

Les ontologies

1. Introduction

L'avènement des ontologies a actionné considérablement de nombreux domaines notamment la recherche d'information avec ses variations sémantiques et lexicales, la gestion des connaissances, les systèmes coopératifs, le e-commerce, et bien sur le Web sémantique.

Notons aussi que les ontologies viennent s'ajouter à cet amalgame de recherches et apportent un plus aux systèmes experts et surtout aux systèmes à base de connaissances (SBC), les successeurs des systèmes experts et répondaient aux besoins d'une coopération et à un dialogue entre le système et l'utilisateur humain. Citons quelques systèmes à titre d'exemple : les systèmes d'aide à la décision, système d'enseignement par ordinateur (EAO) et plus particulièrement la recherche d'information sur le web. Les SBC offraient une représentation des connaissances d'un domaine d'une part et des connaissances de raisonnement qui manipulent et utilisent ces connaissances du domaine d'autre part. L'idée de cette séparation modulaire était de construire au mieux et plus rapidement des systèmes à bases de connaissances en réutilisant le plus possible des composants génériques, que ce soit au niveau du raisonnement ou des connaissances du domaine [BAC03].

C'est dans cette perspective que les chercheurs ont proposé de développer ces connaissances sur la spécification d'une *ontologie*. Ce second chapitre, donne un aperçu horizontal sur la notion d'ontologie et ses constituants, ainsi qu'une description des besoins auxquels elle peut répondre et ces différents domaines d'applications. Nous soulignerons aussi une présentation des différentes classifications des ontologies et en fin, une description des langages utilisés pour les manipuler.

2. La notion ontologie

2.1. L'origine des ontologies

L'ontologie, un mot grec de racine, est composée d'onto (le participe présent du verbe *être*) qui est l'étude de l'être en tant qu'être et logos qui signifie *discours*. Son premier

sens trouve son origine en philosophie depuis *ARISTOTE*, où l'ontologie est l'étude des propriétés générales de ce qui existe.

L'apparition du terme « Ontologie » dans l'informatique a vu le jour au début des années 1990, grâce au projet d'ARPA Knowledge Sharing Effort (Effort de partager la connaissance) [GRU91]. Selon son sens philosophique, une ontologie est une explication systématique de l'être [GOM04].

2.2. Qu'est ce qu'une ontologie ?

En 1991, Neches et ses collègues donne une, des premières, définition du terme « ontologie » en citant : « *An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary* ». A savoir :

« Une ontologie définit les termes et les relations de base du vocabulaire d'un domaine ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire ».

Cette dernière définition nous dicte une méthode d'élaboration d'une ontologie, le repérage des termes de base et les relations entre les termes, ensuite relever les règles qui régissent de cette relation, et en fin donner des définitions à ces termes et ces relations. Le résultat, est une ontologie incluant non seulement des termes qui y sont explicitement définis, mais aussi des termes qui peuvent être créés par déduction en utilisant des règles [GOM99].

Une deuxième définition, au sens strict, est donnée en juin 1993, par Gruber, et qui est la plus citée en informatique plus précisément en intelligence artificielle (IA) : « *An ontology is an explicit specification of conceptualization*. » à savoir : « *Une ontologie est une spécification explicite d'une conceptualisation* ».

En 1995, Guarino donna une troisième définition de l'ontologie, et clarifia la définition de Gruber en énonçant que : « *Les ontologies sont des spécifications partielles et formelles d'une conceptualisation commune* ».

Quand à la notion de « partielles », la conceptualisation n'est jamais formalisée logiquement en sa totalité, du fait d'ambiguïtés ou du fait qu'aucune représentation de leur sémantique n'existe dans le langage de représentation d'ontologies choisi. Le terme « commune » du fait qu'une ontologie partage un savoir consensuel (reconnue par une communauté) [FUR02].

En 1997, Borst la reformula légèrement, « *an ontology is a formal, specification of a shared conceptualization* », puis Studer et ses collègues reprennent les deux définitions de Gruber et de Borst, en 1998, les fusionnent et énoncent la définition suivante :

« *an ontology is a formal, explicit specification of a shared conceptualization* » [OSC05].

Le terme « conceptualisation » oriente les ontologies sur l'aspect *sémantique* c'est-à-dire le sens des termes et que la littérature philosophique désigne par les *intensions*, par opposition aux *extensions*. La conceptualisation est représentée dans un langage, suivant lequel l'ontologie prendra la forme d'une théorie logique (ensemble de formules logiques) ou d'un réseau sémantique.

- Le terme « spécification explicite » oriente, par contre, les ontologies sur le côté syntaxique. Une spécification explicite signifie que les concepts, ainsi que les contraintes qui s'y rapportent sont explicitement définis.

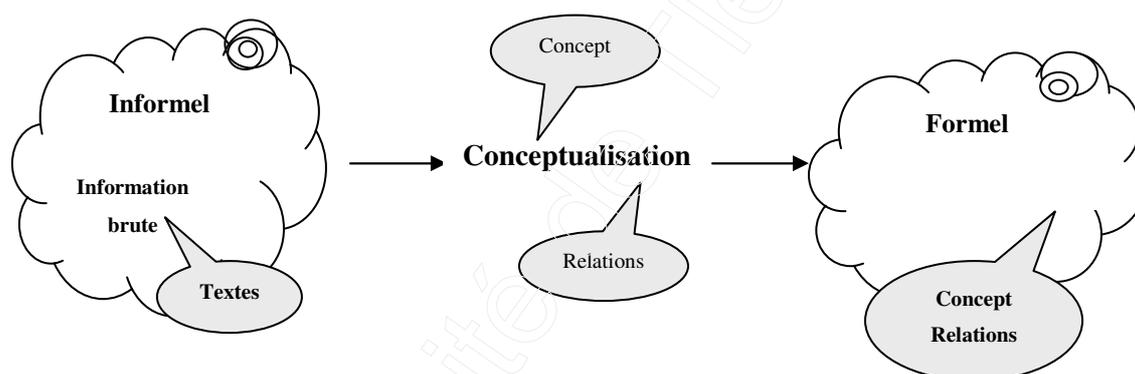


Fig. 14 : Conceptualisation d'une Ontologie

D'autres définitions d'ontologie viennent s'ajouter pour mieux cerner cette notion. On note souvent qu'une ontologie définit un vocabulaire partagé pour aboutir à une compréhension commune d'un domaine donné. Effectivement Dans un sens large, on peut adopter pour la notion d'ontologie la caractérisation suivante [USC98] :

“. . . An ontology may take a variety of forms, but will necessarily include a vocabulary of terms and some specification of their meaning (i.e. definitions)”.

Cette définition met en évidence les différentes formes que l'ontologie peut prendre. Mais souligne aussi que cette dernière, doit inclure nécessairement un vocabulaire de termes et une spécification de leur signification. L'ontologie inclut des définitions et une indication

de la façon dont les concepts sont reliés entre eux, les liens imposant collectivement une structure sur le domaine et contraignant les interprétations possibles des termes.

Exemples de formes d'ontologies : – glossaires – terminologies – thesaurus – Lexicons.

2.3. Pourquoi les ontologies ?

Et quelles sont les raisons de développer des ontologies ?

Une issue pour offrir des structures d'un espace commun, partagé, nécessaire pour une véritable intégration sémantique des sources d'informations. Une description précise d'un univers du discours et affirmé dans une langue utilisée pour le raisonnement. [CHR09]

En réponse, nous dirons que l'une des motivations principales du développement d'ontologie est de communiquer et partager de la compréhension commune d'un domaine entre des humains ou des systèmes. Plus précisément, c'est une solution pour fournir des bases communes d'un domaine, partagées, nécessaires pour une réelle assimilation sémantique des sources d'information comme une description précise d'un monde de discours.

Ce type d'ontologies est considéré comme étant formel. Les ontologies formelles sont, en général, explicites et manipulables par la machine, ce qui permet de gérer la sémantique et l'inférence à travers un moteur de raisonnement [DIB04]. Le développement des ontologies a d'autres objectifs, citons ici quelques uns que nous considérons les plus importantes [Mer09] :

- Les ontologies offre la possibilité de réutilisation du savoir sur un domaine : cette caractéristique était parmi l'une des causes principales qui ont manœuvré la recherche sur les ontologies ces dernières années.
- Expliciter ce qui est considéré comme implicite sur un domaine ; Les spécifications formels du savoir sur un domaine sont, de surcroît, utiles pour les nouveaux utilisateurs qui doivent apprendre seulement la signification des termes du domaine.
- Distinguer le savoir sur un domaine du savoir opérationnel est une autre finalité des ontologies. La tâche de configuration d'un produit final à partir de ses constituants, peut être décrit en considérant les spécifications essentielles et faire un programme qui réalisera cette configuration indépendamment des produits et de leurs composants [WRI98].

- Possibilité d'analyser le savoir sur un domaine dès que la spécification des termes du domaine est faite. L'analyse formelle des termes est extrêmement précieuse aussi bien pour la réutilisation des ontologies existantes, que pour leur extension [FIK00]. De façon plus générale, pour [Usc96], les avantages à tirer d'une ontologie peuvent être réparties en trois catégories :
 - Communication : une fois les termes fixés, et le vocabulaire et la grammaire bien définie, elles permettent de répondre au besoin de communication entre personnes-personnes, personnes- systèmes et systèmes-systèmes ;
 - Interopérabilité : Une seule connaissance commune et partagée, pour la compréhension des concepts, elles permettent de faciliter l'interopérabilité des systèmes et la réutilisation des sources de connaissances ;
 - Amélioration logicielle : au niveau de la spécification (compréhension partagée des problèmes), fiabilité (tests de consistance [semi-automatique]) et de la réutilisabilité (fixe la structure des connaissances).

2.4. La représentation des connaissances et les ontologies

Partant d'une pensée, enrichie par une idée conceptuelle et élaborer dans des laboratoires, les ontologies sont considérées comme des modules logiciels s'insérant dans les systèmes d'information et leur apportant une dimension sémantique qui leur faisait défaut auparavant. [FUR02]

On ne peut parler d'ontologie, sans évoquer la représentation des connaissances, puisque l'ontologie est le fruit de cette représentation au sein des systèmes informatiques. Si nous reprenons la définition de T. Gruber 93, « ... une spécification explicite d'une conceptualisation ... », nous remarquons que dans le processus de construction d'une ontologie, on doit tout d'abord réaliser la première phase qu'est la conceptualisation. Ce travail nécessite une identification, dans un corpus, les connaissances spécifiques au domaine de connaissances à représenter. « A conceptualisation is an abstract, simplified view of the world that we wish to represent for some purpose » [GRU 93].

Les ontologies sont formelles car ils sont logiquement exprimées, et partielles car une conceptualisation ne peut pas toujours être entièrement formalisée dans un cadre logique, du fait d'ambiguïtés ou du fait qu'aucune représentation de leur sémantique n'existe dans le langage de représentation d'ontologies choisi. (Revoir la définition de Guarrino 95 plus haut)

J. NOBÉCOURT présente dans [NOB 00] que les formalismes opérationnels présentent une faible tolérance d'interprétation, ce qui oblige à passer directement d'une ontologie informelle à une ontologie totalement formelle et non-ambiguë. De plus, on doit souvent modifier la base de connaissances¹ au cours de son élaboration, ce qui entraîne des incohérences et des modifications plus larges. [FUR02]

Ainsi, pour minimiser cette incohérence, il est nécessaire d'élaborer des modèles semi-formels, partiellement cohérents, correspondant à une conceptualisation semi-formelle. C'est ce qu'on appelle l'ontologie conceptuelle semi-formelles, et le processus de spécifications de en question est appelé ontologisation [KAS00].

Une ontologie doit représenter des connaissances formelles et opérationnelles pour son utilisation dans des systèmes. En effet, une ontologie n'est pas opérationnelle, dans la mesure où elle ne fait pas référence à des mécanismes de raisonnement. Le langage cible doit donc permettre de représenter les différents types de connaissances (connaissances terminologiques, faits, règles et contraintes) et de manipuler ces connaissances à travers des mécanismes adaptés à l'objectif opérationnel du système conçu. Ce processus de traduction est appelé opérationnalisation. [FUR02]

FRÉDÉRIC FÜRST dans [FUR02] découpe le processus général de représentation des connaissances en 3 phases :

La conceptualisation : identification des connaissances issue d'un corpus représentatif du domaine. Ce travail doit être mené par un expert du domaine, assisté par un ingénieur de la connaissance ;

L'ontologisation : formalisation, autant que possible, du modèle conceptuel obtenu à l'étape précédente. Ce travail doit être mené par l'ingénieur de la connaissance, assisté de l'expert du domaine ;

L'opérationnalisation : transcription de l'ontologie dans un langage formel et opérationnel de représentation de connaissances. Ce travail doit être mené par l'ingénieur de la connaissance. En fin, on peut définir l'ontologie en Ingénierie des Connaissances comme un ensemble d'objets identifiés et reconnus comme existant dans le domaine.

1 - Une base de connaissances contient les connaissances utilisées dans un Système à Base de Connaissances. Ces connaissances sont formalisées et des mécanismes permettent de gérer la base pour consulter des connaissances ou en ajouter.

2.5. Les constituants d'une ontologie

2.5.1. Les connaissances et domaines de connaissance

En ingénierie des Connaissances, le terme de connaissance a un sens limité : ne sont considérées que les connaissances (au sens large) susceptibles d'être formalisées, c'est-à-dire les connaissances peu ou amplement techniques [BAC00] : « For knowledge-based systems, what exists is exactly that which can be represented » [GRU 93].

On ne peut manipuler automatiquement des connaissances que si leurs sens est largement consensuel. Plus précisément, on peut considérer qu'il n'y a connaissance que si l'information présente dans la machine prend un sens pour l'utilisateur, c'est-à-dire qu'il peut établir un lien entre cette information et celles qu'il possède déjà, et ce sens doit être le même pour tous les utilisateurs [CHA01]. C'est dans cette théorie que naissent les modèles de type réseaux sémantiques qui restent une conséquence de ce lien sémantique.

En conclusion, la construction d'une ontologie doit se faire à partir d'un champ de connaissances bien délimité par un objectif opérationnel clair, et portant sur des connaissances objectives dont la sémantique puisse être exprimée rigoureusement et formellement. Partant de là, plusieurs types d'ontologies peuvent être distingués en fonction des différents objectifs opérationnels recensés. En analysons les définitions de J. Charlet, B. Bachimont et R. Troncy, trois caractéristiques principales, nous permettent de préciser les constituants d'une ontologie en tant qu'objet informatique :

- Les concepts
- Les propriétés
- Les relations

2.5.2. Les concepts et les relations

a. Concepts

Plusieurs définitions ont été données pour ce terme, par exemple : « *Un concept est une représentation générale et abstraite d'une réalité. Le terme concept vient du participe passé latin «conceptus» du verbe «concipere», qui signifie « contenir entièrement», «former en soi»¹.*

"Le *concept* du temps."

"Le *concept* de l'espace."

1 - fr.wikipedia.org/wiki/Concept

Un concept est défini comme étant une notion généralement exprimée par un terme, ou plus généralement par un signe. Il représente un groupe d'objets ou d'entités qui ont en commun un ensemble de caractéristiques et qui nous permettent de les reconnaître comme faisant partie de ce groupe [GAN02].

Dans [REN07], A. Renouf définit plus formellement une ontologie comme étant :

- un ensemble de concepts ;
- un ensemble de relations entre ces concepts ;
- un ensemble d'axiomes (transitivité, réflexivité, symétrie des relations...)

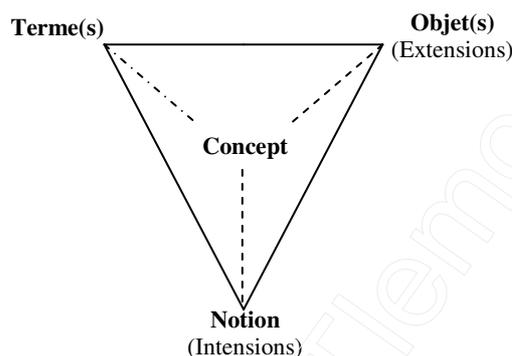


Fig. 15 : Le triangle sémantique (Ogden et Richards, 1923)

Le(s) terme(s) : expriment le concept dans la langue. Le terme en linguistique est un *élément lexical* qui permet d'exprimer le concept en langue naturelle, pouvant admettre des synonymes.

La signification du concept, appelée également « notion » ou « intension » (avec un 's' !) du concept. Notion : également appelée intension du concept, contient *la sémantique du concept*, exprimée en termes de propriétés et attributs, et de contraintes. L'(es) objet(s) dénotés par le concept, appelé(s) également « réalisation » ou « extension » du concept regroupe *les objets manipulés* à travers le concept ; ces objets sont appelés instances du concept. [FUR02]

Par exemple :

Le terme « طاولة » renvoie à la fois à la notion de table comme objet de type « أثاث » possédant « لوحة_مسطحة » et « أعمدة », et à l'ensemble des objets de ce type.

Prenons l'exemple du concept ÉTOILE :

- Son terme est le nom commun « النجم ».
- Sa notion est « نقطة مضيئة في السماء، ليلا »
- Son extension est justement : « جميع النقاط المضيئة في السماء، ليلا » que l'on peut découvrir dans un ciel nocturne ... dégagé !

Soulignons qu'on peut trouver des concepts sans extension, ce sont les concepts génériques comme par exemple le concept « الحقيقة » qui a le sens de « كل ما هو صحيح », cette notion est abstraite. [FAL01] souligne l'existence des concepts partageant la même extension mais pas leur intension et portent le même terme. Ceci correspond à *des points de vue* différents sur un même concept. Mais l'exception ne fait pas la règle, l'exemple du concept de « طاولة », montre que cette notion ne peut se définir qu'en utilisant d'autres concepts comme « أثاث », « لوحة_مسطحة » et « أعمدة » et à l'ensemble des objets de ce type. [FUR02]

- * La subsomption

La subsomption désigne une relation hiérarchique, entre des concepts.

En logique classique, la subsomption est proche de la relation « est impliqué par », ou encore « contient » en logique ensembliste.

Par exemple le concept « الإنسان » subsume le concept « رجل ».

En général, un concept C1 *subsume* un concept C2 si toute propriété sémantique de C1 est aussi une propriété sémantique de C2, c'est-à-dire que C2 est plus spécifique que C1. L'extension d'un concept est forcément plus réduite que celle d'un concept qui le subsume. Son intension est par contre plus riche.

- * La généralité

Un concept est générique s'il n'admet pas d'extension.

Exemple : « الحقيقة » est un concept générique.

- * L'identité

En 1994, N. GUARINO propose cette propriété et souligne qu'un concept porte une *identité* si cette propriété permet de conclure quant à l'identité de deux instances de ce concept. Cette propriété peut porter sur des attributs du concept ou sur d'autres concepts.

Exemple : le concept « تلميذ » porte une propriété d'identité liée au numéro d'élève, deux élèves étant identiques s'ils ont le même numéro.

- * La rigidité

Proposé par N. GUARINO en 1994, un concept est rigide si toute extension de concept en reste extension dans toutes les connaissances du monde possibles.

Exemple 1 : « الإنسان » est un concept rigide, « تلميذ » est un concept non rigide ;

Exemple 2 (Voir fig.12) : « Etre vivant » est un concept rigide. Mais « Humain » ne l'est pas. Car « Humain » est une instance de « Etre vivant ».

L'anti-rigidité : Un concept est anti-rigide si toute instance de ce concept est essentiellement définie par son appartenance à l'extension d'un autre concept.

Exemple 3 : « تلميذ » est un concept anti-rigide car « تلميذ » est avant tout « الإنسان ».

* L'unité

N. GUARINO (1994), un concept est un concept unité si, pour chacune de ses instances, les différentes parties de l'instance sont liées par une relation qui ne lie pas d'autres instances de concepts. Par exemple : les deux parties d'un couteau « السكين », « المقبض » et « النصل » sont liées par une relation « توصيل » qui ne lie que cette « النصل » et ce « المقبض ».

Remarque :

Pour la suite de notre étude, on emploie, tout au long du manuscrit, le mot concept pour désigner l'intension de concept et le mot instance pour désigner un élément de l'ensemble constituant l'extension de concept.

Propriétés :

- * L'équivalence : Deux concepts sont équivalents s'ils ont la même extension;
- * La disjonction (l'incompatibilité) : Deux concepts sont disjoints si leurs extensions sont disjointes. *Exemple* : « رجل » et « امرأة » ; « الليل » et « النهار ».
- * La dépendance : Un concept C1 est dépendant d'un concept C2, si pour toute instance de C1, il existe une instance de C2 qui ne soit ni partie ni constituant de l'instance de C2. *Exemple* : « الأب » est un concept dépendant de « الإبن » et inversement.

b. Relations

Les concepts peuvent être reliés entre eux à l'aide des propriétés décrites ci-dessus, mais il existe d'autres liens représentés par des relations autonomes [FUR02]. Par exemple, la relation « يكتب » lie une instance du concept « إنسان » et une instance du concept « نص », dans cet ordre.

Les concepts (respectivement instance) peuvent être reliés entre eux par des relations au sein d'une ontologie. Une relation est définie comme une notion de lien entre les entités, souvent exprimé par un terme ou un symbole littéral ou autre. En général, ces liens sont classés en deux catégories : *les liens hiérarchiques et les liens sémantiques.*

→ La structure hiérarchique : reprend la structure d'Hyperonymie/Hyponymie. Elle lie un élément supérieur, dit l'hyperonyme, et un élément inférieur, dit l'élément hyponyme, ayant les mêmes propriétés que le premier élément avec au moins une en plus. Dans certains cas, le couple (*Hyperonymie, Hyponymie*) s'interprète par (*Type, Sous-Type*). L'hyperonyme englobe l'hyponyme. On pourra alors écrire «HYPONYME est une sorte de HYPERONYME ».

Exemple :

« الأب » est une sorte « الإنسان »

donc,

« الإنسان » est l'hyperonyme de « الأب ».

→ La relation sémantique : c'est liaison entre les concepts à travers un lien, appelée souvent « Partie-de » ou « Partie_Tout », ce qui correspond à la structuration de HOLONYMIE/MERONYMIE. La relation « partie-Tout » est différente de celle d'hyponymie par le fait qu'un hyperonyme impose ses propriétés à ses hyponymes, par contre le TOUT dispose des propriétés qui ne sont pas obligatoirement transmises à ses parties.

Exemple : Dans le corps humain, « الرأس » et « الساق » font partie du « الجسم » mais elles ne disposent pas des mêmes propriétés. « الرأس » n'est pas une sorte de « الجسم » (voir la Figure 16).

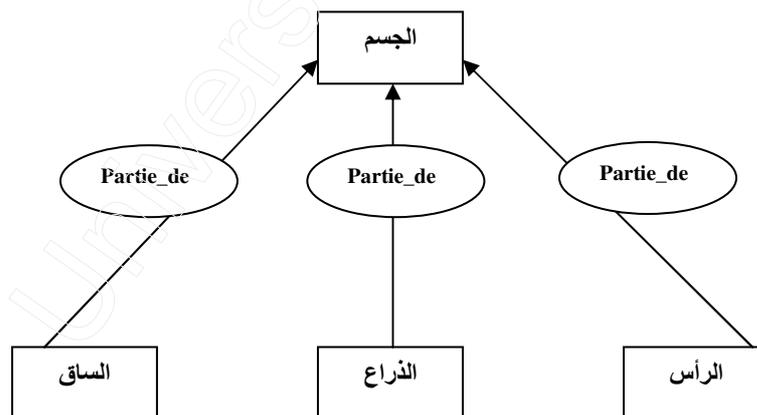


Fig. 16 : Exemple de la relation « Partie-de »

Notons aussi, que les relations peuvent avoir des propriétés. Essentiellement ces dernières peuvent être algébriques (symétrie, réflexivité, transitivité). Mais aussi des propriétés de cardinalité, en général, ces relations sont binaires.

Comme par exemple :

← "الحاسوب" - "يحتوي_على_الأقل" - "قرص_صلب"
 ← "المنزل" - "يحتوي_على_الأقل" - "باب"
 ← "إنسان" - "عنده ما بين صفر واثنان" - "ساق"

- *Les propriétés liant deux relations*

- L'incompatibilité : Deux relations sont incompatibles si elles ne peuvent lier les mêmes instances de concepts.

Exemple : les relations « تكون حمراء » et « تكون خضراء » sont incompatibles ;

- L'inverse : Deux relations binaires sont inverses l'une de l'autre si, quand l'une lie deux instances I1 et I2, l'autre lie I2 et I1.

Exemple : les relations « أب » et « ابن » sont inverses l'une de l'autre ;

- L'exclusivité : Deux relations sont exclusives si, quand l'une lie des instances de concepts, l'autre ne lie pas ces instances, et vice-versa. L'exclusivité entraîne l'incompatibilité.

Exemple : « يحتوي_على » et « لا يحتوي_على » sont exclusives.

2.6. Les formalismes de représentation [GAE02]

Une ontologie, a besoin d'être représentée formellement et explicite (GARINO 1995). Plus encore, une ontologie doit représenter l'aspect sémantique des relations liant les concepts. A cet effet, de nombreux formalismes ont été développés pour représenter les connaissances, de la logique des prédicats jusqu'aux langages sophistiqués basés sur des structures de données. [GAE02]

2.6.1. Les formalismes logiques

Dans ce formalisme, Une base de connaissances rassemble un ensemble d'axiomes décrivant une situation, sur lesquels des règles d'inférence opèrent et produisent de nouvelles formules valides. Celles-ci constituent alors de nouveaux états de choses dans la base. Le langage de programmation Prolog en est un exemple.

2.6.2. Les réseaux sémantiques

En I.A., hommage à QUILLIAN qui fut le premier à développer de tels réseaux en tant que modèles de la mémoire associative humaine. Un réseau sémantique est un modèle de représentation, sous forme de graphe, du contenu sémantique des concepts. Les nœuds du graphe représentent des objets (concepts, situations, événements, etc.) et les arcs expriment des relations entre ces objets. Ces relations peuvent être des liens " sorte - de " exprimant la relation d'inclusion ou des liens " est-un " représentant la relation d'appartenance.

Exemple : on peut dire que : Volkswagen est une marque de voiture (voir la figure 17).

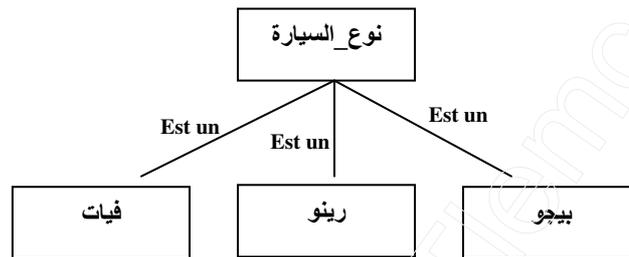


Fig. 17 : Exemple de réseau sémantique utilisant la relation " Est-un "

Exemple : On peut dire que « الصقر » <est un> « طائر_الجوارح » qui <est_une_sorte_de> « طيور » (Voir la figure 18).



Fig. 18 : Exemple de réseau sémantique utilisant la relation " sorte-de "

Remarque :

« Une ontologie peut être considérée comme un réseau sémantique, puisqu'elle regroupe un ensemble de concepts décrivant complètement un domaine, reliés les uns aux autres par des relations, taxonomiques (hiérarchisation des concepts) et sémantiques. »

Parmi les réseaux sémantiques, les plus répandus pour la conceptualisation des ontologies, on trouve *les graphes conceptuels* dont le but fondamental est d'être " un système de logique hautement expressif, permettant une correspondance directe avec la langue naturelle " [SOW92]. Ils sont basés sur la logique pour la représentation des connaissances.

Note : Actuellement, la théorie des graphes conceptuels de [SOWA84], représentant les relations sémantiques, constitue le formalisme le plus répandu pour conceptualiser les ontologies.

Mis au point par John F. Sowa, pour modéliser une ontologie de haut niveau [SOW00]. Un graphe conceptuel est un graphe étiqueté, biparti, connexe et fini. Les sommets représentent les entités, attributs, états ou évènements. Chaque sommet est typé. Ces types sont ordonnés dans une structure de treillis orienté du plus spécifique au plus général avec des relations "sorte-de ". Le langage CGIF (Conceptual Graph Interchange Form) a été développé pour définir des graphes conceptuels. Exemple d'ontologie « Top-Level » de Sowa, treillis :

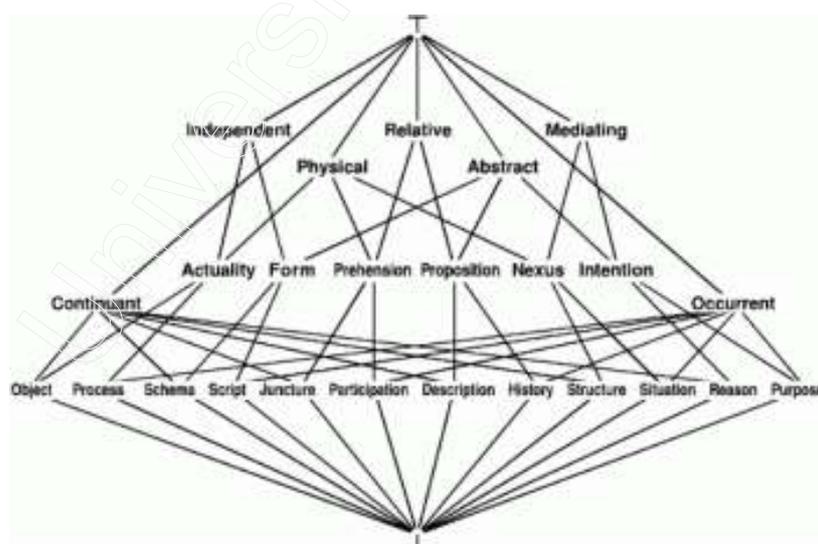
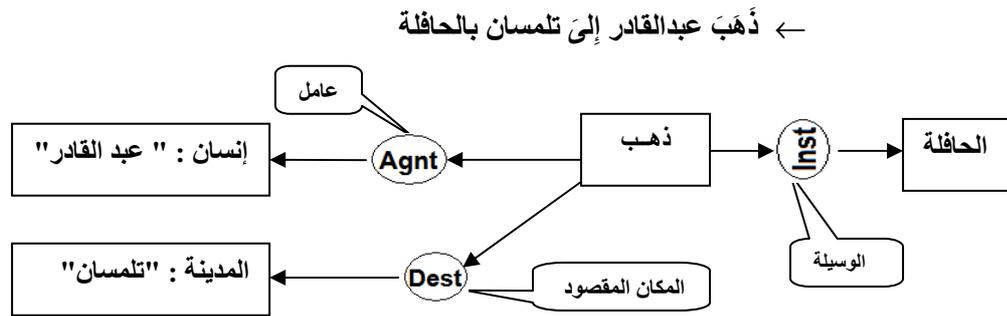


Fig. 19 : Hiérarchie du graphe de Sowa (Top-Level) : Treillis

Exemple : Graphe conceptuel à quatre concepts de la phrase :



2.6.3. Les schémas (Frame)

Les schémas, connus beaucoup plus sous son terme anglais « Frame », sont apparue en 1932 dans les études psychologiques. Plus tard, les schémas ont été introduits en intelligence artificielle par **Minsky** afin de résoudre les problèmes de la vision par ordinateur. Parmi les langages de représentation des connaissances à base de frame : KRL (1977) et KL-One (1985). Les frames ont été définies par MINSKY [MIN75] par: "A frame is a data-structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed. We can think of a frame as a network of nodes and relations. The "top levels" of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals - "slots" that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet (assignments themselves are usually smaller "sub-frames"). Simple conditions are specified by markers that might require a terminal assignment to be a person, an object of sufficient value, or a pointer to a sub-frame of a certain type. More complex conditions can specify relations among the things to several terminals".

Ce qui se traduit par :

« Un frame est une structure de données représentant une situation stéréotypée, comme se trouver dans un certain type de salon ou se rendre à un fête d'anniversaire d'un enfant. Divers types d'informations sont associés à chaque frame. Certaines d'entre elles concernent l'utilisation de ce frame. D'autres portent sur ce que l'on s'attend à ce qu'il arrive par la suite. D'autres encore portent sur ce qu'il faut faire si ces attentes ne sont pas confirmées [...]. »

Donc on peut dire qu'un schéma est une structure de données complexe décrivant une situation ou un objet standard. Un schéma, comme le montre la figure 20, est caractérisé par des attributs, des facettes et des relations.

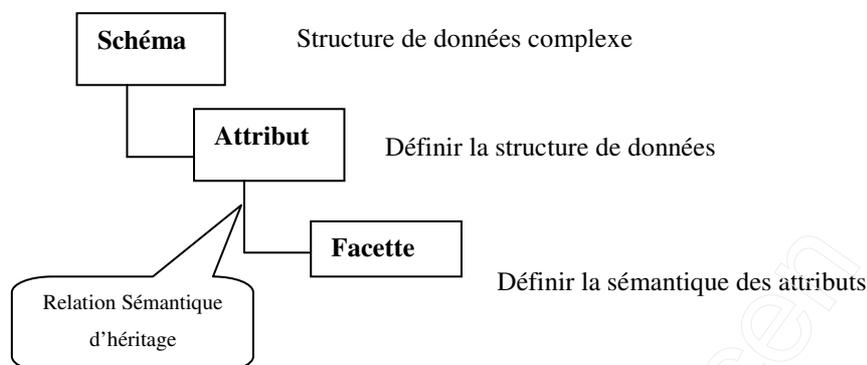


Fig. 20 : Eléments caractérisant un Schéma [MEL07]

2.6.4. Les scripts

La notion de " script " (ou scénario) a été introduite par Schank et Abel, sur le modèle des schémas pour le traitement du langage naturel. Ils ont défini un script par : " *A script is a structure that describes appropriate sequences of events in a particular context. A script is made up of slots and requirements about what can fill those slots. The structure is an interconnected whole, and what is in one slot affects what can be in another. Scripts handle stylized everyday situations. They are not subject to much change, nor do they provide the apparatus for handling totally novel situations. Thus, a script is a predetermined, stereotyped sequence of actions that defines a well-known situation. Scripts allow for new references to objects within them just as if these objects had been previously mentioned; objects within a script may take "the" without explicit introduction because the script has already implicitly introduced them* " [SCH88].

Un script est donc une structure de données regroupant des connaissances relatives à une situation et qui permet de combiner des représentations.

3. Construction d'une ontologie

Soulignons qu'on ne peut construire une ontologie généralisant tous les domaines, pour la simple raison, que beaucoup de termes n'ont pas le même sens d'un domaine à un autre [BAC00]. De ce principe on peut dire que pour construire une ontologie il faut la restreindre dans un domaine bien précis. Ainsi, les connaissances traduites par une ontologie sont à véhiculer à l'aide des deux principaux éléments suivants : *Concepts* et *Relations*.

Il existe trois méthodes pour la construction d'une ontologie [MEL07] :

- La Méthode manuelle : Les experts créent une nouvelle ontologie d'un domaine ou développent une ontologie déjà présente, l'ontologie WordNet¹.
- La Méthode automatique : L'ontologie est construite par des techniques d'extraction des connaissances : Des concepts et leurs relations sont extraits des bases de connaissances et ensuite vérifiés par les inférences.
- La Méthode mixte (Semi-automatique) : Les ontologies sont construites par des techniques automatiques mais elles permettent d'étendre des ontologies qui ont été construites manuellement comme la base des connaissances (Cyc²).

3.1. Le cycle de vie des ontologies

Le cycle de vie de l'ontologie permet d'identifier les différentes étapes de construction d'une ontologie. Uschold et Grüninger dans [USC96], ont mis au point une méthode, qui d'ailleurs est la plus connue, pour la création d'ontologie.

La méthode est générique, c'est pourquoi, ses étapes sont considérées comme la base d'un processus standard de construction. On remarque quatre phases fondamentales (Voir la figure 21) :

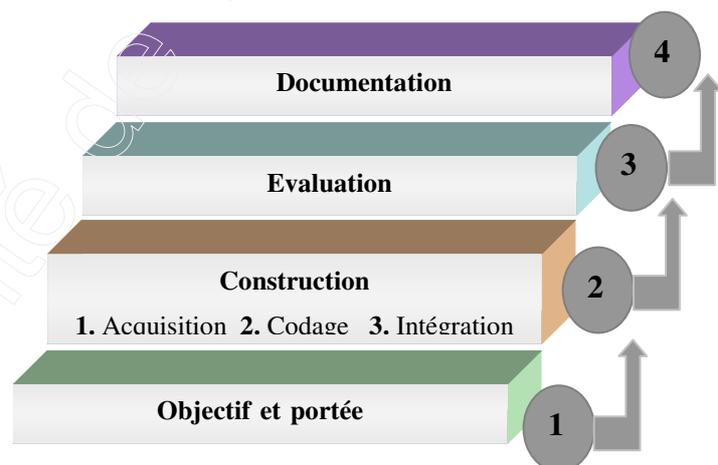


Fig. 21 : Etapes pour la construction des ontologies [MEL07]

Phase 1 : Objectif et portée : Identifier d'une façon générale l'objectif, la portée et les limitations de l'ontologie à construire.

Phase 2 : Construction : Cette étape prend un temps considérable et reste la plus difficile à réaliser, elle comporte :

1 - <http://wordnet.princeton.edu/>

2 - <http://www.opencyc.org>

- L'acquisition des connaissances :

Définir les concepts et les relations entre eux, sans ambiguïtés. Parmi les formes d'acquisition de connaissances qui existent, nous soulignons la forme d'analyse informelle de texte pour définir les concepts généraux ou bien sous forme d'une analyse formelle afin de définir les structures des connaissances [GOM97].

- Le codage :

Après la formalisation des concepts et leurs relations acquises, le codage permet de représenter l'ontologie dans un langage formel. La formalisation de l'ontologie peut être de différents degrés [BRI04] :

Très informel : L'ontologie est exprimée dans le langage naturel ;

Semi-informel : Elle s'exprime dans une forme structurée du langage naturel;

Semi-formel : Elle est exprimée dans un langage artificiel défini formellement;

Rigoureusement formel : l'ontologie est exprimée dans un langage formel utilisant une sémantique formelle avec des théorèmes et preuves.

- L'intégration :

Permettre la réutilisation des concepts déjà définis dans des ontologies existantes.

Phase 3 : Evaluation : Grüber 1993, repris par (Corcho et al. 2002) a proposé quelques critères pour l'évaluation d'une ontologie :

- La clarté : les concepts d'une ontologie doivent fournir le sens attendu de termes;
- La cohérence : l'ontologie ne doit permettre que les inférences qui sont en accord avec les définitions afin de ne pas créer de contradictions.
- L'extensibilité : Réutilisation de l'ontologie, sans remettre en cause le vocabulaire précédemment conçu ;
- Le biais d'encodage minimum¹ (minimal encoding bias) : la spécification de l'ontologie doit être aussi indépendante que possible du choix d'une représentation.

1 - On parle aussi de déformation d'encodage minimale : une déformation d'encodage apparaît lorsque les choix d'une représentation sont faits uniquement pour une commodité d'implémentation. Ceci doit être évité car les agents partageant la connaissance peuvent être implémentés dans différents systèmes de représentation et styles de représentation.

- L'engagement ontologique minimal (minimal ontological commitment) : l'objectif est de permettre la spécialisation des spécifications d'une ontologie donnée selon des besoins réels ; c'est-à-dire l'expressivité maximum de chaque terme.

Phase 4 : Documentation : permet de renseigner les ontologies, leurs concepts importants ainsi que leurs objectifs. Ontolingua représente l'un des éditeurs le plus performant qui peut nous aider à réaliser cette étape en mettant à notre disposition des documentations formelles et informelles.

Par la suite, R.DIENG et al. dans [DIE01], affinent les étapes d'Uschold et Grüninger, et proposent un cycle de vie inspiré du génie logiciel, Il comprend une étape initiale d'évaluation des besoins, une étape de construction, une étape de diffusion, et une étape d'utilisation. Après chaque utilisation significative, l'ontologie et les besoins sont réévalués et l'ontologie peut être étendue et, si nécessaire, en partie reconstruite.

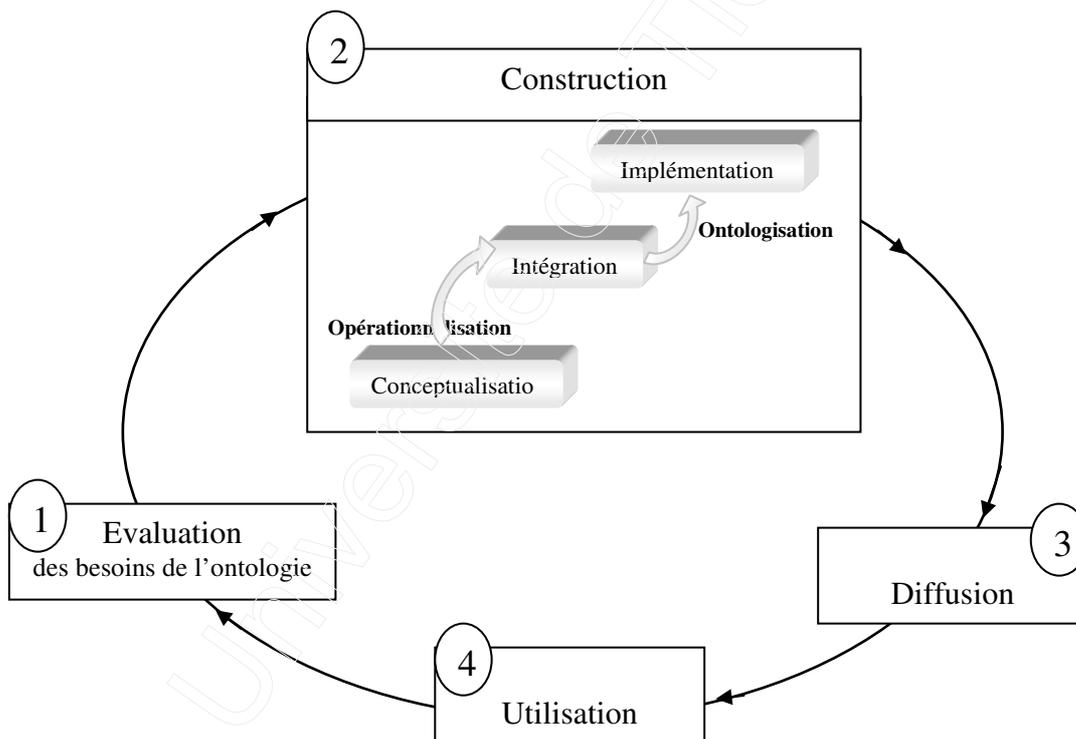


Fig. 22 : Le cycle de vie d'une ontologie [FUR02]

1 - Une ontologie doit faire aussi peu d'affirmations que possible sur ce qui a été modélisé afin de pouvoir instancier et spécialiser l'ontologie suivant les besoins.

3.1.1. Evaluation des besoins

Le cycle Commence par la phase d'évaluation des besoins, la finalité observée par la construction d'une ontologie se résume en 3 points [USC 95] :

L'objectif opérationnel : il est essentiel de souligner l'usage opérationnel de l'ontologie à travers des diagrammes de cas d'utilisation ;

- Le domaine de connaissance : il faut le cerner aussi clairement que possible, et le découper si nécessaire en termes de connaissances du domaine, connaissances de raisonnement, connaissances de haut niveau (communes à plusieurs domaines) ;
- Les utilisateurs : doivent être ciblés et identifiés autant que possible, pour ne pas s'éloigner de l'objectif opérationnel et de la granularité du formalisme choisie qui indique les propriétés d'intérêt pour l'utilisateur.

Après cette phase, les besoins sont définis, ainsi l'étape de la construction de l'ontologie peut démarrer, en commençant par la phase de conceptualisation.

3.1.2. Conceptualisation [FUR02]

Cette étape, permet de faire ressortir, à partir des données brutes, d'un corpus¹ l'ensemble des concepts et les relations entre eux décrivant ainsi les connaissances informelles.

Pour commencer, la première tâche est de trier les connaissances se trouvant dans le corpus spécifique (séparer les termes spécifique du domaine de ceux qui sont présents juste pour l'expression du domaine). La seconde tâche vient ensuite, pour spécifier ces concepts, ces relations, les propriétés des concepts et des relations, les règles et les contraintes, etc. c'est à-dire, tout l'aspect conceptuel. Le choix d'usage de l'ontologie peut donc être indiqué dans cette étape.

Notons qu'un texte ou corpus, n'a un sens que lorsqu'il est lu par un expert² [BAC99]. C'est pour cela que ce travail doit être mené par un expert du domaine, assisté par un ingénieur de la connaissance.

1 - Le corpus est souvent composé de textes d'auteurs extraits d'œuvres littéraires ou d'ouvrages critiques, etc.

2 - Le terme expert désignant justement ici une personne pour qui le corpus fait sens.

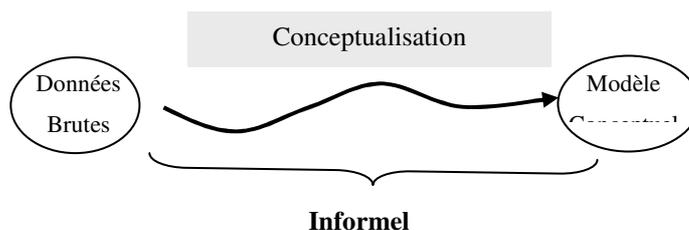


Fig. 23 : Processus de conceptualisation

En achevant cette étape, on obtient un ensemble de termes désignant les entités d'un domaine de connaissances (relatif au corpus). Ainsi, la conceptualisation est couronnée par un modèle informel, donc sémantiquement ambiguë et exprimé en langage naturel.

La fouille des connaissances d'un domaine peut se reposer à la fois sur l'analyse de documents et sur l'interview d'experts du domaine. Ces activités doivent être raffinées au fur et à mesure que la conceptualisation se dévoile. [FUR02]

3.1.3. Ontologisation

La deuxième phase est l'*ontologisation*. C'est une formalisation partielle de connaissances, sans perte d'information (formalisme générique avec une sémantique claire), du modèle conceptuel obtenu dans l'étape de conceptualisation.

Cette formalisation partielle va faciliter la représentation de l'ontologie dans un langage totalement formel et opérationnel. [FUR02]

Mais, essayer de préserver toutes les connaissances du domaine, nous amène à ajouter, à l'ontologie, des connaissances que nous ne pouvons formalisées c'est-à-dire dont la sémantique est ambiguë. Cependant, le modèle ainsi obtenu est *semi-formel*.

Une ontologie semi-formelle n'est pas autorisée à être utilisée dans un système de base de connaissances. En revanche, une ontologie, contenant toutes les connaissances d'un domaine, constitue le support idéal de communication et de partage des connaissances.

Le résultat final d'une ontologisation est une hiérarchie de concepts, de relations, mais aussi d'attributs des concepts.

USCHOLD (1996) conseille de construire ces hiérarchies de bas en haut, pour donner ainsi la priorité aux concepts de bas niveau réellement utilisés dans le domaine, par rapport aux concepts qui ne sont souvent qu'ajouter artificiellement pour bâtir la hiérarchie.

Par exemple, pour construire une ontologie en rapport avec « الهندسة », les concepts de « النقطة » et « خط » vont, entre autres, être pris en compte, mais le concept

« مجموعة من النقط » peut être ajouté, concept qui subsume « النقطة » et « خط », et ceci uniquement pour structurer l'ontologie. Cette étape d'ontologisation donne un résultat en deux parties :

- Une partie formelle avec une sémantique claire ou du moins consensuelle ;
- Une partie informelle, avec une sémantique fixée *a priori* et donc exprimée dans un langage naturel ou semi-structuré.

Cette ontologisation est évalué par les critères d'évaluations de T.Grüber (1993), mentionné plus haut à l'étape d'évaluation du model de Uschold et Grüninger.

B. BACHIMONT propose de commencer à bâtir d'abord une ontologie différentielle¹ en organisant les concepts à l'aide des quatre principes suivant [BAC01] :

- Communauté avec le père ou principe de similarité : un concept partage l'intension de son concept père ;
- Différence avec le père ou principe de différence : l'intension d'un concept est différente de celle de son concept père, sinon il n'y aurait pas besoin de définir le concept fils ;
- Communauté avec les frères ou principe de sémantique unique : une propriété est commune aux concepts frères issus du même concept père mais s'exprime différemment pour chaque frère, exemple : Les concepts « رجل » et « امرأة » portent la propriété « جنس » héritée de leur concept père « إنسان », mais cette propriété vaut « ذكر » chez « رجل » et « أنثى » chez « امرأة » ;
- Différence avec les frères ou principe d'opposition : les frères doivent tous être incompatibles, sinon il n'y aurait pas besoin de tous les définir.

3.1.4. Opérationnalisation

Appelée parfois *représentation*, est une transcription de l'ontologie dans un langage formel et opérationnel de représentation de connaissances (*possède une syntaxe et une sémantique*). L'opérationnalisation consiste à donner la main à une machine pour manipuler les connaissances d'une ontologie.

Ce travail doit être mené par l'ingénieur des connaissances pour traduire le model conceptuel structuré, issu de la phase d'ontologisation, dans un langage semi formel de représentation par exemple, les langages à base des frames, le modèle des graphes conceptuels (traitant des opérations telles que la jointure, la projection, etc.) ou les logiques de description (opérations de subsomption). L'ontologie opérationnelle est implantée en machine au sein d'un système manipulant le modèle de connaissances utilisé via le langage opérationnel choisi. A l'étape de diffusion l'ontologie est testée par rapport au contexte d'usage pour lequel

1 - Ontologie différentielle : permet de préciser le sens des concepts (concepts sémantiques) de manière non ambiguë

elle a été construite. Une fois cette étape achevée, l'ontologie, en question, peut être mise à la disposition des utilisateurs. [FUR02]

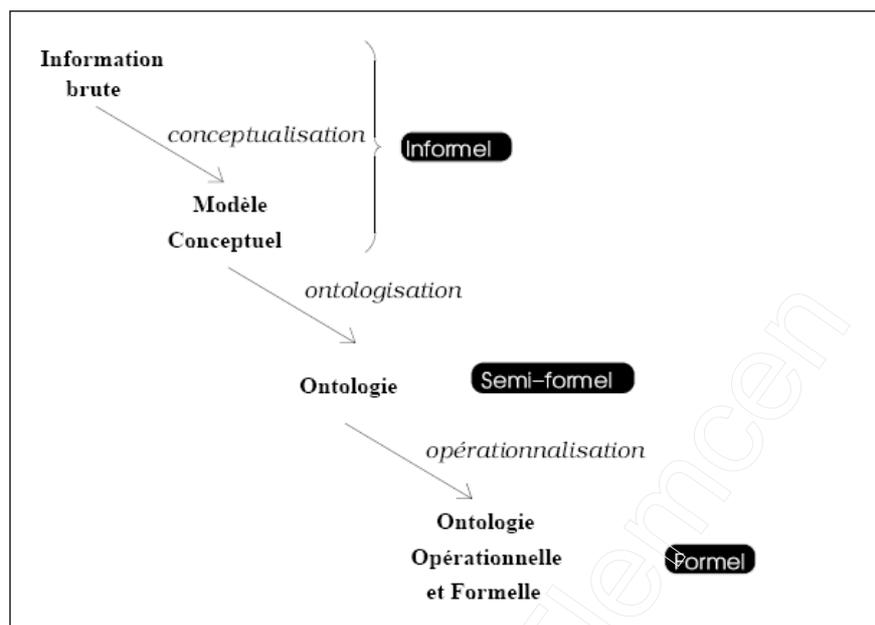


Fig. 24 : Construction d'une ontologie opérationnelle. [FUR02]

3.2. L'évaluation et l'évolution d'une ontologie

M. GRUNINGER et M.S. FOX proposent d'utiliser des questions de compétences permettant de tester l'ontologie. Ainsi cette phase, d'évaluation d'une ontologie se fait a priori par des tests correspondants à l'objectif opérationnel de l'ontologie. [GRU95]

La validation formelle consiste à vérifier l'absence de cycle, c'est-à-dire de définition en boucle, s'il n'y a pas redondance de concepts ou de relations, si chaque hiérarchie est bien connexe¹ [GOM 96].

La validation sémantique permet de contrôler que la structure des hiérarchies est correcte vis-à-vis des principes différentiels utilisés.

Si l'évaluation de l'ontologie a échoué, il faut donc remonter jusqu'à l'étape de conceptualisation, afin d'éviter des modifications qui ne respectent pas la sémantique du domaine, et faire évoluer l'ontologie.

3.3. La fusion des ontologies

Un domaine de recherche encore peu exploré. La fusion consiste de construire une nouvelle ontologie et l'intégrer dans une ontologie déjà existante. Soit un simple alignement

1 - hiérarchie connexe : il n'y a pas de concept ou de relation isolé des autres et donc sans aucun sens

dans le cas où aucune partie n'est commune aux ontologies, soit une véritable fusion [NOY02]. La fusion ne peut se faire que si les deux ontologies présentent des entités conceptuelles (concepts ou relations) communes afin de pouvoir les exprimer et les identifier dans un même formalisme.

3.4. Méthodologie et outils de construction d'ontologies

Beaucoup de méthodologies et d'outils ont été développés dans le but de créer des ontologies. Nous allons citer à titre d'information, que les plus importants c'est-à-dire ceux qui offrent des implémentations de méthodologies, bien qu'ils offrent un processus d'ontologisation mais ils restent beaucoup à faire du côté de conceptualisation.

Méthodologies :

- Approche de construction d'ontologie de domaine à partir de grandes ontologies (SENSUS, Cyc, AKT,...) : Construire des ontologies de domaine à partir des grandes ontologies déjà existantes.
- Méthode de Uschold et King's [USC95] : Basé sur l'expérience acquise lors du développement de l'ontologie d'Entreprise, the enterprise ontology.
- La méthodologie On-To-Knowledge¹ (OTK) : Donc, la méthodologie On-To- Knowledge propose de construire une ontologie en tenant compte de comment l'ontologie va être utilisée par l'application plus tard (cette méthodologie est très dépendantes de l'application).
- METHONTOLOGY : Cette méthodologie développée A.GOMEZ-PEREZ au sein du groupe d'ontologie à l'université polytechnique de Madrid. METHONTOLOGY est né dans les travaux de software development process [IEEE96] et dans knowledge engineering methodologies [GOM97] et [WAT86].

Outils :

- TERMINAE, qui intègre des outils linguistiques, permet la visualisation des résultats des extracteurs de candidats-termes Lexter et/ou Syntex développés par D. Bourigault. Ces concepts doivent ensuite être triés par un expert et organisés hiérarchiquement, puis la sémantique du domaine est précisée à travers des axiomes.
- DOE (DIFFERENTIAL ONTOLOGY EDITOR) 1ère version en 2002, dernière version (v2.0) 2008. DOE n'est pas complet, il est simplement un éditeur

1 - <http://www.ontoknowledge.org>.

d'ontologie qui applique la méthodologie des principes différentiels de Bruno Bachimont.

- ODE (ONTOLOGY DESIGN ENVIRONMENT), construction des ontologies au niveau connaissances (Méthodologie METHONTOLOGY).
- ONTOEDIT (ONTOLOGY EDITOR), est indépendant de tout formalisme. Il permet l'édition des hiérarchies de concepts et de relations et l'expression d'axiomes algébriques portant sur les relations, et de propriétés telles que la généralité d'un concept.
- PROTEGE2000 issu du modèle des frames, c'est un outil non lié à des formalismes de représentation. Il permet aussi l'édition, le contrôle, la visualisation et l'extraction d'ontologie à partir des textes.
- ONTOLINGUA qui constitue en fait une extension du langage KIF (KNOWLEDGE INTERCHANGE FORMAT). Ce serveur d'édition permet la fusion d'ontologie. L'ontologie est immédiatement représentée dans un formalisme.
- OILED (OIL EDITOR) est un éditeur d'ontologie s'inspirant du formalisme OIL. Un éditeur de petites ontologies avec un moteur d'inférence pour tester la cohérence de l'ontologie construite.

Plusieurs méthodes en Ingénierie Ontologique mais encore loin d'être complètes des méthodes du Génie Logiciel. Il reste beaucoup à faire pour unifier toutes ces méthodes. Cette situation justifie les travaux de recherche en cours dans différentes équipes de laboratoires.

L'article, dans [FER99], présente un premier état de l'art, relativement complet, sur ces méthodes de développement. Ainsi que l'article [AUS00] qui exploite des résultats de travaux menés par le groupe français TIA (Terminologie et IA) pour proposer une méthode basée sur l'analyse de corpus.

4. Classification des ontologies

Les ontologies peuvent être de nature très diverses. Afin de mieux s'y retrouver, un certain nombre de classifications ont été proposées.

La plus courante des classifications d'ontologies est la classification selon l'objet de conceptualisation [PSY03]. On peut ainsi distinguer sept catégories [GOM99b] :

- Les ontologies de représentation des connaissances : les ontologies de représentations des connaissances sont utilisées pour formaliser un modèle de représentation des connaissances. On peut par exemple citer l'exemple de l'ontologie de *frame* [GRU93], qui définit les primitives de représentation des langages à base de frames (classes, instances, slots, facettes, etc.).

➤ les ontologies supérieures (aussi appelées ontologies de haut niveau):

Une ontologie de haut niveau décrit des concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc. Ces concepts ne dépendent pas d'un problème ou d'un domaine particulier. Ces ontologies doivent être, du moins en théorie, consensuels à de grandes communautés d'utilisateurs [GUA98]. Des exemples d'ontologies de haut niveau sont « Upper Cyc ».

➤ Les ontologies génériques (méta-ontologie) : Elles contiennent des concepts généralistes, mais moins abstraits que ceux contenus dans les ontologies de haut niveau. On pourra réutiliser l'ontologie dans plusieurs domaines [PSY03]. Un exemple d'une telle ontologie : SUMO (Suggested Upper Merged Ontology), une autre ontologie générique a été développée dans cette classe citons : Wordnet.

➤ Les ontologies de tâches [MIZ03] : Ce type d'ontologie sert à modéliser les tâches d'un problème ou d'une activité donnée. Ce type d'ontologie est utile pour décrire la structure d'une tâche de résolution de problème de manière indépendante du domaine concerné.

➤ Les ontologies de domaine : Elles sont réutilisables à l'intérieur d'un domaine donné et modélisent le vocabulaire à l'intérieur de ce domaine [GOM99b]. La plupart des ontologies existantes sont des ontologies de domaine [PSY03].

➤ Les ontologies de tâches-domaine : ce sont des ontologies de tâches spécifiques à un certain domaine. Un exemple d'une telle ontologie est celui d'une ontologie des termes liés à la planification chirurgicale. [GOM99b]

➤ Les ontologies d'application. Il s'agit du type d'ontologie le plus spécifique [PSY03]. Les concepts que l'on trouve dans ce genre d'ontologies modélisent les concepts d'un domaine particulier dans le cadre d'une application donnée.

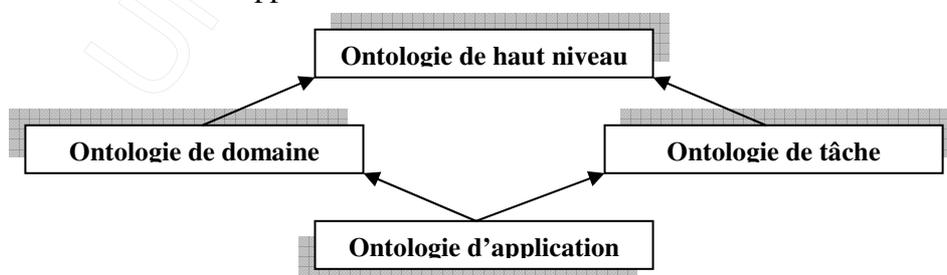


Fig. 25 : Type d'ontologie selon Guarino

On peut aussi classer les ontologies selon le niveau de formalisme du langage que l'on utilise pour les modéliser. Uschold et Gruninger proposent [USC96] à ce niveau quatre types

d'ontologies (Figure 26) :

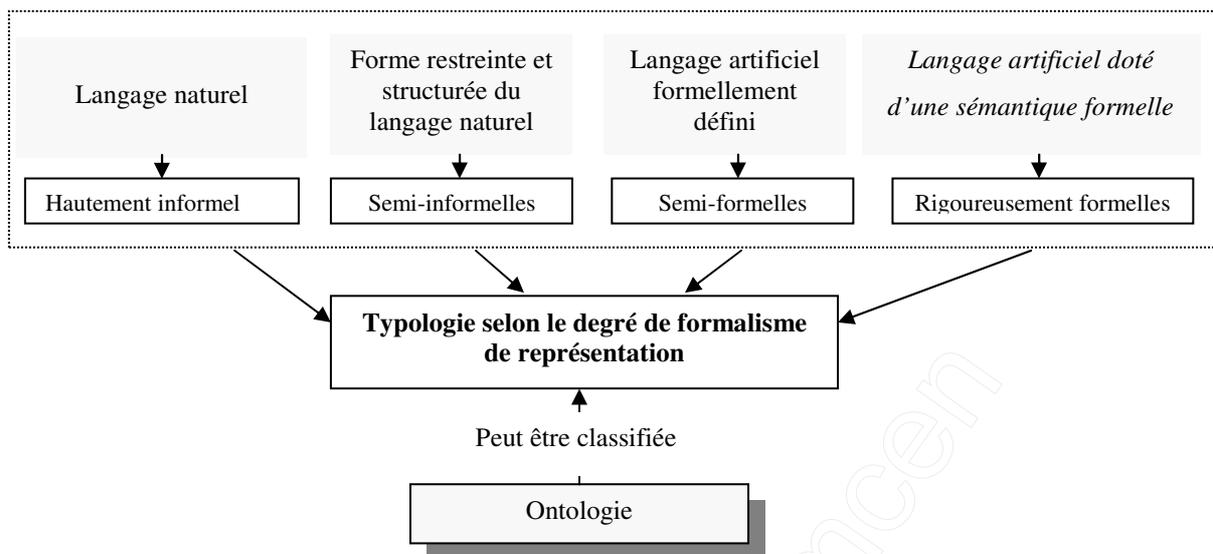


Fig. 26 : Différents types d'ontologies selon le degré de formalité

- Les ontologies informelles, exprimées en langage naturel.
- Les ontologies semi-informelles, écrites dans un langage naturel, mais sous une forme limitée et structurée, permettant d'augmenter la clarté et la lisibilité.
- Les ontologies semi-formelles, exprimées dans un langage artificiel défini de manière formelle.
- Les ontologies strictement formelles, définies elles aussi dans un langage artificiel, mais avec des théorèmes et des preuves sur des propriétés de l'ontologie, telles que la robustesse ou la complétude.

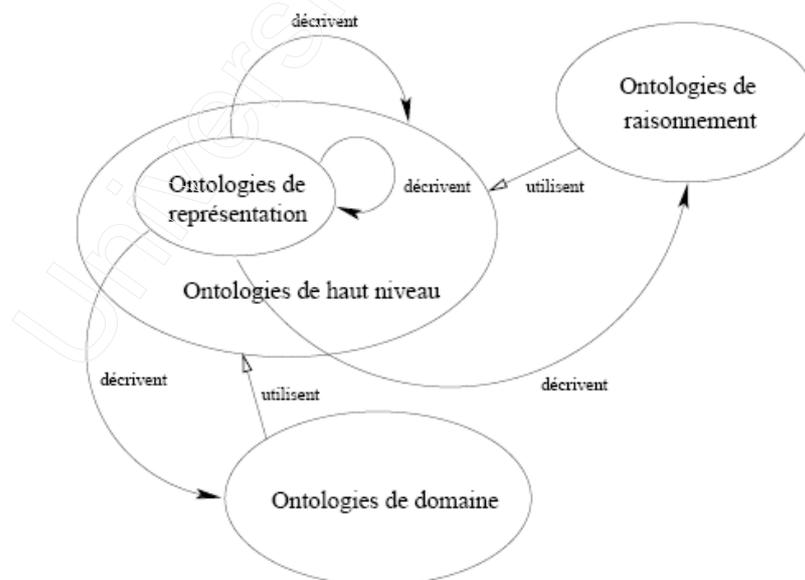


Fig. 27 : Différents types d'ontologies [MIZ 97]

5. Conclusion

Dans ce chapitre, nous avons précisé les différentes définitions de la notion d'ontologie. Nous avons aussi souligné les différents éléments dont elle est constituée et une description des besoins auxquels elle répond, ainsi que les différents domaines d'applications.

Les méthodes et les outils de construction ont été bien détaillés, car ils ont une relation importance avec le thème de notre thèse. La création d'ontologie est expliquée par la représentation d'un cycle de vie d'une ontologie. Le chapitre est achevé par une classification des ontologies.

Le chapitre suivant sera consacré à la description de la base lexicale WordNet, désormais appelée ontologie lexicale, [GRA04] énonce « ... *un lexique peut néanmoins servir à l'élaboration d'une ontologie...* ». Cette dernière sera notre modèle de construction d'ontologie.