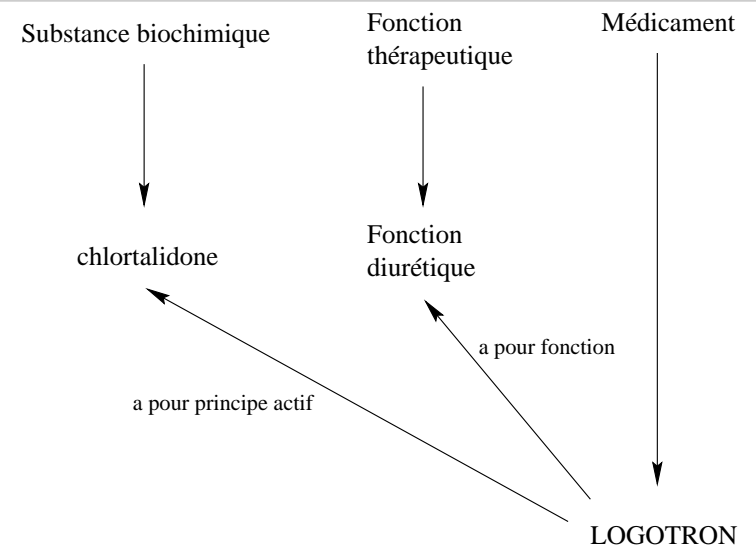
Le sexe est-il un attribut ? (2)

- 1 Deux types de concept, *patient* et *patiente* sur lesquels le système ne peut rien «dire» sauf qu'ils sont différents ;
- 2 Un concept «primitif» (*patient*), deux caractéristiques différentes, différenciant ainsi les deux concepts «définis».
- 3 idem 2^e cas **mais en plus** la différence est explicitement liée au sexe puisque c'est une valeur de l'attribut.

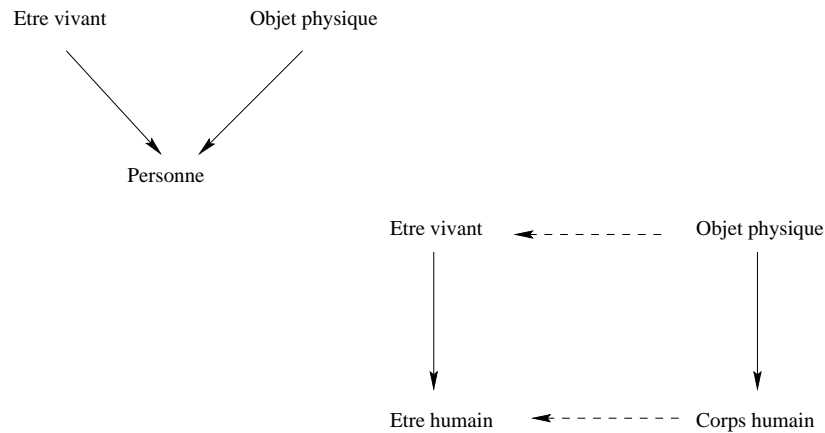
Taxinomie intuitive



Taxinomie fonctionnelle



Normaliser : séparer conceptuellement les notions (1)



Normaliser : séparer conceptuellement les notions (la maladie)

- Non stabilité des objets dans le discours : processus ou état ?
«l'évolutivité de la maladie est rapide» *versus* «la maladie est un état morbide»
- Non stabilité des objets dans le discours : processus physiologique ou mesure diagnostique ?
«le transit intestinal est bon» *versus* «le transit intestinal est de xx.xx»

Méthodologie

(Bachimont 2000)

- Hypothèse
- Principes
- Description

Hypothèse

- ☞ Des domaines se formalisant *a priori* peu ou pas et dont le mode d'accès aux informations et connaissances se fait essentiellement par le vecteur de la langue naturelle

La médecine : une science pour ses principes biochimiques, physiques et ses protocoles d'études épidémiologiques ; une pratique quotidienne où les informations se consignent dans des «dossiers », regroupements cohérents de documents essentiellements textuels

→ Accès à des corpus (livres de cours, comptes rendus, etc.)

La primauté du corpus

- Une expression des connaissances à travers une langue de *spécialité*, avec un vocabulaire, une terminologie et une sémantique spécifiques
- Utiliser une démarche, *de corpus*, et des outils terminologiques pour modéliser le domaine
- Description de l'utilisation de ces outils au sein d'une démarche méthodologique claire précisant à quel stade du processus et selon quelles modalités il convient de les utiliser

La normalisation

- Choisir un contexte de référence des objets dont on parle pour fixer et stabiliser leur interprétation \Rightarrow *stabilité référentielle* et *stabilité relationnelle*
- Structurer et organiser les *concepts* ainsi créés

Langue naturelle vs langage formel

Langue naturelle Pas de primitives, des définitions circulaires, des énoncés imprécis, contextuels, des objets «instables»



Langage formel Des expressions logiques construites avec des primitives, des connecteurs, des quantificateurs...

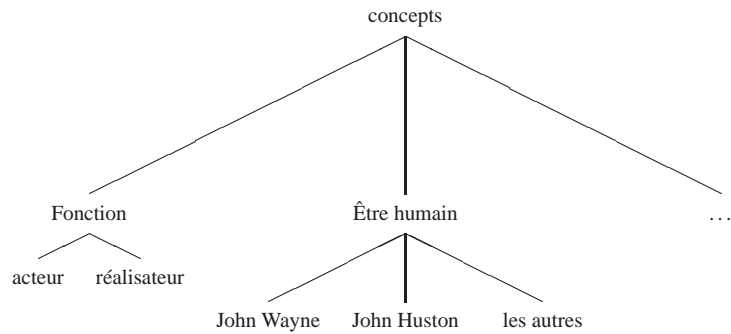
- Choisir un contexte de référence des objets dont on parle pour fixer et stabiliser leur interprétation \Rightarrow *stabilité référentielle* et *stabilité relationnelle*
- Structurer et organiser les *concepts* ainsi créés

Méthode (Bachimont, *synthèse TIA*, etc.)

- 1 **Analyse de corpus.** Extraire des candidats-termes, des relations, des verbalisations, ...
 \Rightarrow base de connaissances terminologique (BCT)
 Des travaux sur la manière de développer des «produits intermédiaires» qui permettraient de mieux «gérer» le passage du linguistique au conceptuel

② **Normalisation sémantique.** Selon le contexte et la tâche

➤ arbre de concepts



③ **Engagement ontologique.**

➤ treillis formel

④ **Opérationnalisation** dans un langage de représentation des connaissances – e. g. les graphes conceptuels

➤ treillis computationnel

[Etre-humain: JohnWayne]–(a-pour-fonction)→[acteur]

[Etre-humain: JohnHuston]–(a-pour-fonction)→[realisateur]

Engagement sémantique

Ces primitives n’existent pas *a priori*

- Repartir des corpus textuels du domaine
- Dégager un ensemble de termes en construisant un système de *différences* entre ces termes (structure d’arbre)
- «Normalisation sémantique» suivant un point de vue particulier – ex : un médicament défini suivant son principe actif (pharmacologue), ou de sa fonction (clinicien) : contexte fixé par la normalisation

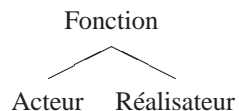
Construction de l’ontologie régionale qui n’est valable que localement, qui comporte un contexte privilégié pour déterminer des significations adaptées à une tâche.

Principes différentiels (1)

Construire l’arbre en respectant des principes différentiels :

- ① **Principe de communauté avec le père**
- ② **Principe de différence avec le père**
- ③ **Principe de différence avec les frères**
- ④ **Principe de communauté avec les frères** (explicitation de l’attribut choisi)

Principes différentiels (2)



- ❶ Fonction d'un être humain dans un film
- ❷ Jouer un rôle du film
- ❸ Le réalisateur qui fait le film, l'acteur qui y joue
- ❹ apparaître *dans* la film *versus* travailler *sur* le film
(fonction sur le plateau)

Engagement ontologique

Maintenant, les primitives existent...

- ☞ les concepts formels vérifient les relations d'identité et d'inclusion unissant les concepts sémantiques
- ☞ ils ne vérifient pas les relations d'exclusion car on considère maintenant les extensions des objets
- ☞ un treillis (structure algébrique des ensembles)

Construction de l'ontologie formelle ou référentielle qui n'est (toujours) valable que localement.

Retour sur les étapes

- ❶ **Analyse de corpus.** (Choix et) description linguistique du corpus caractérisant le domaine ; reflète les normes sociales, techniques, pratiques. . .
Corpus → Signifié
- ❷ **Normalisation sémantique** du signifié linguistique pour dégager les objets du domaine et leur type
Signifié → Signifié normé (ou Concept linguistique)
- ❸ **Engagement ontologique**
Concept linguistique → Concept formel
- ❹ **Opérationnalisation** dans un langage de représentation des connaissances

Terminologie

(Bourigault & Slodzian 1999)

- Principes (le corpus)
- Outils
- Ressources

Principe

L'hypothèse selon laquelle l'expert d'un domaine est dépositaire d'un système conceptuel qu'il suffit de mettre à jour est non productive.

La tâche d'analyse terminologique vise alors la construction d'une description des structures lexicales à l'œuvre dans le corpus textuel à partir d'une analyse réglée de ce corpus.

(Bourigault & Slodzian 1999)

Corpus

- Le texte est le point de départ de la description lexicale à construire
- Le corpus doit être soigneusement choisi en fonction du domaine et de l'application visée (la tâche)
- La taille des corpus et la masse d'informations contenues implique l'utilisation d'outils de terminologie textuelle (concordancier, extracteur de candidats termes, extracteur de relations, etc.)
- Description de l'utilisation de ces outils au sein d'une démarche méthodologique claire précisant à quel stade du processus et selon quelles modalités il convient de les utiliser

Travail

- La médiation d'un analyste entre l'étude du corpus et l'expert est nécessaire pour se soustraire aux usages langagiers idiosyncrasiques
- La différence de pratiques à l'intérieur d'un domaine induit des points de vue différents qu'il faut arbitrer
- Ce médiateur doit être en charge de l'application
- L'expert est un partenaire du linguiste terminologue. Il est sollicité pour valider les descriptions construites

Terme

- Les termes ne sont pas des «unités de connaissances» qui viendraient «habiter la langue». Il n'y a donc pas de redécouverte d'un système notionnel préexistant
- Les notions n'ont pas d'antériorité ou de priorité sur les mots : la terminologisation est un processus parallèle à l'élaboration conceptuelle
- Le terme doit être attesté en corpus (il peut exister, pour des raisons de structuration, des termes non attestés en corpus mais cela doit être assumé)

Un terme (ou concept linguistique) est construit. Il est le produit d'un travail d'analyse et d'une négociation avec l'utilisateur. Il doit être pertinent vis-à-vis du corpus comme de l'application visée.

Ressources

- Index structurés pour les documentations électroniques
- Thesaurus pour les systèmes d'indexation automatiques
- Terminologies de référence pour les systèmes d'aide à la rédaction
- Référentiels terminologiques pour les systèmes de gestion de données techniques
- Ontologies pour les SBC
- Ontologies pour la gestion des connaissances
- Réseaux lexicaux spécialisés pour les moteurs de recherche thématique sur le Web
- Bases de connaissances terminologiques
- ...

Thesaurus : Langage documentaire fondé sur une structuration hiérarchisée. Les termes y sont organisés de manière conceptuelle et reliés entre eux par des relations sémantiques. Organisé alphabétiquement, il forme un répertoire alphabétique de termes normalisés pour l'analyse de contenu, le classement et donc l'indexation de documents d'information.

Définitions

Indexation : Processus destiné à représenter par les éléments d'un langage documentaire ou naturel des données résultant de l'analyse du contenu d'un document ou d'une question. On désigne également ainsi le résultat de cette opération (voir Index).

Index : Table alphabétique des mots, des termes correspondant aux sujets traités, des noms cités dans un livre. Dans le domaine technique, l'index d'un document est alors son analyse sommaire présentée sous forme de mots-clés, rubriques, etc.

Langage documentaire : Ensemble organisé de termes normalisés, utilisé pour représenter le contenu des documents à des fins de mémorisation pour une recherche ultérieure. On distingue essentiellement, dans les langages documentaires, les classifications et les thesaurus.

Pré-coordination : Principe suivant lequel les combinaisons entre les termes d'un langage documentaire s'effectuent au cours de son élaboration, par exemple la création des termes composés dans un thesaurus.

Réutilisabilité (1)

Réutilisation d'ontologies pour la spécification et l'analyse de systèmes : leçons provisoires

- ☞ Différente suivant leur niveau de préoccupation et de généralité
- ☞ Retours d'expériences : l'utilisation d'ontologies implique une adaptation souvent importante
- ☞ Facilité si les ontologies sont modulaires, bien délimitées, organisées en hiérarchies
- ☞ Facilité si elles incluent les buts de leur conception et une documentation des choix faits dans sa réalisation
- ☞ Une ontologie peut aider au *processus* de modélisation en proposant un mode de découpage et de structuration des connaissances du domaine MAIS le *problème d'interaction* implique la mise en place de meilleurs processus
- ☞ Comment coupler les approches par ontologies avec celles par *patterns* d'analyse ou de spécification ?

Réutilisabilité (2)

- ☞ Il n'existe pas d'ontologies universelles
- ☞ La réutilisation d'une ontologie implique de connaître les buts pour lesquels elle a été réalisée
- ☞ Dépendance entre la conceptualisation du monde et une notion de point de vue sur le monde, dépendant d'un mode d'action sur celui-ci : le raisonnement organise une description pertinente du monde, compte tenu de ses fins
- ☞ Intérêt pour l'acquisition : guider précisément le type de découpage et de description du monde qui correspond à un raisonnement donné