

CHAPITRE 1

LE WEB SEMANTIQUE ET LES ONTOLOGIES

1 Introduction

En philosophie, l'ontologie est l'étude de l'être en tant qu'être, c'est-à-dire l'étude des propriétés générales de ce qui existe.

Ce chapitre s'intéresse à la théorie du web sémantique et les ontologies, nous commencerons par définir les limites de web actuel, nous présenterons ensuite le web sémantique et enfin nous terminerons ce chapitre par des théories sur les ontologies.

2 Le web syntaxique

2.1 Historique

Le Word Wide Web (www) a été mis en place par Tim-Berners Lee qui est considéré comme le père fondateur du web. Au milieu des années 1990, Internet fait son apparition au grand public en version 1.0 via des pages statiques codés en HTML.

Il s'agit de sites non interactifs principalement destinés à la recherche d'informations : encyclopédies, etc.

Au début des années 2000, Le web a évolué et est devenu dynamique. Il s'agit de sa version 1.5. Il est maintenant possible de consulter du contenu dynamique en ligne, via des bases de données : boutique en ligne, etc. (A noter que c'est aussi l'avènement des start-up qui surfent sur la vague du web dynamique, accessible à tous) [1].

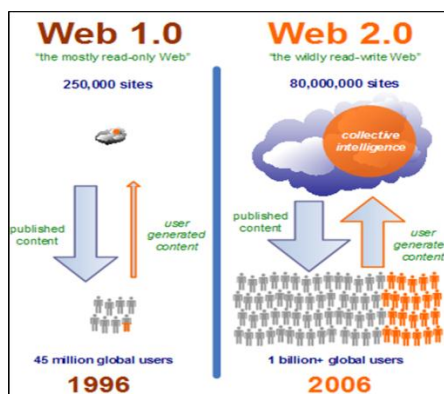


Figure 1.1 web 1.0 vs web 2.0 [2]

En 2004, Dale Dougherty utilise le terme « Web 2.0 » qui sera vite repris par Tim O'Reilly, spécialiste du World Wide Web. Le web 2.0 se caractérise par la prise de pouvoir des internautes sur internet, grâce notamment aux réseaux sociaux. Plus qu'un bouleversement technologique (PHP5, AJAX, etc.), l'apparition du web 2.0 prend une véritable dimension sociologique puisqu'il replace le consommateur à la source de l'information.

Pour Joshua Porter, le web 2.0 « c'est le partage de l'information, fondé sur des bases de données ouvertes qui permettent à d'autres utilisateurs de les employer. » L'internaute est donc enfin devenu le centre d'intérêt d'internet.

Depuis quelques années, on entend de plus en plus parler du Web 3.0, c'est-à-dire un web encore plus humain, encore plus intelligent (citons notamment l'exemple devenu célèbre du frigo intelligent qui fait les courses tout seul selon vos habitudes lorsqu'il est vide).

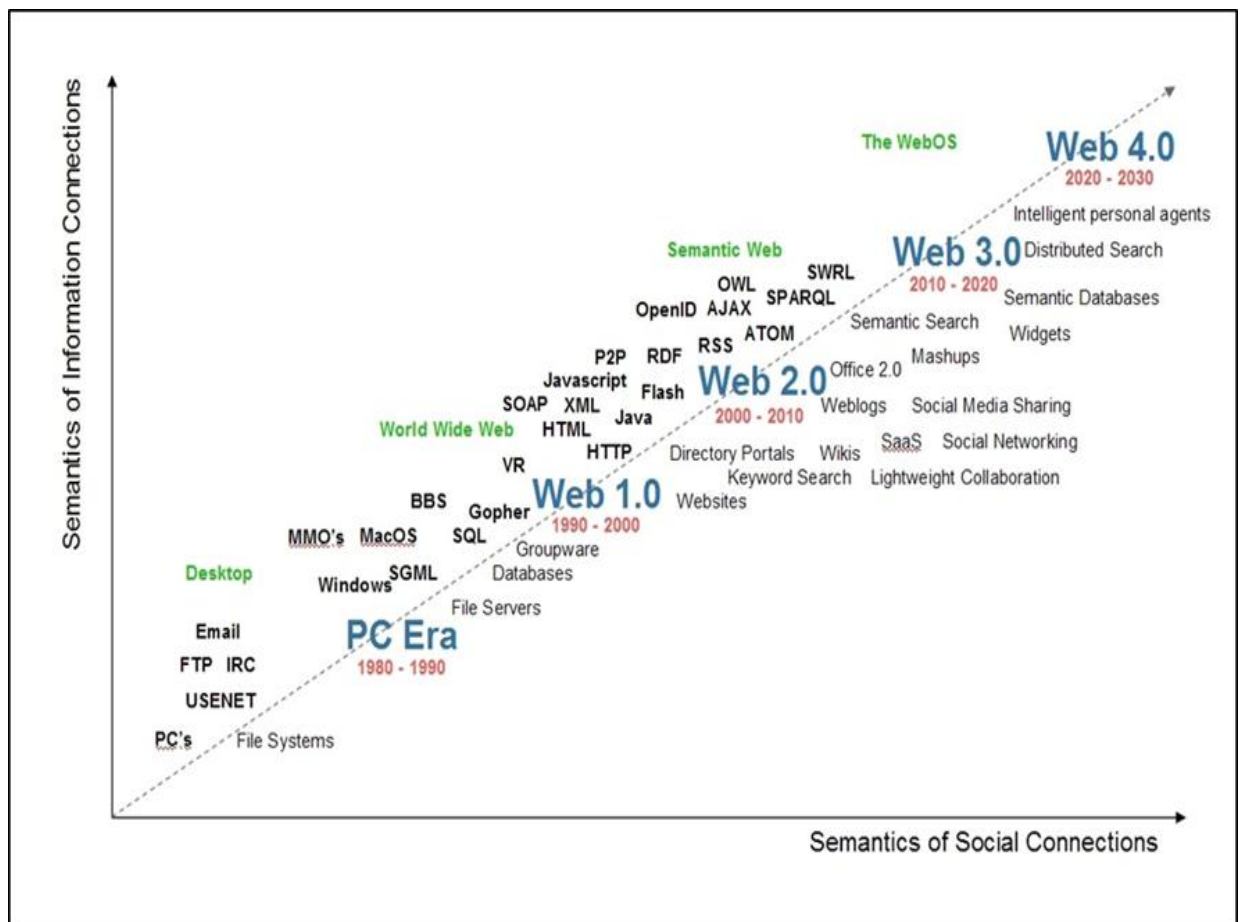


Figure 1.2 l'évolution de web [2]

2.2 Définition de web (World Wide Web)

L'appellation Internet est l'abréviation Interconnections of networks. L'Internet est un réseau informatique mondial constitué d'un ensemble de réseaux nationaux, régionaux et privés. C'est un moyen d'accès à des ressources de toutes sortes, un moyen de communication et un lieu d'échange. Le service Internet le plus largement utilisé est sans doute le web ou le www [3].

Le Web (nom anglais signifiant toile), contraction du World Wide Web que l'on pourrait traduire en français par toile d'araignée mondiale (d'où l'acronyme www), une des possibilités offertes par le réseau Internet de naviguer entre des documents reliés par des liens hypertextes. Ces liens existent à travers tout le réseau, ils constituent une base de connaissances multimédias distribuée à grande échelle.

Le concept du Web a été mis au point au CERN (Centre Européen de Recherche Nucléaire) en 1991 par une équipe de chercheurs à laquelle appartenaient Tim-Berners , le créateur du concept d'hyperlien.

Le principe de Web repose sur l'utilisation d'hyperliens pour naviguer entre des documents (appelés pages web) grâce à un logiciel appelé navigateur ou en anglais browser [3].

Au-delà des liens reliant des documents formatés, le web prend tout son sens avec le protocole HTTP permettant de lier des documents hébergés par des ordinateurs distants (appelés serveurs). Sur Internet les documents sont ainsi repérés par une adresse unique appelée URL, permettent de localiser une ressource sur n'importe quel serveur du réseau internet.

2.3 Les lacunes du web actuel

Le Web contient de vastes bases de données avec plusieurs buts et de multiples sources non homogènes. Ceci prouve qu'il y a nécessité d'améliorer l'accessibilité à cette importante masse d'informations et disposer d'outils plus sophistiqués pour une meilleure recherche et organisation au sein du Web [4].

Il est clair que le Web actuel doit changer d'orientation d'un Web "présentable" qui se contente de Présenter ou d'afficher les informations disponibles, a un Web "intelligent" ou compréhensible.

Cette réorientation doit s'effectuer pour ne pas ruiner les moteurs de recherche et le e-business, surtout avec la quantité croissante des informations sur le Web.

Ceci nous montre que le Web dans son état actuel souffre de plusieurs lacunes, le problème qu'on doit résoudre est: " Comment trouver, de façon précise et rapide, l'information que l'on recherche sur la toile ? " [5].

La réponse à cette question passe par la nécessité de changer la structure du Web actuel. Cette dernière se contente d'établir des liens entre des fichiers et de les formater en vue de les afficher. En changeant cette structure en une autre, qui permet de relier le sens des informations se trouvant à l'intérieur des fichiers, on va permettre aux machines de mieux exploiter ces informations. Cette vision correspond bien au principe du Web Sémantique.

En plus des points évoqués plus haut et qui ont poussé les chercheurs à vouloir améliorer le Web actuel, une autre raison a fait son apparition durant l'année 2001.

Pendant ces dernières années, un nouveau terme a vu le jour et a eu beaucoup de succès, il s'agissait du terme " portail ". Par ce mot, on désigne une énorme base de données de ressources documentaires (textes, images, sons, vidéos, etc.) comme Lycos, Yahoo, etc.

Ces bases de données ne peuvent pas être exploitées efficacement avec les techniques standards de la recherche d'informations (RI) [6].

Ce besoin a renforcé l'utilité du Web Sémantique dans le domaine de la RI.

3 le Web Sémantique

3.1 Définition

Le Web Sémantique (*SW*) est une extension ou amélioration du Web actuel, dans lequel la signification des données est prise en compte, afin de permettre aux utilisateurs et aux machines de travailler en collaboration en utilisant des règles d'inférence. Tout cela aura pour conséquence de réduire les données non pertinentes et accroître l'utilité du Web [7].

Le Web Sémantique est défini comme étant un concept dans lequel le Web comporte des données définies et reliées entre elles à l'aide de règles d'inférence. Ces données doivent être utilisées par des machines non seulement à des fins d'affichage et de présentation, mais pour " automatiser, intégrer et réutiliser ces données entre applications " [8].

Le but du Web sémantique est de rendre explicite le contenu sémantique des ressources dans le Web (documents, pages web, services...). Les ordinateurs et les agents logiciels

pourraient donc « comprendre » les informations contenues dans ces ressources grâce à un système de métadonnées formelles et aider les utilisateurs à exécuter et compléter leurs tâches, leurs requêtes de façon plus automatique et plus efficace [5].

3.2 Web Actuel versus Web Sémantique

Le tableau au-dessous explique la différence entre le web actuel et le web sémantique

Web actuel	Web Sémantique (WS)
Ensemble des documents	Ensemble des connaissances et documents
Basé essentiellement sur HTML	Basé essentiellement sur XML et RDF(S)
Recherche par mots-clés	Recherche par concepts
Utilisable par l'être humain	Utilisable par la machine et l'être humain

Table 1.1 Web Actuel Versus Web Sémantique

3.3 Les principales composantes du Web sémantique

Le Web Sémantique a été proposé en se basant sur les critiques adressées au web actuel :

- (i) certes HTML a permis de tisser tout un réseau d'informations par ses liens hypertextes, mais il n'a donné aucune sémantique à ces liens ce qui les rend pratiquement inexploitable par les machines, (ii) les métadonnées utilisées sont non structurées et limitées dans leurs usages, et (iii) il est difficile de faire des inférences et des raisonnements sur les connaissances décrites dans les documents publiés sur le web vu l'absence de modèles permettant la représentation sémantique de ces connaissances [9].

Tous ces problèmes ont fait l'objet de différents travaux de recherche qui ont convergé vers plusieurs solutions parmi lesquelles nous présentons celles qui nous semblent les plus essentielles:

- Proposer des langages et des formalismes de représentation et de structuration des connaissances. Ces langages permettent de modéliser et de représenter le contenu sémantique des ressources du web.
- Rendre disponibles des ressources conceptuelles (des modèles) représentées dans ces langages modélisant les connaissances et facilitant leur accès et leur partage : les ontologies.

- Proposer des métadonnées explicites, c'est-à-dire qui suivent un modèle et qui sont exprimées dans des langages définis formellement.

3.4 Architecture du Web sémantique

Le Web sémantique nécessite une architecture partagée par tous pour échanger des ressources sur l'Internet. Il a besoin aussi des standards pour :

- Les ontologies et mécanismes d'inférence associés.
- La sémantique explicite des ressources ou méta données sémantiques liées à des ontologies.
- Le format des ressources ou documents.
- L'adressage des ressources ou documents.

Il y a des nombreux des langages sémantique marqueur dans la pile de Web sémantique. D'abord, XML fournit une syntaxe extérieure pour les documents structurés mais n'impose aucune contrainte sémantique à la signification de ces documents. Le schéma de XML est une langue pour limiter la structure des documents de XML. Puis, RDF est une langue pour créer un modèle de données pour des objets (ou des ressources) et des relations parmi eux. Il peut également fournir une sémantique simple pour ce modèle de données. Les modèles de données sont représentés en syntaxe de XML. Le schéma de RDF est un vocabulaire pour décrire des propriétés et des classes des ressources de RDF, avec la sémantique pour des hiérarchies de généralisation de telles propriétés et classes. Ensuite, le 'OWL ajoute plus de vocabulaire pour décrire des propriétés et des classes. En outre, il peut ajouter les relations parmi des classes. On peut dire qu'OWL ajoute L'expression de la signification et la sémantique au XML, RDF et RDF Schéma. Donc, on peut représenter le contenu compréhensible par une machine [10,11].

Nous allons maintenant présenter les trois niveaux importants du Web Sémantique : Le niveau de l'adressage, le niveau syntaxique et le niveau sémantique.

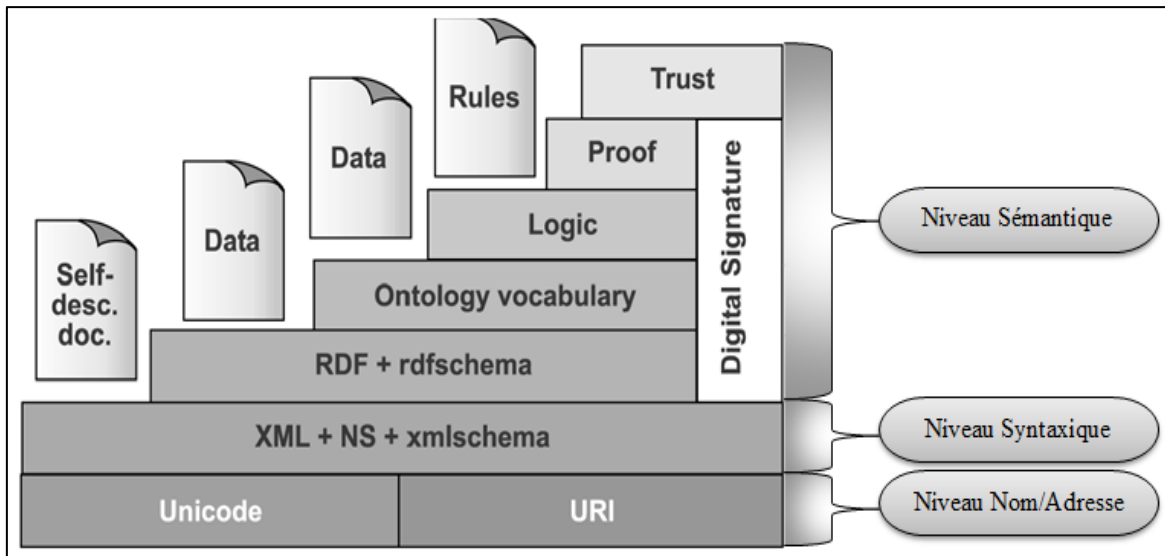


Figure 1.3 Le trois niveaux du Web Sémantique

3.4.1 Niveau de Nommage/Adressage

Le World Wide Web repose sur un concept important qu'est l'URI (Uniform Resource Identifier). Tout ce qui est disponible sur Internet doit être identifié par un URI. Un URI identifie de manière unique et non ambiguë chaque ressource du Web, comme une page, une adresse email, ou une image. Le point central des URIs est l'URL (Uniform Resource Locator) traditionnelle utilisée pour définir les liens du Web.

Ces URLs sont utilisées pour référencer des fichiers Web à travers un protocole particulier, comme HTTP ou FTP [3].

3.4.2 Niveau Syntaxique

Le niveau syntaxique est le niveau de la structuration des documents. La spécification de la structure logique des documents repose sur XML.

XML est un métalangage qui permet de définir d'autres langages. Les langages définis par XML sont des langages de présentation de documents. Il est aussi particulièrement adapté à l'échange de données. Son champ d'application est très vaste. XML permet de stocker des données structurées dans un fichier de texte. Elles y sont stockées dans un format simple et lisible relativement facilement [8].

XML fournit un ensemble de règles pour la création de vocabulaires qui structurent à la fois les documents et les données sur le Web. XML donne des règles très claires sur la syntaxe. Les schémas XML sont utilisés ensuite comme méthode autorisant la composition de vocabulaires XML. XML est une syntaxe puissante et flexible pour les documents structurés, mais n'impose aucune contrainte sémantique sur la signification de ces documents.

XML est utilisé sur le Web car il permet la description de documents électroniques par l'intermédiaire d'une DTD (Document Type Definition). Il fait le lien entre les documents et les logiciels qui les utilisent. En effet, le but de XML est de faciliter la diffusion et l'échange d'informations sur Internet.

Les points forts de XML en plus de sa simplicité sont :

- Son indépendance quant à la plateforme utilisée
- Son exploitation possible par un système informatique
- La séparation de la présentation et du contenu
- C'est un langage de description facilement extensible en fonction des besoins des applications
- La gestion de la cohérence grâce aux DTD

L'avantage de XML est la possibilité de personnaliser la présentation des documents en utilisant XSL (XML Stylesheet Language) qui permet de transformer automatiquement un fichier

XML en une page HTML qui est consultable via un navigateur Internet. Cependant, son inconvénient est : Il n'a pas de sémantique formelle permettant l'interprétation par la machine. XML décrit uniquement la structure de l'information, sa syntaxe.

3.4.3 Niveau Sémantique

RDF (Resource Description Framework) est un standard permettant la mise en place de descriptions simples. XML est à la syntaxe, ce que RDF est à la sémantique. RDF Schéma permet ensuite de combiner ces descriptions en un seul vocabulaire. tout ceci, il manque la possibilité de décrire des vocabulaires spécifiques à des domaines bien particuliers. C'est là que les ontologies jouent leur rôle [10].

Une ontologie définit les termes utilisés pour décrire et représenter un champ d'expertise. Les ontologies sont utilisées par les personnes, les bases de données, et les applications qui ont besoin de partager des informations relatives à un domaine bien spécifique, comme la médecine, la fabrication d'outils, l'immobilier, la réparation d'automobiles, la gestion de finances, etc... Les ontologies associent les concepts de base d'un domaine précis et les relations entre ces concepts, tout cela d'une manière compréhensible par les machines. Elles encodent la connaissance d'un domaine particulier ainsi que les connaissances qui recouvrent d'autres domaines, ce qui permet de rendre les connaissances réutilisable [11].

3.4.4 Ontologie pour le Web sémantique

Un des objectifs principaux du Web sémantique est le partage et la réutilisation d'information pour l'homme et machine. Donc, on doit ajouter la connaissance à l'information pour améliorer la compréhension pour la machine. Pour le faire, simplement, on annote des ressources avec la signification des concepts. Donc, on a besoin de même concepts pour représenter la même chose dans l'indexation de différentes ressources. Dans ce cas, on a besoin le terme 'Ontologie'.

L'ontologie a joué un rôle très important pour le Web sémantique parce qu'il va représenter la sémantique des documents en permettant son exploitation par les applications et les agents intelligents. Elle est très utile pour structurer et définir la signification des termes de métadonnées actuellement collectés et normalisés. Donc, à l'aide des ontologies, les applications sur Web de demain pourront devenir intelligentes, au sens où elles pourront opérer plus précisément au niveau conceptuel humain.

4 Les ontologies

4.1 Définition

Le terme Ontologie est emprunté à la philosophie. Il est défini par Aristote par « la science de l'être ». La discipline analyse et représente les propriétés des êtres et leur existence. Non seulement le sens du terme a évolué dans le domaine de la philosophie, d'autres sciences se le sont par ailleurs approprié : logique, informatique, sciences du langage, sciences de l'information, etc. Il est actuellement fortement corrélé au web sémantique qui manipule des concepts ou toutes sortes de données informatiques structurées [12].

En informatique, l'ontologie s'inscrit dans le cadre de l'intelligence artificielle où les raisonnements humains sont simulés par une machine. Le terme ontologie emprunte à la philosophie la démarche de définition, de formalisation, d'identification ou encore de représentation des connaissances. Bien que ce point soit sujet à débat, on retrouve communément l'idée qu'en philosophie les ontologies se veulent absolues alors qu'en informatique elles dépendent d'un consensus.

Les ontologies sont une formalisation logique, structurée et relativement exhaustive des connaissances pour un domaine donné et sont réalisées en fonction d'un usage spécifique. Ces connaissances, également appelées concepts, peuvent être des idées, des principes, des catégories d'objet du monde. Elles ont des propriétés, peuvent être mises en relation et sont

soumises à variations selon des événements et des processus. « Ontologie du domaine » ou « modèle conceptuel » peuvent également servir à nommer le résultat [13].

La définition informatique est très fréquemment associée à Gruber qui affirme : « Une ontologie est une spécification explicite d'une conceptualisation ».

Damers propose une définition de l'ontologie en 4 points [21] :

- ontologie comme étude de ce qui est : « nature et organisation de la réalité » ;
- ontologie comme (représentation d'une) conceptualisation ; conceptualisation des objets, de leurs propriétés et des relations qu'ils entretiennent ainsi que conceptualisation relative à un consensus ;
- ontologie formelle ; nécessitant une représentation rigoureuse, selon des standards de description ;

Une ontologie est un ensemble structuré sous la forme d'un graphe orienté qui contient :

- des nœuds représentant le vocabulaire d'un domaine particulier.
- des arcs représentant les relations (ou rôles) nommées entre les concepts.

A partir de cette structure, la sémantique de chaque mot est déduite par les relations que ce mot possède dans l'ontologie, ce qui permet de restreindre les interprétations possibles.

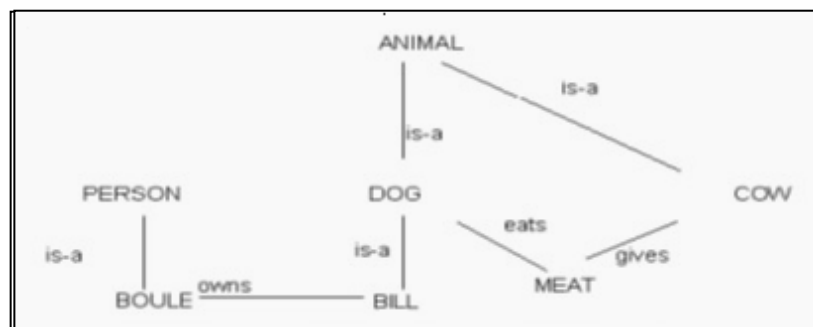


Figure 1.4 Réseau Sémantique avec les concepts en nœud et les relations en arc

Cependant pour être exploitable par une machine, une ontologie doit respecter certaines règles :

- être définie par une syntaxe formelle et une sémantique non ambiguë.
- permettre la déduction de nouvelles connaissances présente implicitement dans l'ontologie.

4.2 Les constituants d'une ontologie

Une ontologie ne peut être construite que dans le cadre d'un domaine de connaissance précis, ne serait-ce que parce que beaucoup de termes n'ont pas le même sens d'un domaine à l'autre, et qu'une sémantique non ambiguë doit être intégrée à l'ontologie. Un domaine de connaissances est donc constitué par les objets du domaine et par un contexte d'usage de ces objets [14].

Délimiter rigoureusement un domaine de connaissances peut par contre se révéler ardu à cause de la nature holistique de la connaissance. Certaines connaissances, qui peuvent constituer en elles-mêmes un domaine, sont utilisées dans tous les autres domaines.

C'est le cas des notions générales liées à la causalité, au temps, à l'espace, etc. Savoir borner le domaine des connaissances à représenter demande donc une délimitation extrêmement précise de l'objectif opérationnel de l'ontologie [14].

Les connaissances traduites par une ontologie sont à véhiculer à l'aide des éléments suivants : Concepts ; Relations ; Fonctions ; Axiomes; Instances. Ces notions nécessitent cependant d'être explicitées.

4.2.1 Les concepts

Les concepts, correspondent aux abstractions pertinentes d'un segment de la réalité, retenues en fonction des objectifs qu'on se donne et de l'application envisagée pour l'ontologie. Un concept peut représenter un objet matériel, une notion ou une idée [16].

Un concept peut être divisé en trois parties : un terme (ou plusieurs), une notion et un ensemble d'objets. La notion, également appelée "intension" du concept, contient la sémantique du concept, exprimée en termes de propriétés et d'attributs, de règles et de contraintes. L'ensemble d'objets, également appelé "extension" du concept, regroupe les objets manipulés à travers le concept ; ces objets sont appelés "instances" du concept. Par exemple, le terme « table » renvoie à la fois à la notion de table comme objet de type «meuble» possédant un plateau et des pieds, et à l'ensemble des objets de ce type [16].

Les deux aspects d'un concept sont assez différents, en particulier par le fait que deux concepts peuvent partager la même extension sans pour autant avoir la même intension. C'est le cas des concepts d'« étoile du matin » et d'« étoile du soir », qui désignent tous deux

Vénus. De plus, des concepts partageant la même extension mais pas leur intension peuvent être désignés par le même terme. Ceci correspond à des points de vue différents sur un même objet. Par exemple, les chiens peuvent être considérés comme des animaux de compagnie, ou comme des ressources culinaires [18].

Bien que le langage naturel contienne de nombreux termes désignant plusieurs concepts sémantiques différents (par exemple « table » pour un meuble et « table » pour un tableau de valeurs numériques), de telles ambiguïtés ne sont pas gérables en machine, où on identifie généralement un concept à l'aide de ses termes.

Mais la restriction à un domaine de connaissance permet généralement d'éviter les homonymies de concepts.

Une liste des principales propriétés pouvant être associées aux concepts est donnée ci-dessous [15] :

- **la généricité** : un concept est générique s'il n'admet pas d'extension. Exemple: la "vérité" est un concept générique, prise dans le sens de « ce qui est vrai » et non pas du «degré de vérité »;
- **la rigidité** : (propriété proposée par N. GUARINO) un concept est rigide si toute instance de ce concept en reste instance dans tous les mondes possibles.
Exemple: "humain" est un concept rigide, "étudiant" est un concept non rigide ;
- **l'unité** : (propriété proposée par N. GUARINO) 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.
Exemple: les deux parties d'un "couteau", manche et lame sont liées par une relation «emmanché » qui ne lie que cette lame et ce manche.
- **l'équivalence** : deux concepts sont équivalents s'ils ont la même extension. Exemple: "étoile du matin" et "étoile du soir" ;
- **la disjonction** : (on parle aussi d'incompatibilité) deux concepts sont disjoints si leurs extensions sont disjointes. Exemple: "homme" et "femme";
- **la dépendance** : (propriété proposée par GUARINO) 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: "parent" est un concept dépendant de "enfant" (et vice-versa).

4.2.2 les relations

Si certains liens conceptuels existant entre les concepts peuvent s'exprimer à l'aide de propriétés portées par les concepts, d'autres doivent être représentés à l'aide de relations autonomes. Une relation permet de lier des instances de concepts, ou des concepts génériques. Elles sont caractérisées par un terme (voire plusieurs) et une signature qui précise le nombre d'instances de concepts que la relation lie, leurs types et l'ordre des concepts, c'est-à-dire la façon dont la relation doit être lue. Par exemple, la relation « écrit » lie une instance du concept « personne » et une instance du concept « texte », dans cet ordre [17].

Tout comme les concepts, les relations peuvent être spécifiées par des propriétés dont une liste, non exhaustive, est donnée ci-dessous :

- **les propriétés algébriques** : symétrie, réflexivité, transitivité ;
- **la cardinalité** : nombre possible de relations de ce type entre les mêmes concepts (ou instances de concept). Les relations portant une cardinalité représentent souvent des attributs. Exemple: une pièce a au moins une porte, un humain a entre zéro et deux jambes [15].
- **l'incompatibilité** : deux relations sont incompatibles si elles ne peuvent lier les mêmes instances de concepts. Exemple: les relations «être rouge» et «être vert»
- **l'inverse** : deux relations sont inverses, quand l'une lie deux instances I1 et I2, l'autre lie I2 et I1. Exemple: les relations «a pour père» et «a pour enfant»
- **l'exclusivité** : deux relations sont exclusives si, quand l'une lie des instances de concepts, l'autre ne lient pas ces instances, et vice-versa. L'exclusivité entraîne l'incompatibilité.
exp: l'appartenance et le non appartenance sont exclusives.

4.2.3 Les fonctions

Les fonctions constituent des cas particuliers de relations, dans laquelle un élément de la relation, le nième (extrant) est défini en fonction des n-1 éléments précédents (intrants) [14].

4.2.4 Les axiomes

Les axiomes permettent de modéliser des assertions acceptées comme vraies, à propos des abstractions du domaine traduites par l'ontologie. Leur inclusion dans une ontologie peut avoir plusieurs objectifs: interviennent dans la définition des significations des composants d'ontologie, les contraintes sur les valeurs des attributs, les arguments de relations et dans l'inférence de nouvelles informations, etc. D'après, Les axiomes formels sont utilisés pour vérifier la consistance de l'ontologie... [15].

4.2.5 Les instances

Les instances constituent la définition extensionnelle de l'ontologie ; ces objets véhiculent les connaissances (statiques, factuelles) à propos du domaine du problème [14].

4.3 Différents types d'ontologies

Les éléments de l'ontologie (concepts, relations, axiomes et instances) sont définis explicitement avec un langage dont la sémantique est plus ou moins formelle, dépendant du degré d'abstraction voulu. Par conséquent, les ontologies présentent des degrés d'abstraction différents dépendamment du domaine auquel elles sont dédiées.

On identifie trois types d'ontologies selon un niveau décroissant d'abstraction : les ontologies globales, les ontologies de domaine, ou dédiées à une tâche plus spécifique, et les ontologies d'application.

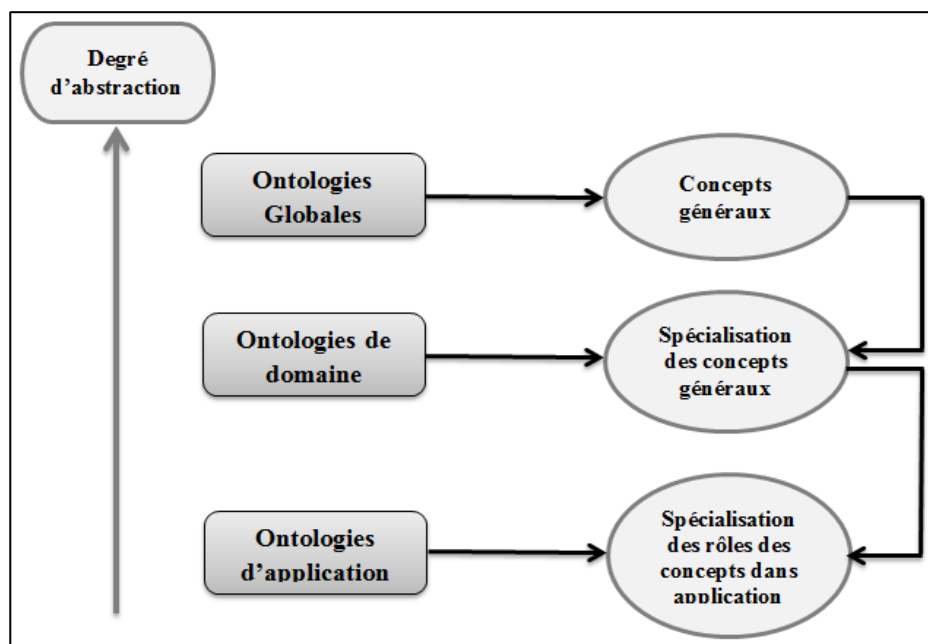


Figure 1.5 Types d'ontologies

Les ontologies globales, c'est-à-dire celles qui présentent le plus haut niveau d'abstraction et de généralité, sont les ontologies formelles, car elles sont issues d'un développement systématique, rigoureux et axiomatique de la logique de toutes les formes et modes d'existence. L'adoption de principes rigoureux dans la conception de l'ontologie formelle répond au besoin de disposer de connaissances pouvant être partagées et transférées d'un contexte à l'autre. Elles sont dédiées à des utilisations générales (ex : WordNet). Une ontologie formelle est donc une théorie des distinctions formelles entre les éléments d'un

domaine, indépendamment de leur réalité. Au second degré d'abstraction, on trouve les ontologies de domaine qui sont limitées à la représentation de concepts dans des domaines donnés (géographie, médecine, écologie, etc.) et qui spécialise les concepts de l'ontologie globale. Finalement, les ontologies d'applications offrent le plus fin niveau de spécificité, c'est-à-dire qu'elles sont dédiées à un champ d'application précis à l'intérieur d'un domaine et décrivent le rôle particulier des entités de l'ontologie de domaine dans ce champ. Par exemple, l'ensemble des spécifications sur la forêt de Montmorency constitue une ontologie d'application qui spécifie les concepts généraux pouvant provenir d'une ontologie de domaine forestier générale [9].

4.4 Les différentes approches pour la conception d'ontologie

En général, il existe trois principales approches pour la conception d'une ontologie de domaine : inspirationnelle, synthétique et collaborative. La table 2 ci-dessous décrit ces trois approches.

Approches de conception d'ontologie	Base de conception
Inspirationnelle	Point de vue individuel sur le domaine.
Synthétique	Ensemble d'ontologies existantes, chacune fournit une caractérisation partielle du domaine.
Collaborative	Multiple points de vue sur le domaine.

Table 1.2 Les trois principales approches pour la conception d'ontologie

Dans l'approche *inspirationnelle*, un concepteur seul prend les décisions de rassembler les conditions de l'analyse de domaine, de la conception et de la vérification de l'ontologie. De ce fait, le développeur doit être à la fois expert du domaine et expert en conception d'ontologie (i.e. ontologistes) afin d'assurer le succès de la conception, et surtout, pour assurer l'adoption de l'ontologie par la communauté d'utilisateurs. Ainsi, ce type d'approche dépend fortement de la créativité d'une seule personne, de son inspiration et de sa perception personnelle du domaine. Cette approche est souvent appliquée dans un domaine qui peut être cerné (compris) par un seul concepteur, où ce dernier est capable de dominer le processus de conception de l'ontologie [12].

Dans l'approche *synthétique*, un ensemble d'ontologies liées est identifié. Le concepteur d'ontologie synthétise alors les éléments de ces ontologies préexistantes avec les concepts du nouveau domaine à représenter, produisant ainsi une nouvelle ontologie unifiée.

L'approche *collaborative*. La marque d'une ontologie "moderne" est sa grande taille et sa haute complexité ; ces ontologies englobent un ensemble si riche de la connaissance que leur compréhension (complète) dépasse celle de n'importe quel développeur ou concepteur seul ou même d'une petite équipe de concepteurs. De ce fait, le développement d'une ontologie à grande échelle doit être le fruit d'un effort commun de plusieurs experts de domaine et des ingénieurs de la connaissance, voir même les futurs utilisateurs de l'ontologie [12].

4.5 Cycle de vie d'une ontologie

Puisque les ontologies sont destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel. Ainsi, les ontologies doivent être considérées comme des objets techniques évolutifs et possédants un cycle de vie qui nécessite d'être précisé. Dans ce contexte, les activités liées aux ontologies sont, d'une part, des activités de gestion de projet (planification, contrôle, assurance qualité), et, d'autre part, des activités de développement (spécification, conceptualisation, formalisation) ; s'y ajoutent des activités transversales de support telles que l'évaluation, la documentation, la gestion de la configuration [19].

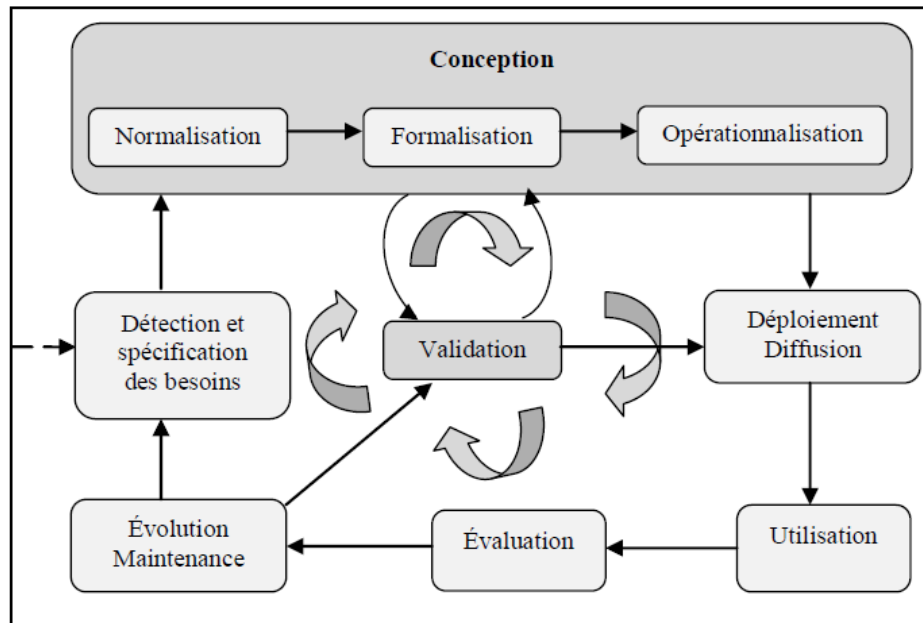


Figure 1.6 Cycle de vie d'une ontologie

Un cycle de vie, voire *figure 1. 6* [20] Il comprend une étape initiale de détection et de spécification des besoins qui permet notamment de circonscrire précisément le domaine de connaissances, une étape de conception qui se subdivise en trois phases, une étape de déploiement et de diffusion, une étape d'utilisation, une étape, incontournable, d'évaluation, et enfin, une sixième étape consacrée à l'évolution et à la maintenance du modèle. Après chaque utilisation significative, l'ontologie et les besoins doivent être réévalués et l'ontologie peut être étendue et, si nécessaire, en partie reconstruite. La validation du modèle de connaissances est au centre du processus et se fait de manière itérative.

4.6 Niveaux de formalisation des ontologies

Les ontologies peuvent être modélisées avec des langages de différents niveaux de formalité :

- Informelles : ontologies opérationnelles dans un langage naturel (sémantique ouverte);
- Semi-informelles : utilisation d'un langage naturel structuré et limité ;
- Semi-formelles : langage artificiel défini formellement;
- Formelles : utilisation d'un langage artificiel contenant une sémantique formelle, ainsi que des théorèmes et des preuves de propriétés telles la robustesse et l'exhaustivité.

4.7 Processus de développement des ontologies

Les ontologies nous a permis de dégager un consensus sur le processus de conception d'une ontologie exploitable dans un système à base de connaissances [19]. La **figure 1.7** illustre ce consensus qui repose sur un enchaînement de trois étapes, permettant de passer des données brutes à l'ontologie opérationnelle. Les données brutes, constituant un corpus (exprimé *priori* en langage naturel). En effet, pour construire une ontologie, il faut tout d'abord recueillir son corpus. La nature du corpus dépend du domaine d'application de l'ontologie. En effet, un corpus peut être sous forme de documents textuels, des interviews, des représentations conceptuelles, de sites Web, etc., ... Une fois le corpus prêt, la première étape de construction de l'ontologie peut être déclenchée:

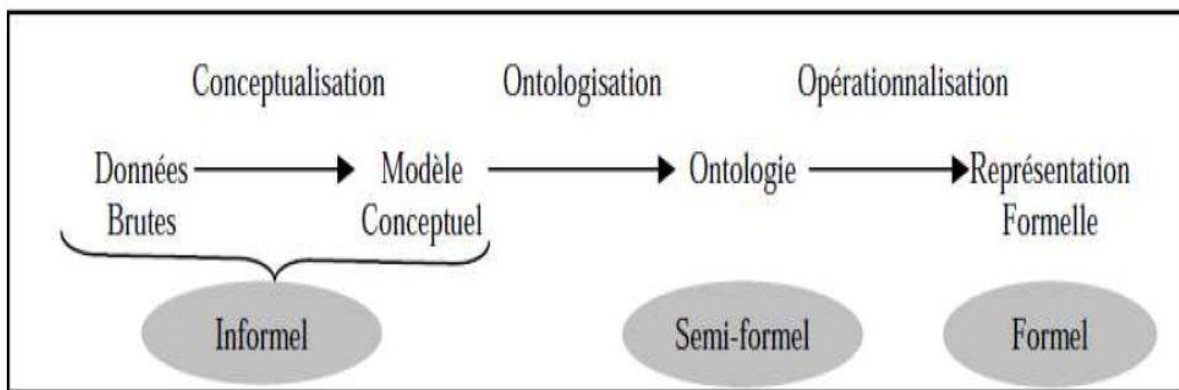


Figure 1.7 Processus général de construction d'une ontologie

Bien que les ontologies obtenues en amont du processus. Par la suite nous allons clarifier les diverses opérations menées à chaque étape ainsi que le statut de l'ontologie obtenue à chacune de ces étapes. Ces étapes soient effectuées en séquence, il est évident que des « *retours en arrière* » sont possibles ; en d'autres termes, la réalisation d'une étape peut provoquer la remise en cause des choix adoptés aux étapes précédentes et, par voie de conséquence, conduire à compléter ou modifier [22].

a) Conceptualisation

La première étape, appelée conceptualisation, permet d'aboutir à un modèle informel, donc sémantiquement ambiguë et généralement exprimé en langage naturel. Cette étape, consiste, à partir des données brutes, à dégager les concepts et les relations entre ces concepts permettant de décrire de manière informelle les entités cognitives du domaine. Elle est réalisée par un expert du domaine assisté de l'ingénieur de la connaissance et aboutit à un modèle conceptuel. Ainsi, le modèle obtenu consiste en un ensemble de termes désignant les entités du domaine

de connaissances (concepts, relations, propriétés des concepts et des relations, etc.), assortis d'informations exprimant leur sémantique. La découverte des connaissances d'un domaine peut s'appuyer 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 émerge.

Il est à noter que certaines connaissances implicitement utilisées dans le domaine ne sont cependant jamais exprimées, ni dans le corpus, ni par les experts [9]. Un des points les plus délicats de la conceptualisation consiste donc à identifier ces connaissances. La mise en évidence de ces connaissances implicites ne peut a priori se faire que lors de l'utilisation de l'ontologie, lors d'une phase de test opérationnel et/ou de test des questions de compétences. Ceci montre que le processus de construction des ontologies ne peut être séquentiel et que des allers retours entre les différentes étapes du processus sont à prévoir.

b) L'ontologisation

La première étape de cette formalisation est l'ontologisation, qui consiste en une formalisation partielle, sans perte d'information, du modèle conceptuel obtenu dans l'étape précédente. Cette formalisation partielle facilite sa représentation ultérieure dans un langage complètement formel et opérationnel [22]. De plus, il faut bien voir que l'ontologisation est une transcription des connaissances dans un certain formalisme de connaissances, ce formalisme devant être aussi générique que possible, mais présentant une sémantiquement clair.

S'imposer de conserver toutes les connaissances conduit à intégrer, à l'ontologie du domaine, des connaissances qui ne peuvent être formalisées, ou dont la sémantique est ambiguë.

Cependant, le modèle obtenu pendant à ce stade est souvent qualifié de semi-formel pour justifier le fait que certaines connaissances ne peuvent pas être totalement formalisées. Le caractère semi-formel d'une ontologie lui interdit d'être utilisée telle quelle dans un SBC. En revanche, une ontologie, contenant toutes les connaissances d'un domaine, constitue le support idéal de communication et de partage des connaissances de ce domaine. En fait, cette étape produit un résultat en deux parties :

- Une partie formelle, disposant d'une sémantique précise ou du moins consensuelle;
- Une partie informelle qui ne dispose pas d'une sémantique claire ou consensuelle, ou tout ou moins d'une sémantique fixée a priori et donc exprimée dans un langage naturel ou semi-structuré.

c) L'opérationnalisation

La dernière étape, appelée parfois représentation, consiste à formaliser complètement l'ontologie obtenue précédemment dans le cadre d'un langage de représentation de connaissances, formel et opérationnel. Le terme opérationnel à la fois pour caractériser un modèle de représentation doté de mécanismes de raisonnement et d'une sémantique opérationnelle, et pour caractériser tout langage exécutable qui implémente un tel modèle.

Cette étape est menée par un spécialiste du langage de représentation et par l'ingénieur de la connaissance. On obtient alors une représentation formelle des connaissances du domaine.

Ainsi, le caractère formel de l'ontologie permet à une machine, via cette ontologie, de manipuler des connaissances du domaine. Il faut noter que la formalisation totale des connaissances peut imposer à ce stade une perte d'information, du fait qu'aucune représentation de leur sémantique n'existe dans le langage de représentation d'ontologies choisi [20].

Finalement, l'ontologie opérationnelle est intégrée en machine au sein d'un système manipulant le modèle de connaissances utilisé via le langage opérationnel choisi. Mais avant d'être livrée aux utilisateurs, l'ontologie doit, bien sûr, être testée par rapport au contexte d'usage pour lequel elle a été bâtie [22].

4.8 Les outils de construction d'ontologies

Les outils de construction d'ontologies ont été considérablement améliorés depuis la création du premier environnement. Si on prend en considération l'évolution des outils de construction d'ontologies depuis les milieux des années 1990 [11], on peut distinguer deux familles d'outils:

- Les outils de construction d'ontologie dépendants du formalisme de représentation, Ces outils ont été créés pour permettre l'édition et la navigation sur des ontologies selon leurs langage de développement, ainsi pour permettre l'importation et l'exportation d'ontologies à partir de / vers d'autres langages, mais ils exigent que les utilisateurs aient connaissance du langage d'ontologie correspondant, dans cette catégorie on peut citer les outils suivants : Ontolingua , OntoSaurus , WebOnto et OilEd [15].
- Les outils de construction d'ontologie indépendants du formalisme de représentation, dont la principale caractéristique est qu'ils ont une architecture extensible, et dont le modèle de connaissance est souvent indépendant du langage de l'ontologie. Ces outils ont été construits pour supporter une large variété d'activités du processus de développement d'ontologies. Pour ce faire, ils ont une extensible architecture à base de

composants, où de nouveaux modules peuvent facilement être ajoutés pour fournir plus de fonctions. Dans cette catégorie on peut citer les outils suivants : Protégé2000, ODE et son adaptation pour le Web WEBODE et Onto Edit [15].

4.9 Les méthodologies pour la construction d'ontologies

L'absence de directives structurées et communes ralentisse le développement d'ontologies à l'intérieur et entre les équipes, l'extension de n'importe quelle ontologie, la possibilité de réutilisation de l'ontologie. On attend par méthodologie, les procédures de travail, les étapes, qui décrivent le pourquoi et le comment de la conceptualisation puis de l'artefact construit. L'objectif de cette sous-section est de présenter les principales méthodologies utilisées pour construire les ontologies [23].

a) La méthode SENSUS

C'est une méthode qui consiste à construire une ontologie de domaine à partir d'une plus grande ontologie, l'ontologie SENSUS. La méthode propose de relier les termes spécifiques du domaine à cette ontologie et de tailler dans SENSUS, les termes qui ne sont pas relevant dans la nouvelle ontologie qu'on souhaite construire. Le résultat de ce processus est le squelette de cette nouvelle ontologie, qui est générée automatiquement en utilisant ce processus et l'outil OntoSaurus [24].

b) La méthodologie On-To-Knowledge(OTK)

Le but du projet On-To-Knowledge est d'appliquer les ontologies aux informations disponibles électroniquement pour améliorer la qualité de la gestion de la connaissance au sein des organisations larges et distribuées. Dans le cadre de ce projet, ils ont développés une méthodologie et des outils pour l'accès intelligent aux sources de large volume d'informations textuelles et semi-structurés dans des environnements basés sur l'intranet, extranet et internet. Cette méthodologie inclut une méthodologie pour la construction d'ontologies qui vont être utilisées par les applications de management de connaissances [23]. 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. Par conséquent les ontologies développées en suivant cette méthodologie sont très dépendantes de l'application [24].

c) methontology

Cette méthodologie a été développée au sein du groupe d'ontologie à l'université polytechnique de Madrid. METHONTOLOGY est née dans les travaux de software développent procès et dans knowledge engineering méthodologies [23].

Cette méthodologie inclut : l'identification du processus de développement de l'ontologie, le cycle de vie basé sur l'évolution de prototypes et les techniques de gestion de projet (planification, assurance qualité), de développement (spécification, conceptualisation, formalisation, implémentation, maintenance) et des activités de support (intégration, évaluation, documentation).

Les outils ODE et WeODE ont été construit pour donner un support technique à METHONTOLOGY. D'autres outils peuvent aussi être utilisés pour construire des ontologies en suivant cette méthodologie, par exemple, Protégé, Onto Edit, etc [24].

5 Conclusion

Le Web sémantique est une vision de l'information qui permet d'être lisible par les humains et par les machines. Cela permettra d'effectuer les travaux fastidieux et répétitifs dans le domaine de la recherche d'information par des machines tout en améliorant et consolidant l'information sur le Web pour ses utilisateurs. Les ontologies sont un des concepts de base du Web sémantique. Une ontologie est l'ensemble structuré des termes et concepts représentant le sens d'un champ d'informations, que ce soit par les métadonnées d'un espace de noms, ou les éléments d'un domaine de connaissances.

Table de matière

1 Introduction	4
2 Le web syntaxique.....	4
2.1 Historique	4
2.2 Définition de web (World Wide Web)	6
2.3 Les lacunes du web actuel.....	6
3 le Web Sémantique	7
3.1 Définition	7
3.2 Web Actuel versus Web Sémantique	8
3.3 Les principales composantes du Web sémantique	8
3.4 Architecture du Web sémantique	9
3.4.1 Niveau de Nommage/Adressage	10
3.4.2 Niveau Syntaxique	10
3.4.3 Niveau Sémantique	11
3.4.4 Ontologie pour le Web sémantique	12
4 Les ontologies	12
4.1 Définition	12
4.2 Les constituants d'une ontologie.....	14
4.2.1 Les concepts	14
4.2.2 les relations	16
4.2.3 Les fonctions	16
4.2.4 Les axiomes	16
4.2.5 Les instances	17
4.3 Différents types d'ontologies	17
4.4 Les différentes approches pour la conception d'ontologie	18
4.5 Cycle de vie d'une ontologie.....	19
4.6 Niveaux de formalisation des ontologies	20
4.7 Processus de développement des ontologies	21
a) Conceptualisation	21
b) L'ontologisation	22
c) L'opérationnalisation	23
4.8 Les outils de construction d'ontologies.....	23
4.9 Les méthodologies pour la construction d'ontologies.....	24
5 Conclusion.....	25

Les Figure

Figure 1.1 web 1.0 vs web 2.0 [1]	4
Figure 1.2 l'évolution de web [1]	5
Figure 1.3 Le trois niveaux du Web Sémantique	10
Figure 1.4 Réseau Sémantique avec les concepts en nœud et les relations en arc	13
Figure 1.5 Types d'ontologies.....	17
Figure 1.6 Cycle de vie d'une ontologie	20
Figure 1.7 Processus général de construction d'une ontologie	21

Les Tableaux

Table 1.1 Web Actuel Versus Web Sémantique	8
Table 1.2 Les trois principales approches pour la conception d'ontologie.....	18