

Ontologies informatiques: dans un système d'information, un petit peu de sémantique peut vous emmener très loin.

Introduction

A faire à la fin.

De l'Ontologie à l'ontologie ou l'histoire d'une notion à la recherche d'un mot.

"Ontologie" est un mot que l'informatique a emprunté à la philosophie au début des années 90. Si vous ouvrez un dictionnaire tel que le petit Larousse Illustré la définition n'éclaira probablement pas beaucoup votre lanterne quant à l'intérêt d'importer cette notion en informatique: "(1) Etude de l'être en tant qu'être, de l'être en soi (2) Etude de l'existence en général, dans l'existentialisme."

Pourtant, à y regarder de plus près nous pourrions, à l'extrême inverse, penser que beaucoup de d'ingénieurs en informatique sont des messieurs Jourdain de l'ontologie. A titre d'exemple, lorsqu'ils conçoivent un schéma de classes pour implanter une application, les ingénieurs en informatique s'interrogent sur les objets que cette application va manipuler, les classes qui les regroupent, les caractéristiques communes à tous les objets de chaque classe, les relations qui peuvent exister entre ces objets, etc. En d'autres termes, ces ingénieurs s'interrogent sur ce qui définit ces classes d'objets, ce qui permet d'identifier qu'un objet appartient à une classe, ce que cette appartenance signifie en terme de contenu ou de manipulations possibles, bref ils s'interrogent sur la définition existentielle des classes d'objets mobilisés dans les scénarios de l'application qu'ils développent.

En philosophie, l'Ontologie est une branche fondamentale de la Métaphysique qui s'intéresse à la notion d'existence, aux catégories fondamentales de l'existant et étudie les propriétés les plus générales de l'être. L'Ontologie a donc un rapport direct avec notre conception de la réalité. L'ontologue qui interroge nos conceptualisations du monde et l'ingénieur qui conçoit ses représentations logicielles ne sont-ils pas plus proches qu'il n'y semblait initialement?

Si la notion d'ontologie informatique et l'exemple de la programmation orientée objets ont cette ressemblance, c'est qu'ils ont un ancêtre commun: les systèmes de l'intelligence artificielle symbolique. Les débuts de cette branche de l'intelligence artificielle se confondent avec les débuts de l'informatique car dès ses prémices, l'informatique a perpétué le rêve d'automate de simuler et dépasser l'intelligence humaine avec des systèmes artificiels.

Si la branche de l'intelligence artificielle à laquelle nous nous intéressons ici est qualifiée de symbolique c'est parce qu'elle repose sur des représentations formelles des connaissances sous la forme de symboles que le système peut stocker et manipuler (ex: langages et opérations logiques, structures et opérations de graphes). Contrairement à d'autres approches, ces représentations sont à la fois compréhensibles par les humains et manipulables par les systèmes en appliquant des règles de manipulations définies sur les symboles de ces représentations et dont l'interprétation simule, par exemple, un raisonnement.

La notion d'ontologie existait sans être nommée et de façon transversale dans les différents systèmes de représentation de connaissances dès les années 70: c'est la TBox des logiques de description où l'on décrit les types de termes qui existent dans notre représentation et leurs caractéristiques; c'est le support des graphes conceptuels où l'on décrit des hiérarchies de multi-héritage entre des types de concepts ou des types de relations; c'est les schémas des Frames et les classes des langages de représentation par objets. Il faudra pourtant attendre les années 90 pour que le mot ontologie soit adopté par toute la communauté et sa définition fait encore couler de l'encre (électronique).

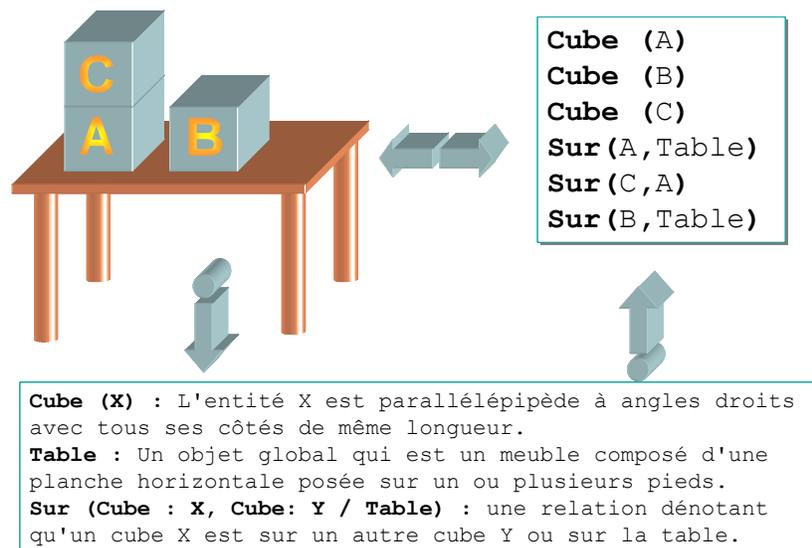
Aux grands mots les grands remèdes

Comble de l'histoire donc, cette notion d'ontologie qui s'attache tant à la définition précise des concepts que nous manipulons a longtemps cherché une définition et un mot. Ce retard est probablement largement dû à la nature abstraite même de la notion d'ontologie. Pour définir la notion d'ontologie informatique, commençons par une pirouette:

Tout le monde connaît la *Paléontologie*, cette science qui étudie les êtres et organismes vivants ayant existé au cours des temps géologiques. Si on enlève le préfixe *palé* il nous reste les racines grecques *ontos* (ce que existe, l'existant) et *logos* (le discours, l'étude). L'Ontologie désigne donc l'étude des propriétés générales de ce qui existe. En important cette notion en informatique nous sommes passés de la science (l'Ontologie) à est un objet (une ontologie). Une ontologie informatique est une représentation de propriétés générales de ce qui existe dans un formalisme supportant un traitement rationnel. C'est le résultat d'une formulation exhaustive et rigoureuse de la conceptualisation d'un domaine. Cette conceptualisation est souvent qualifiée de partielle car, en l'état de l'art, il est illusoire de croire pouvoir capturer dans un formalisme toute la complexité d'un domaine.

C'est pour cet aspect de description de l'existant et de ses catégories que les ontologies informatiques ont emprunté leur nom à l'Ontologie philosophique. De ce rapprochement vient aussi la possibilité d'adapter des méthodes de la philosophie pour proposer des méthodes d'ingénierie d'ontologies.

Pour faire simple, considérons l'exemple de la Figure 1. On y voit une certaine scène du monde. La description de cette scène demande deux choses: (1) un vocabulaire non ambigu aussi appelé vocabulaire conceptuel ou ontologie (2) une énonciation des faits de la scène, reposant en utilisant le vocabulaire de l'ontologie.



1. L'incontournable exemple des cubes.

Pour mieux saisir cette notion d'ontologie, répondons à une question récurrente: **Que met-on dans une ontologie?**

Une ontologie inclut généralement une organisation hiérarchique des concepts pertinents et des relations qui existent entre ces concepts ainsi que des règles et axiomes qui les contraignent. L'ontologie définit des concepts (principes, idées, catégorie d'objet, notions potentiellement abstraites) et des relations.

L'ensemble des propriétés d'un concept s'appelle sa compréhension ou son *intension* et l'ensemble des êtres qu'il englobe, son *extension*. Prenons un concept volontairement anonyme *C*, nous pouvons lui associer:

- une intension: c'est un ensemble de propriétés qualitatives ou fonctionnelles communes aux individus auxquels le concept s'applique, et permettant de définir le concept, ex: "*C* est une sous catégorie de véhicules de transports automobiles conçus et aménagés pour le transport d'un petit nombre de personnes (7 ou moins) ainsi que d'objets de faible encombrement et dotée d'au minimum trois roues".
- une extension: un ensemble d'entités qui rentrent dans cette catégorie, ex: {la twingo de Rose, le kangoo d'Olivier, la clio d'Alain, ... }

Pour exprimer, communiquer un concept nous choisissons une représentation symbolique, souvent linguistique et verbale, parfois iconique. Dans le cas précédent nous pouvons donner comme exemples de représentations linguistiques, les termes de "voiture", "automobile", "auto" ou "véhicule auto", "tacot", "bagnole", "tire" ou "caisse". Nous dissociions donc les concepts et leurs manifestations linguistiques. Un terme n'est pas un concept et vice-versa. Un terme peut être ambigu alors qu'un concept n'a qu'un seul sens une seule définition. Il faut alors gérer les problèmes de synonymie (un concept dénoté par plusieurs termes) et d'homonymie (un terme dénotant plusieurs concepts).

De la même façon que pour les concepts, l'ontologie définit des relations pouvant exister entre les instances de ces concepts. Prenons une relation volontairement anonyme *R*, nous pouvons aussi lui associer:

- une intension, ex: "*R* est une relation entre une personne ou un groupe qui a créé un document, son contenu intellectuel, son arrangement ou sa forme";
- une extension, ex: { (Hugo, Notre Dame de Paris), (Jean Markale, Le cycle du Graal), ... }
- des représentations linguistiques: "a écrit", "auteur de", "écrivain de"

Les relations possèdent en plus une signature spécifiant les types d'instances qu'elles relient ex: (Personne ou groupe ; Document).

Dans une ontologie, les intensions sont organisées, structurées et contraintes pour représenter notre conception du monde et de ses contraintes (ex: une voiture est forcément un véhicule).

L'ontologie capture les intensions et les lois qui les régissent afin de rendre compte des aspects de la réalité choisis pour leur pertinence dans les scénarios d'application considérés.

La représentation des intensions et de l'ontologique peut faire appel à des langages plus ou moins formels (graphes, logiques, langue restreinte) selon l'opérationnalisation envisagée pour l'ontologie. La construction formelle de l'intension donne une représentation précise et non ambiguë de la manière dont on peut concevoir son sens, ce qui permet sa manipulation logicielle et son utilisation comme une primitive de représentation de connaissances pour décrire et structurer, par exemple, des données, des logiciels, des utilisateurs des communautés, etc.

Dans l'ontologie, les intensions sont habituellement organisées en taxonomie ou hiérarchie de types. L'importance de l'organisation taxinomique se justifie par le fait que la classification ou

identification (le fait de déterminer si quelque chose appartient à une classe) et la catégorisation (le fait d'identifier les catégories existantes) sont des inférences élémentaires que nous faisons à longueur de journée. Prenons l'exemple simple d'une conversation entre deux personnes:

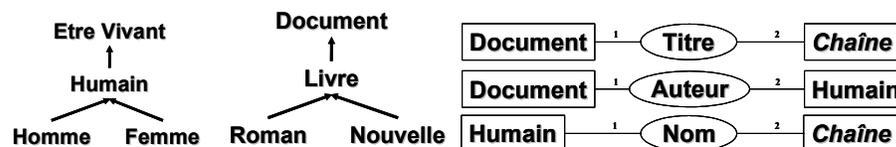
- " - Tu connais un *restaurant* proche ?
- Il y a un *pizzeria* au coin de la rue.
- Merci."

Dans une conversation aussi banale, la première personne a généralisé sa requête au concept de *restaurant*, qui représente la catégorie la plus abstraite recouvrant toutes les formes de réponses acceptables. La deuxième a, probablement sans même y prêter attention, utilisé sa taxonomie de concepts pour en déduire qu'une *pizzeria* est un *restaurant* et que par conséquent sa réponse est pertinente. Le fait que cette connaissance taxonomique soit partagée est implicite puisque la deuxième personne suppose que sa réponse sera comprise sans préciser qu'une *pizzeria* est un *restaurant* et que c'est effectivement le cas. Le recours à des conceptualisations partagées et aux inférences qu'elles permettent est donc au cœur d'activités aussi simple que cet échange d'information. Le fait de rendre explicite les connaissances ontologiques et de s'assurer de leur nature consensuelle est un des problèmes majeurs de l'ingénierie ontologique.

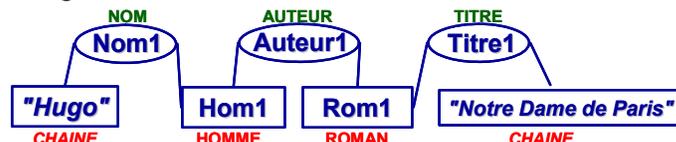
Ainsi dans un système d'information le simple ajout de cette connaissance peut permettre d'améliorer considérablement les capacités des machines. Prenons l'exemple très simplifié où l'on recherche des livres écrits par un certain "Hugo". Si votre système d'information se contente de travailler au niveau textuel avec les mots clefs "Hugo" et "Livre", vous verrez apparaître plusieurs problèmes:

- la précision: le système ne saura pas faire la différence entre le nom de famille "Hugo", le prénom "Hugo" ou le nom de rue "Hugo";
- le rappel: le système s'il rencontre le terme "R-o-m-a-n" ne saura pas qu'il est pertinent pour votre requête car il cherche le mot "L-i-v-r-e"

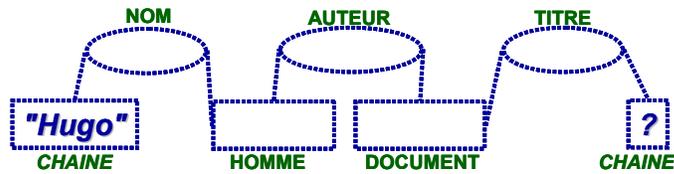
Si maintenant vous expliquez au système quelques aspects de notre réalité sur les humains, les documents et les relations entre les deux:



Puis que vous utilisez ce vocabulaire pour décrire la réalité ex: qu'un homme dont le nom est "Hugo" a écrit un roman intitulé "Notre Dame de Paris".



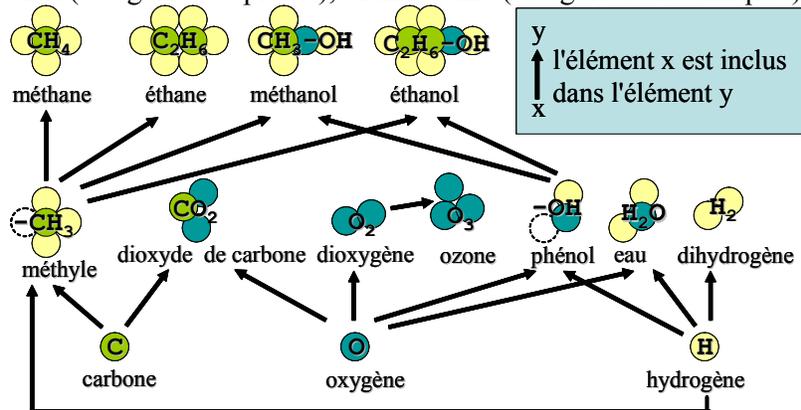
Alors vous pouvez formuler une requête non ambiguë avec ce même vocabulaire pour rechercher les documents écrits par un certain "Hugo":



Et en utilisant la logique de votre langage le système peut inférer qu'un roman est un livre, un livre est un document donc un roman est un document et que la réponse "Hugo à écrite le roman Notre Dame de Paris" est valide.

En voyant cet exemple, il ne faut cependant pas confondre ontologie et taxonomie. Les connaissances ontologiques dépassent largement les connaissances taxonomiques. Ainsi, on peut trouver dans une ontologie:

- des connaissances de composition: par exemple en chimie (catégories d'éléments), en production (catégories de pièces), en médecine (catégories anatomiques), etc.



- des définitions complètes, ex: une personne est un directeur si et seulement si il existe une organisation qui soit dirigée par cette personne;
- des contraintes d'intégrité, ex: un livre édité a un et un seul ISBN, un parent ne peut pas être plus jeune que ses enfants;
- des fonctions de calcul, ex: le rythme cardiaque conseillé pour une personne lors d'un effort cardio-vasculaire est $(220 - \text{âge}) * 0.65$;
- des propriétés algébriques: la relation "est marié avec" est symétrique i.e. si Thomas est marié avec Stéphanie alors le système peut aussi déduire que Stéphanie est mariée avec Thomas et vice-versa;
- des connaissances par défaut, ex: par défaut une voiture a quatre roues;
- des relations inverses, ex: "faire partie de" et l'inverse de "inclure" i.e. si une portière fait partie d'une voiture alors la voiture inclut la portière et vice-versa;
- des règles du domaine, ex: en biologie, pour chaque récepteur qui active une fonction moléculaire, si cette fonction joue un rôle dans le fonctionnement de l'organisme alors le récepteur joue le même rôle;

Le contenu d'une ontologie varie aussi avec le type d'ontologie considéré. Une ontologie de domaine contiendra des connaissances propres à un domaine de connaissances (ex: l'aviation). Une ontologie de tâche contiendra des connaissances propres à une activité (ex: le diagnostique). Une ontologie de haut niveau contiendra des connaissances abstraites très générales destinées à rassembler d'autres ontologies (ex: notions d'entité, d'évènement, de rôle, etc.). Le contenu dépendra aussi du degré de formalisation (langue naturelle, langage restreint, formalisme simple, logiques complexes).

Une autre question récurrente liée à la nature des ontologies est: **pourquoi avoir séparé ce type de connaissances des autres?** On peut donner plusieurs raisons:

- Factoriser des connaissances: dans un modèle les connaissances ontologiques sont des connaissances toujours vraies quelque soit l'état du système et les descriptions faites. L'ontologie permet de les factoriser et de ne pas avoir à les répéter pour chaque occurrence, ex: on dira dans l'ontologie qu'une voiture est un véhicule (car c'est toujours vrai) mais on ne lui donnera pas une couleur car cela change d'une voiture à une autre.
- Réutiliser et échanger des connaissances: les connaissances ontologiques étant séparées elles peuvent être réutilisées dans plusieurs applications et ces réutilisations (totales ou partielles) peuvent être la base d'une interopérabilité entre différents systèmes.
- Compiler les connaissances et optimiser les inférences: les connaissances ontologiques peuvent faire l'objet de traitement particulier pour leur donner des structures efficaces, certifier leur cohérence et optimiser les inférences qui les exploitent, ex: le calcul d'une fermeture transitive comme "si un coupé est une voiture et une voiture est un véhicule alors un coupé est un véhicule".

Une dernière question récurrente est à: **à quoi servent ces ontologies?**

Une ontologie informatique permet, en particulier grâce aux travaux de l'intelligence artificielle symbolique sur les systèmes à base de connaissances et les moteurs d'inférence, d'implanter des mécanismes de raisonnement déductif, de classification automatique, de recherche d'information et d'assurer l'interopérabilité entre plusieurs systèmes de ce type. Nous allons regarder quelques exemples de systèmes d'information.

Rendre l'implicite explicite: l'exemple des ontologies dans les systèmes d'information

Beaucoup d'entités sociales sont des communautés épistémiques: que ce soit leur raison d'être (réseaux d'intérêt, équipes de recherche, écoles, etc.) ou un résultat de leur fonctionnement (entreprises, administrations, associations, etc.), elles ont à gérer et maintenir des connaissances. De l'agilité de ces entités à détecter, mémoriser, se remémorer et activer leurs connaissances dépend leur agilité à répondre au monde extérieur (innovation des recherches, temps de réponse au marché, qualité des formations, etc.). La connaissance est un capital et un système d'information performant est un atout. **Au royaume de l'information, les connaisseurs sont rois.**

Les systèmes d'informations collectifs sont des applications singulièrement contraintes: les différences d'expérience, les différentes formations, les différentes cultures, les différents besoins, les différents points de vue, les différentes langues ou jargons, les différents médias et formats, les différents contextes d'utilisation, les différents droits d'accès, etc. posent une diversité de contraintes qui peuvent entraîner un système dans un cercle vertueux ou dans un cercle vicieux selon quelles sont importantes pour les usages envisagés et respectées par les solutions choisies.

L'introduction d'une ontologie dans un système d'information vise à réduire, voire éliminer, la confusion conceptuelle et terminologique et de tendre vers une compréhension partagée pour améliorer la communication, le partage, l'interopérabilité et la réutilisabilité. Une ontologie informatique offre un cadre unificateur et fournit des primitives améliorant la communication entre les personnes, entre les personnes et les systèmes, et entre les systèmes.

Intégrer une ontologie à un système d'information permet donc de déclarer formellement un certain nombre de connaissances utilisées pour caractériser les informations gérées par le système et de reposer sur ces caractérisations et la formalisation de leur signification pour automatiser des tâches de traitement de l'information.

Dans un moteur de recherche c'est, par exemple, pouvoir améliorer: la précision de cette recherche d'information, en évitant des ambiguïtés au niveau terminologique (ex:

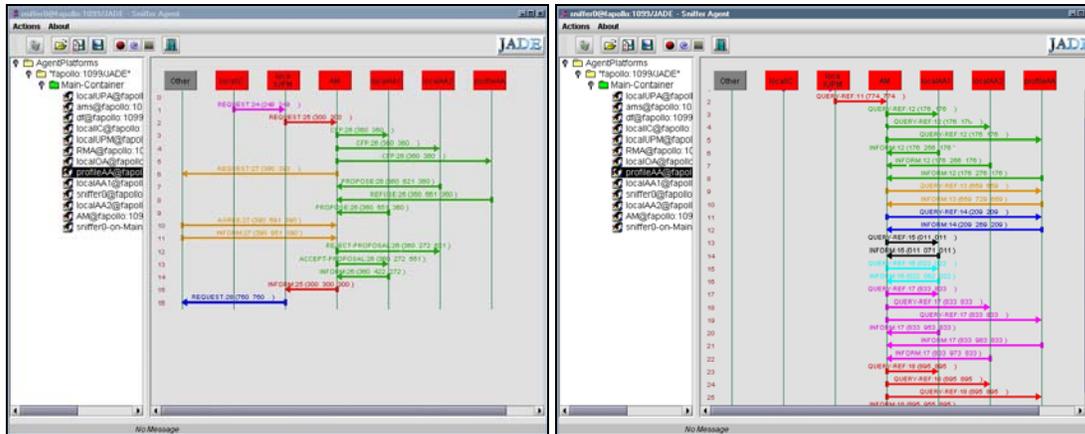
homonymie); le taux rappel de cette recherche d'information, en intégrant des notions plus précises ou équivalentes (ex: synonymie, hyponymie) ou en déduisant des connaissances implicites (ex: règles d'inférence); en relaxant des contraintes trop strictes en cas d'échec de la requête (ex: hyperonymie); en regroupant des résultats trop nombreux selon leur similarité pour les présenter de façon plus conviviale (ex: clustering). Le moteur de recherche Coresé utilise cette approche; voyons quelques exemples d'applications.

L'une de nos premières applications fut la gestion de documents dans le domaine du bâtiment. Aprobatiom (figure 2) est un logiciel permettant de guider la recherche de documents lors de la conception d'un nouveau bâtiment afin d'identifier et de réutiliser des documents traitant de la réalisation de bâtiments ressemblant au nouveau projet. L'ontologie est utilisée pour guider l'utilisateur dans l'expression d'une requête et pour caractériser et comparer les différents documents gérés.



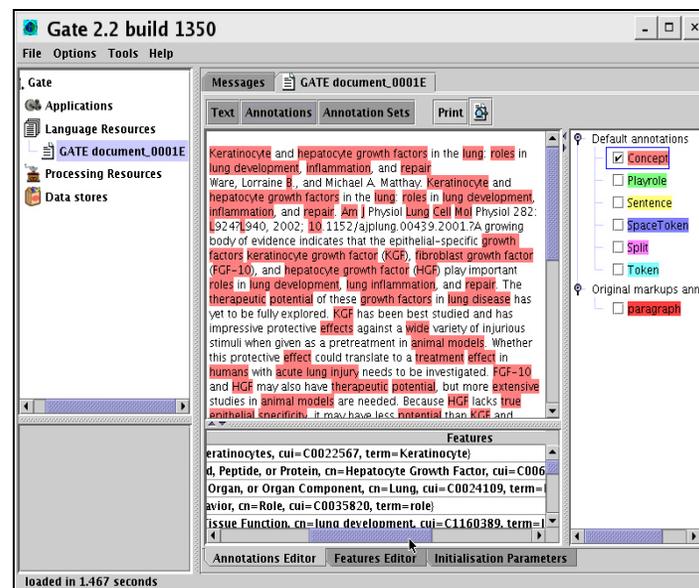
2. Aprobatiom: gestion de documents dans la conception d'un bâtiment.

Dans une entreprise, les connaissances, leurs sources et leurs utilisations sont distribuées. CoMMA (figure 3) est un système permettant de gérer l'annotation distribuée de documents. Donnons deux scénarios où l'ontologie est utilisée: (1) dans le cadre d'une activité de veille technologique, un employé identifie et annote un document intéressant pour son entreprise; l'ontologie est utilisée pour caractériser ce document, archiver cette annotation avec des annotations similaires pour maintenir la spécialisation des archives, identifier les profils de personnes potentiellement intéressées par ce document et leur envoyer un message de notification; (2) un nouvel employé recherche des informations sur les activités et structures de son entreprise; sa requête met en jeu différentes sources distribuées et l'ontologie est utilisée pour découper et distribuer la requête aux sources pertinentes.



3. CoMMA: négociation entre agents logiciels gérant la distribution des connaissances (à gauche des enchères pour décider de la meilleure base d'archivage pour une nouvelle annotation; à droite une coopération pour la résolution d'une requête dont les solutions requièrent plusieurs sources de connaissances).

Dans une communauté comme celle de la biologie cellulaire, la quantité de publications scientifiques est telle que l'on conçoit aisément qu'améliorer la recherche d'information l'intégration des connaissances est un réel enjeu. Les communautés de la biologie et de la médecine sont d'ailleurs parmi les pionnières du développement d'ontologies. Cependant les larges corpus existant font qu'une annotation manuelle ne passerait pas à l'échelle. En couplant les ontologies avec des outils d'analyse de la langue naturelle l'outil MeatAnnot (Figure 4) permet d'annoter formellement des textes avec des connaissances qu'ils décrivent. Par exemple, nous pouvons extraire d'un article une relation de causalité entre un gène et une maladie et annoter cet article avec cette connaissance afin (1) de l'utiliser dans des raisonnements d'analyses d'expériences et la confrontation de résultats (2) de garder une trace de la source de chaque connaissance.



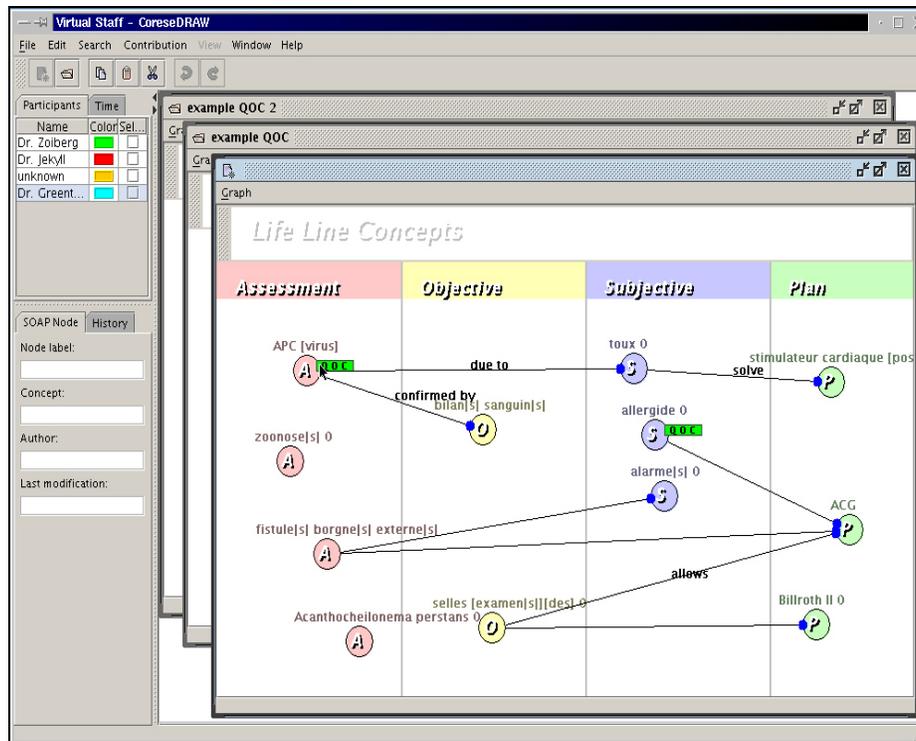
4. MEAT: annoter automatiquement des articles en biologie.

Chaque étudiant d'un cours l'assimile selon un parcours qui lui est propre. Les acquis et les impasses sont différents d'un étudiant à un autre. Dans QBL (Figure 5) la structure d'un support de cours est enrichie à l'aide d'une ontologie des éléments pédagogiques (ex: définition, théorème, exemple, question, etc.) et nous pouvons générer dynamiquement une interface de travaux dirigés ou travaux pratiques assurant une grande souplesse de navigation et suivant la progression de l'étudiant dans l'ensemble des notions à acquérir.

<p>Questions à résoudre :</p> <p>Glossaire</p>	<p>Question : Savoir lire un chronogramme (éditer)</p>	
	<p>Énoncé Étapes de résolution Réponses 2</p>	
	<p>Pour répondre aux questions posées, on suggère</p> <ol style="list-style-type: none"> 1. de lire sur le chronogramme en ordonnée (sur l'échelle verticale) les valeurs maximale et minimale, car elles délimitent l'intervalle demandé 2. de suivre le tracé du signal et lire sur l'axe vertical la plus grande valeur atteinte et la plus petite 3. de lire les graduations de l'axe horizontal, axe des abscisses, et soustraire la plus petite valeur de cet axe à la plus grande valeur atteinte. 4. de consulter les informations sur le blocage d'ordre zéro, et le signal bloqué. 5. de mesurer la durée des segments horizontaux, ou puisque cette durée est identique, de mesurer le nombre de paliers dans une division temporelle et d'en déduire la durée d'un seul palier. 	
	<p>chronogramme (éditer)</p>	<p>Historique des notions consultées:</p> <ul style="list-style-type: none"> • chronogramme
	<p>Définition (éditer)</p> <p>Quand on représente au cours du temps un signal audio, avec le temps en abscisse et l'intensité en ordonnée, on obtient un chronogramme .</p>	<p>Exemple</p>

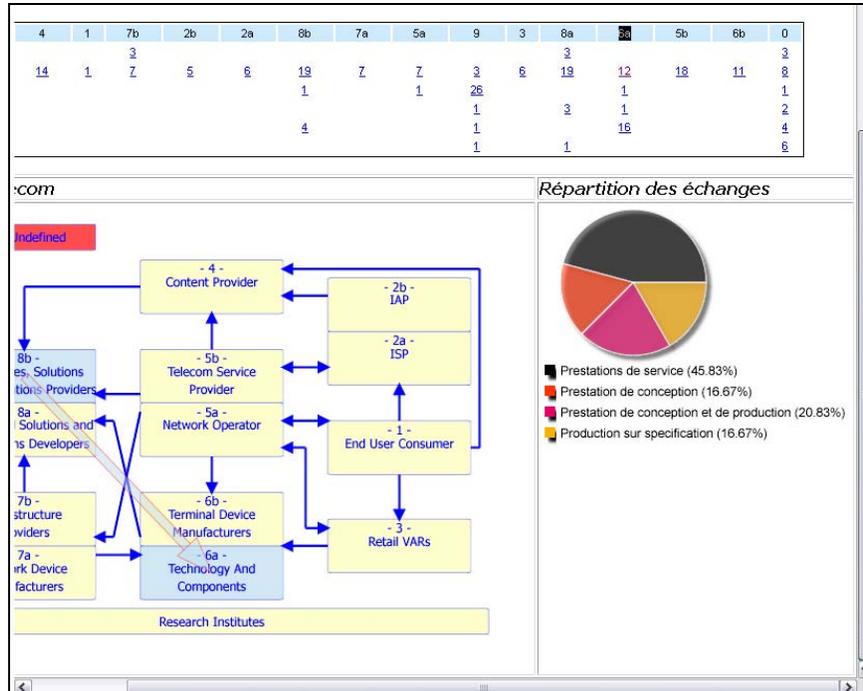
5. QBL: générer des supports de cours dynamiquement.

Une décision thérapeutique peut impliquer plusieurs experts médicaux et peut aussi nécessiter une connaissance détaillée des antécédents du patient. Ligne de vie (Figure 6) est système reposant sur une ontologie pour décrire les symptômes, les diagnostics, les options et les choix thérapeutiques, intégrer différentes contributions au dossier d'un patient et permettre une collaboration non ambiguë des personnes et des systèmes.



6. Ligne de vie: structurer le dossier patient pour aider la coopération médicale.

La Telecom Valley de Sophia Antipolis compte environ 70 membres, représentant plus de 10 000 emplois. La variété des compétences et des échanges possibles dans la chaîne de valeurs de ce parc des télécoms est non seulement élevée mais aussi dynamique et hautement spécialisée. Dans ce contexte, comprendre le paysage industriel pour les institutionnels régionaux ou trouver un partenaire pour les acteurs locaux sont de véritables problèmes d'intercompréhension et d'intégration d'informations. KmP est un portail public reposant sur une ontologie des compétences pour permettre à chaque acteur de décrire ses atouts et ses besoins afin d'améliorer la visibilité du parc et de permettre de nouveaux partenariats. Ainsi, les échanges peuvent être analysés (Figure 7), le système peut trouver des partenaires possibles même si le partenaire parfait n'existe pas (Figure 8) et l'état du parc, ses points forts et ses niches peuvent être visualisés en temps réel (Figure 9).



7. KmP: analyser les échanges.

([Détailer](#) / [Résumer](#) les réponses)

Pas de réponses exactes, mais **11 réponses approchées**

Guide de lecture

- d^m est la distance sémantique de la réponse approchée
- la valeur de d^m représente la distance sémantique entre la réponse approchée et une réponse exacte
- la différence de deux valeurs de d^m représente la distance sémantique entre deux réponses approchées

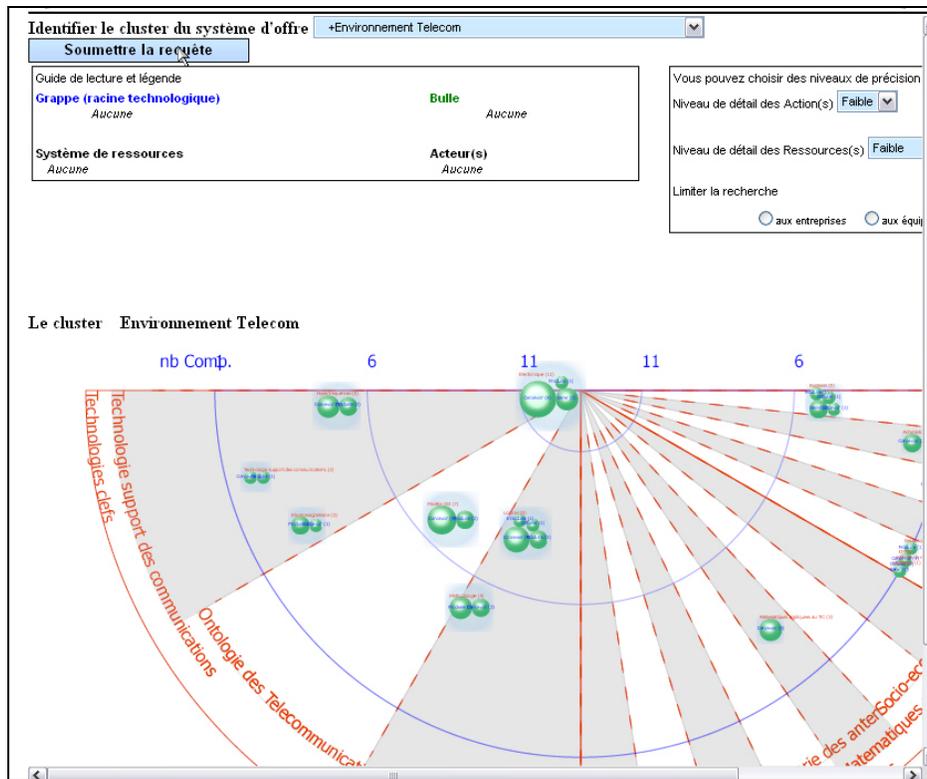
Entreprise [Philips Semiconductors](#)
Distance sémantique (d^m=39)

- [Concevoir , Design de test , 3G](#)

Entreprise [Philips Semiconductors](#)
Distance sémantique (d^m=39)

- [Spécifier / Caractériser , Produits GSM , 3G](#)

8. KmP: recherche d'un partenaire industriel le plus adapté.



9. KmP: visualiser les groupes de compétences.

La vie rêvée des ontologies

Les ontologies sont des objets vivants et chaque étape de leur cycle de vie pose des problèmes de recherche. La figure 9 résume ce cycle de vie qui rassemble sept activités:

L'activité de **détection des besoins**, lors de la conception, et l'activité d'**évaluation**, lorsqu'une ontologie est utilisée, ont en commun un certain nombre de tâches et de méthodologies de recueil (analyse d'entretiens, questionnaires et sondages, étude de l'ergonomie et des usages) et d'identification (ex: modélisation par scénarios). En complément la phase de détection des besoins demande un état des lieux initial approfondi car elle ne peut reposer sur des études précédentes ou des retours d'utilisation comme c'est le cas pour l'évaluation.

La phase de **conception initiale** et la phase d'**évolution** lors du cycle de vie ont en commun un certain nombre d'activités:

- spécification des solutions (ex: conception participative, maquettage, prototypage);
- acquisition des connaissances nécessaires (ex: analyse de textes, traitement automatique de la langue naturelle, plateformes collaboratives);
- conceptualisation/modélisation (ex: design pattern ontologiques, méta-ontologies, entretien avec les experts);
- formalisation (ex: méthodes et outils de l'Ontologie formelle, logiques de description et algorithmes de tableaux, analyse formelle de concepts, graphes conceptuels, formalismes du web sémantique RDF/S et OWL)
- intégration de ressources existantes (ex: alignement automatique d'ontologies, traduction) ;
- implantation (ex: graphes conceptuels, logiques de description, formalismes objets, logiques)

Notons aussi que l'évolution pose le problème de la maintenance de ce qui repose déjà sur l'ontologie: une ontologie est à la fois un objet vivant intéressant en soi et un ensemble de primitives pour décrire des faits du monde et des algorithmes sur ces faits. Lorsque l'ontologie change, ses changements impactent tout ce qui a été construit au dessus. Le maintien de la cohérence dans une ontologie et au dessus d'une ontologie, l'historique et la gestion des versions, la ré-ingénierie et la propagation des changements après modification, sont des questions de recherche encore ouvertes.

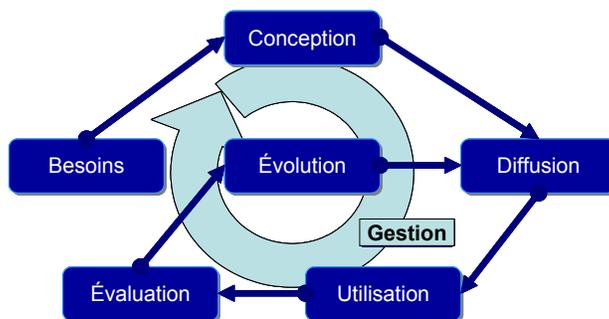
Un autre aspect du problème de conception et d'évolution est l'obtention et le maintien d'un consensus sur les choix de représentation et de conceptualisation faits dans l'ontologie. Suivant les usages, ce problème appelle des collecticiels et des outils de gestion des points de vue, des terminologies, des langues et des jargons différents.

La phase de **diffusion** s'intéresse au déploiement et la mise en place de l'ontologie. Les problèmes de cette phase sont fortement contraints par l'architecture des solutions. Dans un contexte d'application web on reposera sur des technologies idoines. Dans le partage de fichiers des architectures pair à pair ou autres architectures distribuées peuvent être utilisées. Dans l'intégration d'applications des architectures de services webs peuvent être une solution. Dans toutes ces architectures (serveurs web, web services, pair à pair, agents, etc.) la distribution des ressources (données, modèles, applications et utilisateurs) et leur hétérogénéité (syntaxes, sémantiques, protocoles, contextes, etc.) posent des problèmes de recherche sur l'interopérabilité (alignement et médiation) et le passage à l'échelle (larges bases, optimisation d'inférences, propagation de requêtes, syndication de données, composition de services, etc.).

La phase d'**utilisation** regroupe toutes les activités reposant plus ou moins directement sur la disponibilité de l'ontologie, par exemple: l'annotation de ressources (traitement de la langue, rétroingénierie de base de données, etc.), la résolution de requête (algorithme de projection de graphes avec contraintes), la déduction de connaissances et l'aide à la décision (moteurs d'inférence à base de règles), la navigation assistée et les services contextuels (analyse de contexte, identification et composition de services), l'analyse de gros volumes de connaissances (clustering, recherche de motifs récurrents, veille), etc.

Toutes ces activités ont en commun de poser le problème de la conception des interactions avec l'utilisateur et de leur ergonomie (interfaces dynamiques, lien sémiotique-sémantique, profils et contextes d'utilisation)

L'activité permanente de **gestion & planification** souligne qu'il est important d'avoir un travail de suivi et une politique globale pour détecter ou déclencher, préparer et évaluer les itérations du cycle et s'assurer que l'on reste dans le cercle vertueux des systèmes d'information contribution-utilisation-crédation.



10. Cycle de vie d'une ontologie

Conclusion

La notion d'ontologie en informatique prédate largement l'utilisation du mot. Cette idée existe depuis les années 70 en intelligence artificielle symbolique mais si le mot est parfois pris pour un buzzword et si l'appellation d'ontologie évoluera peut être encore, en particulier en donnant des raffinements (ex: folksonomie), la notion, elle, est utilisée depuis longtemps et ne semble pas prête à disparaître.

Au contraire, chaque année, le spectre d'applications et de domaines s'intéressant aux ontologies s'élargit. Anciennement réservée aux systèmes experts simulant des raisonnements humains dans des domaines spécifiques, l'ontologie se retrouve maintenant dans une large famille de systèmes d'information et est utilisée pour: décrire et traiter des ressources multimédia, assurer l'interopérabilité d'applications en réseaux, piloter des traitements automatiques de la langue naturelle, construire des solutions multilingues et interculturelles, permettre l'intégration de sources hétérogènes d'information, décrire des protocoles d'interactions complexes, vérifier la cohérence de modèles, raisonnements temporel et/ou spatial, approximation logique, etc. Les applications se retrouvent dans de nombreux domaines: systèmes d'intégration d'informations géographiques, systèmes de gestion de ressources humaines, système d'aide à l'analyse en Biologie, systèmes de commerce électronique, systèmes d'enseignement assisté par ordinateur, système de gestion de connaissances, bibliothèques électroniques, système d'échanges entre partenaires industriels, etc.

Grâce au web sémantique, l'ontologie a trouvé un formalisme standard à l'échelle mondiale et s'intègre dans de plus en plus d'applications web sans même que les utilisateurs ne le sachent et au profit des logiciels qui, à travers elles et les descriptions qu'elles permettent, peuvent proposer de nouvelles fonctionnalités. On ne compte plus non plus les ontologies de domaines: ontologie de la génétique, ontologie de la géométrie, ontologie pour les musées, ontologie médicale, ontologie pour l'enseignement, ontologie pour le bâtiment, ontologie de systèmes documentaires, ontologie pour la gestion, ontologie dans le secteur automobile, etc.

Cette expansion est loin d'être finie; à l'heure où l'ontologie se dote d'une ingénierie, elle est loin d'avoir révélé toutes ses applications. Parmi ses dernières évolutions, l'ontologie qui s'appliquait essentiellement à des données (documents, images, vidéos, etc.) est maintenant utilisée pour décrire des logiciels (ex: services web) leurs caractéristiques fonctionnelles (ex: types d'entrées, types de sorties) et non fonctionnelles (ex: coût, qualité) et pourrait ainsi permettre l'identification, l'invocation et la composition dynamique d'applications à l'échelle du web. De même, l'ontologie commençait à être utilisée pour décrire les utilisateurs et s'étend maintenant à la description du contexte d'interaction pour doter les applications de ce que l'on appelle une conscience du contexte: préférences de l'utilisateur (langue, goûts, droits, etc.), caractéristiques du terminal (mobile, vocal, etc.), situation géographique (à l'étranger, dans une salle avec imprimante, etc.), activité en cours (au volant, en présentation, etc.), historique d'utilisation, etc. Autre exemple, si l'ontologie est actuellement utilisée pour faciliter l'accès à des informations et des applications, on pressent à l'inverse son utilisation dans la description et l'application de règles de sécurité et confidentialité décrites à de hauts niveaux d'abstraction et permettant de restreindre les accès avec une grande flexibilité.

Nos systèmes d'information sont de plus en plus complexes. Cette complexité, même si elle est artificielle puisqu'il s'agit de technologie, pose des défis scientifiques ardues qu'il nous faudra relever pour voir l'expansion technologique continuer, par exemple, dans l'assistance des sciences de la vie.

La possibilité de concevoir des systèmes autonomiques (i.e. qui se reconfigurent, s'adaptent au contexte, détectent leur fautes, se corrigent dans une certaine mesure) apparaît comme un facteur de passage à l'échelle pour la croissance technologique.

Rendre explicites les conceptualisations du monde sur lesquelles se basent les architectures logicielles, les structures de données, les choix de conception c'est aussi participer à cette évolution des applications et de leur programmation. Le défi actuel des ontologies est de passer dans les pratiques d'ingénierie logicielle pour que les conceptualisations actuellement sous-jacentes, en filigrane dans le code, dans ses commentaires ou dans sa documentation dans le meilleur des cas, soient le plus souvent possible rendues explicites et capturées dans des formalismes. Ainsi exposées, elles offrent des possibilités d'inférence aux systèmes informatiques, de réflexivité sur leurs connaissances et leurs traitements, d'introspection, d'alignement dynamique pour permettre l'interopérabilité, d'évolution dynamique, de description de l'affordance des composants logiciels pour rendre leurs interactions plus dynamiques et et leur gestion plus automatique. A quand la programmation orienté ontologie?

Remerciements

A tous les membres d'Acacia, en particulier aux autres membres permanents (Olivier Corby, Rose Dieng-Kuntz Alain Giboin) et aux doctorants dont les travaux sont cités ici (Sylvain Dehors, Khaled Khelif)

Notes et annexes de construction (à ignorer).

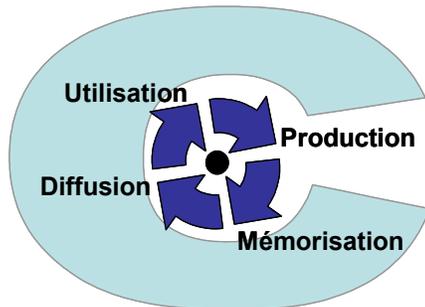


voiture / automobile



voiture / wagon passager

VOITURE: n.f. Véhicule de transport automobile conçu et aménagé pour le transport d'un petit nombre de personnes : *les enfants et moi avons pris la voiture pour arriver à l'école à l'heure.* Syn. de automobile
|| Dans le domaine des chemins de fer une voiture est un véhicule (*généralement non automobile*) transportant des voyageurs avec une capacité pouvant aller bien au-dessus de la *cinquantaine de personnes* (en particulier une voiture à deux étages) ; elle a au minimum quatre roues et *en général huit* et elle s'oppose au wagon qui transporte des marchandises.
|| Dans le domaine hippique la voiture est véhicule de transport hippomobile ex: un attelage, un fiacre, une roulotte || Dans le domaine minier la voiture désigne un chariot



Le presque logiquement juste, apprendre à relaxer intelligemment des contraintes logiques pour données des réponses approchées là où il n'y a pas de réponse exacte.

L'échec logique n'est pas synonyme d'absence de réponse.

Regroupement. KmP

Organisation de Corpus MeatAnnot

Quand des personnes s'accordent sur l'utilisation et la théorie spécifiée par l'ontologie, elles prennent "un engagement ontologique". L'idée est que pour représenter un domaine donné, il est nécessaire de se restreindre à un certain nombre de concepts significatifs suffisants pour interpréter ce domaine (Valente et Breuker, 1996). Une ontologie se fonde sur un certain engagement ontologique permettant de choisir un ensemble donné de concepts plutôt qu'un autre. Une ontologie comporte des définitions fournissant le vocabulaire conceptuel permettant de communiquer au sujet d'un domaine. Cela permet de définir (a) les concepts utilisables pour décrire la connaissance, (b) des relations entre de tels concepts et (c) leurs contraintes d'utilisation. Une ontologie est donc particulièrement intéressante quand elle décrit les connaissances *consensuelles* partagées par une communauté.

Les ontologies sont à la fois un modèle de notre monde en terme de catégories d'objets et de relations possibles entre ces objets (ex: il existe des véhicules, des personnes, et une personne peut avoir un véhicule) et un ensemble de primitives permettant de décrire des situations

particulières de ce monde (ex: la personne "Alain Giboin" possède la voiture "4563 VXT 06"). Les ontologies capturent un vocabulaire conceptuel fournissant une base cohérente non ambiguë et une référence commune utilisable pour décrire et échanger des modèles et implanter les inférences qui s'en nourrissent. Pour être utilisables comme références communes, les ontologies doivent résulter d'un consensus; elles réclament donc une construction et une maintenance collectives et sont dépendantes d'un point de vue socioculturel.

Les spécifications d'une ontologie soulèvent les questions de la couverture/portée et de la granularité qu'elle doit offrir.

Exhaustivité: c'est la largeur du domaine couvert *i.e.* l'étendue couverte par l'ontologie en terme de concepts et de relations nécessaires.

Spécificité: la profondeur du domaine couvert *i.e.* dans quelle mesure l'ontologie fournit précisément les concepts et relations nécessaires.

Granularité: le niveau de détail dans la définition formelle des concepts et des relations

Selon (Van Heijst et al, 1997), on peut distinguer plusieurs types d'ontologies :

- les ontologies du domaine expriment des conceptualisations spécifiques à un domaine. Elles sont réutilisables pour plusieurs applications sur ce domaine.
- les ontologies applicatives contiennent des connaissances du domaine nécessaires à une application donnée, elles sont spécifiques et non réutilisables
- les ontologies génériques ou ontologies de haut niveau (top-ontologies) expriment des conceptualisations valables dans différents domaines.

Dans cette l'optique d'une intégration aux conception logicielles, l'Ontologie en tant que branche de la philosophie n'est pas directement utilisable et nous devons développer une ingénierie des ontologies.

Types de connaissances

Arbre de la connaissance.

Arbre avec "vous êtes ici" sur la classe humain

interaction très fructueuse entre plusieurs domaines (en particulier la philosophie, l'informatique, la logique et les sciences cognitives

De par leur nature d'objet de modélisation de connaissances, les ontologies héritent d'un cycle de vie inspiré par le cycle de vie des (bases de) connaissances. De par leur nature de primitives de modélisation, leur cycle de vie se répercute sur tous les modèles et les algorithmes qui les exploitent.

BB

E-commerce

Spatial ontology geographic systems

Méréologie

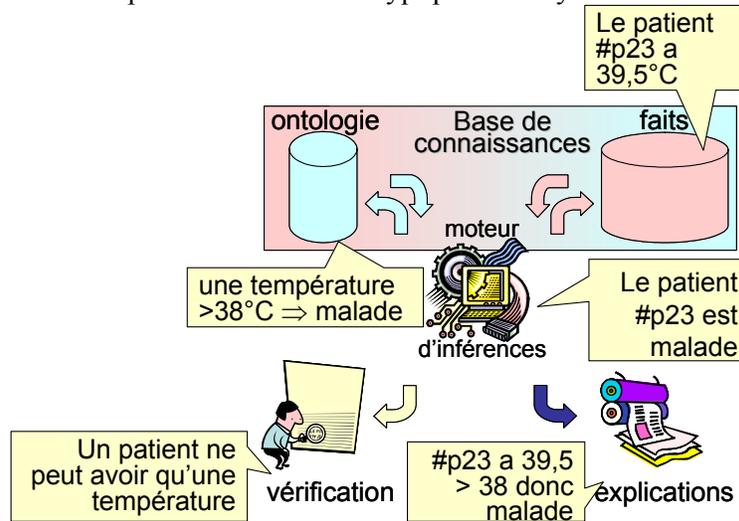
Méréo-topologie

Biomédical

Taxonomie lexicales

Approximation logique
Raisonnement temporel

La figure 2 donne les composants et les tâches typiques d'un système



L'incontournable exemple des cubes. Concevoir au niveau des connaissances, les workflow, les décisions, les tests, les opérations, etc. Conception et développement orientés ontologies.