



UNIVERSITÉ LUMIÈRE LYON 2
ÉCOLE DOCTORALE DE SCIENCES COGNITIVES

MÉMOIRE DE MASTER 2 RECHERCHE EN INFORMATIQUE
spécialité
EXTRACTION DES CONNAISSANCES À PARTIR DES DONNÉES

UTILISATION DE TECHNIQUES DE FOUILLE DE DONNÉES POUR ENRICHIR
UNE ONTOLOGIE D'OBJETS GÉOGRAPHIQUES

par

FADI BADRA

le 26 Juin 2005

Encadré par :

OMAR BOUSSAID

Maitre de Conférences Université Lumière Lyon2

et

ANNE PUISSANT

Maitre de Conférences Université de Caen

Stage réalisé au sein du Laboratoire ERIC
(ÉQUIPE DE RECHERCHE EN INGÉNIERIE DES CONNAISSANCES)

Université Lyon 2

5 Avenue Pierre Mendès-France

69676 BRON Cedex - France

<http://eric.univ-lyon2.fr>

Remerciements

Je tiens tout d'abord à remercier Omar et Anne pour avoir rendu possible cette étude et m'avoir fait confiance tout au long de ce travail.

Plus largement, je souhaite saluer toute l'équipe de FoDoMuST pour la grande convivialité des réunions et pour m'avoir donné à vivre ces débats passionnés. J'espère que ce travail vous servira !

Merci aussi infiniment à Jean pour ses précieuses relectures et à Mathieu et Sylvain pour leurs conseils avisés, sans qui je n'aurais jamais su à quel point l'homme est un animal p«*OWL*»itique...

Merci également à tous ces joyeux travailleurs qui ont eu le courage de braver comme moi la chaleur étouffante des salles de l'université pour rédiger leur rapport et bon courage à tous ceux qui continueront à la braver l'été durant.

Enfin (surtout !), je ne peux oublier ici mes petites femmes qui m'ont soutenues tout au long de cette année d'éloignement... qui sait, peut-être seront nous un jour de grands chercheurs ???

Résumé

La recherche sur les ontologies suscite un intérêt croissant dans le domaine du traitement de l'information géographique. Celles-ci apparaissent comme un composant essentiel pour réaliser l'intégration de sources de données hétérogènes et faire face à la considérable augmentation de la quantité de données d'observation de la Terre aujourd'hui disponibles. Cette étude s'inscrit dans le projet multidisciplinaire FoDoMuST, qui vise à extraire et qualifier des objets urbains à partir d'images satellites multi-spectrales et multi-résolution. Un objet urbain est un concept de la géographie urbaine, discipline scientifique qui étudie les dimensions spatiales du phénomène urbain et dont l'objet d'étude est la ville. Nous nous sommes concentrés essentiellement sur les aspects conception et implémentation d'une ontologie d'objets urbains. Notre démarche a abouti à la proposition d'un modèle conceptuel de l'ontologie ainsi qu'à une étude de son opérationnalisation et des perspectives de son enrichissement par des processus d'extraction de connaissances à partir de données.

Table des matières

Introduction	1
I Contexte	2
1 Le projet FoDoMuST	2
1.1 Objectifs de l'ACI	2
1.2 Une masse de données	2
1.3 Une ontologie pour formaliser les connaissances du domaine	2
2 Cahier des charges	2
II Les ontologies	5
3 Qu'est-ce qu'une ontologie ?	5
3.1 Un artefact informatique	5
3.2 Une représentation...	5
3.3 ...des connaissances	5
3.4 Les briques de la maison ontologie	6
3.5 Ontologies et bases de données	7
4 Conception d'une ontologie	7
4.1 La démarche de conception	7
4.2 Conceptualisation	8
4.2.1 Une spécification d'une conceptualisation	8
4.2.2 Engagement ontologique	8
4.3 Formaliser le modèle conceptuel	9
4.3.1 Un langage	9
4.3.2 Une sémantique	9
4.3.3 Un raisonnement	10
4.4 Les systèmes classificatoires	10
4.4.1 Les cadres (<i>frames</i>)	10
4.4.2 Les logiques de descriptions	11
5 Les ontologies géographiques	13
5.1 Définition	13
5.2 Régions de l'espace ou objets dans l'espace ?	13
5.3 Objets microphysiques, objets mésoscopiques	14
5.4 Revue de presse	14
5.5 Le modèle conceptuel de SPRING	15
5.6 Etat de l'art : bilan	16
III L'ontologie de FoDoMuST	17

6	Conceptualisation	17
6.1	Les images	17
6.2	Différents types d'objets géographiques	17
6.3	La photointerprétation	18
6.4	Variabilité des concepts	20
6.5	Le modèle conceptuel proposé	21
7	Ontologisation	23
7.1	Le langage retenu : OWL	23
7.2	Un langage riche	23
7.3	Un mécanisme d'inférence efficace	24
7.4	Des limitations concernant les domaines de définition des attributs	25
8	Opérationnalisation	25
8.1	Utilisation d'un raisonneur	25
8.2	Déroulement du processus de reconnaissance	25
9	Implémentation	27
9.1	Choix d'un raisonneur : Racer	27
9.2	Solutions techniques pour la déclaration de domaines de définition des attributs	28
IV	Enrichissement de l'ontologie	29
10	Motivations	29
10.1	Ajouter des connaissances décisionnelles à l'ontologie	29
10.2	Quelles sont les données d'apprentissage ?	30
10.3	Quelle stratégie adopter ?	31
	Conclusion et perspectives	32
V	Annexes	33

Introduction

L'apparition de nouvelles générations de satellites d'observation de la Terre met aujourd'hui à disposition des géographes de grandes quantités d'images prises à très haute résolution. Un satellite comme Quickbird, lancé en 2001, fournit en effet des images d'une résolution pouvant aller jusqu'à 70 cm en panchromatique. Cette profusion d'images constitue une source de données complexes, puisque les prises de vues satellitaires varient tant en fonction de la couverture spectrale que de la résolution, de la date ou du lieu où elles sont prises.

Pour peu que l'homme parvienne à les exploiter, ces nouveaux moyens technologiques offrent la possibilité d'obtenir des informations précises, détaillées et régulières sur l'ensemble de la surface du globe. La géographe Anne Puissant a en particulier mis en évidence, au terme d'une analyse des besoins menée auprès de différents acteurs de l'information géographique, l'apport de la très haute résolution pour des applications comme la prévention des risques ou l'urbanisme [Pui03].

Actuellement, la réalité mesurée par les capteurs satellitaires est extraite des données par photointerprétation. Celle-ci se fait par confrontation des différentes images collectées se rapportant à un même lieu géographique. C'est un processus long et difficile, qui met en jeu des connaissances empiriques accumulées par l'expert géomaticien, et qu'il est peu envisageable d'appliquer pour traiter une grande masse de données.

La recherche sur les ontologies occupe quant à elle une place centrale dans les travaux actuels sur le Web sémantique. Les ontologies sont des ressources particulières du Web, qui ont vocation à fournir l'interprétation sémantique à donner aux annotations des autres ressources. Elles constituent donc l'artefact qui doit permettre de manipuler des ressources par leur contenu.

Dans ce contexte a débuté en à l'été 2004 l'Action Concertée Incitative FoDoMuST, qui réunit les laboratoires Image et Ville (LIV, Strasbourg), des Sciences de l'Image, de l'Informatique et de la Télé-détection (LSIIT, Strasbourg) et d'Ingénierie des Connaissances (ERIC, Lyon), avec pour objectif d'extraire et de qualifier la végétation urbaine sur des images satellites multi-spectrales et multi-résolution.

Notre étude s'inscrit dans ce projet. Elle répond à un besoin, qui est d'évaluer le rôle que pourrait jouer une ontologie dans l'automatisation — au moins partielle — du processus de photointerprétation des images satellites. Outre une réflexion sur ce que devrait contenir une telle ontologie, nos travaux ont visé à identifier la manière dont elle pourrait s'articuler au sein du projet. Dans cette optique, les principaux objectifs poursuivis ont été la proposition d'un modèle conceptuel pour cette ontologie, l'examen des modalités de son opérationnalisation, puis une réflexion sur les possibilités d'enrichissement de cette ontologie par des processus d'extraction de connaissances.

Nous ferons donc dans un premier temps un point sur l'existant en matière d'ontologies géographiques, de façon à mettre à jour dans la littérature les différents travaux pouvant servir de base à la création d'une ontologie pour FoDoMuST. Seront ensuite examinés les enjeux relatifs à sa conception et à son implémentation, à la suite de quoi un prototype d'ontologie sera proposé conformément aux besoins exprimés par les acteurs du projet. Enfin, la troisième partie dégagera quelques perspectives d'enrichissement de cette ontologie pour aboutir à la proposition d'un premier processus de fouille.

Première partie

Contexte

1 Le projet FoDoMuST

1.1 Objectifs de l'ACI

L'ACI FoDoMuST¹ a pour objet la définition de méthodes et d'outils permettant une utilisation conjointe de plusieurs sources de connaissances et d'images lors de l'identification, la localisation et la formalisation des éléments du tissu urbain. Elle se donne comme thématique d'étude la végétation en milieu urbain, car celle-ci offre suffisamment de variabilité pour servir de référence à l'étude. Son objectif est alors de proposer un processus complet de sélection, d'extraction et d'interprétation de connaissances à partir de bases de données d'images et de connaissances du domaine considéré en quatre phases :

- Structuration et organisation des données
- Construction des objets
- Classification multi-stratégie
- Aide à l'interprétation

1.2 Une masse de données

L'approche est motivée par la grande diversification de l'offre en images satellites depuis deux décennies, avec notamment l'arrivée des images satellites très haute résolution et des capteurs hyperspectraux qui disposent de plusieurs centaines de longueurs d'ondes. L'utilisateur doit alors faire face à une profusion d'images satellites, provenant de différents capteurs, et variant par leurs résolutions spectrales, spatiales et temporelles. On cherchera alors à exploiter ces données complexes pour reproduire l'approche multi-stratégie développée par l'expert lorsqu'il interprète les images (lorsqu'il procède par emboîtement des échelles de résolution ou par intégration des abords des objets par exemple).

1.3 Une ontologie pour formaliser les connaissances du domaine

Si les bases de données d'images sont une source très riche de connaissances, elles ne permettent pas de donner un sens aux régions identifiées sur les images. Elles doivent pour ce faire être complétées par une formalisation des connaissances du domaine, qui intègre notamment différents objets urbains (parc, alignement d'arbres, jardin ouvrier...). On s'appuiera sur ces connaissances pour qualifier les régions obtenues au terme d'un processus de traitement des images (figure 1).

Se tournant vers les travaux effectués en intelligence artificielle, et plus particulièrement dans sa branche représentation des connaissances, les chercheurs ont donc envisagé de bâtir une ontologie pour représenter ces connaissances et de s'en servir ensuite dans l'interprétation des images.

2 Cahier des charges

L'ontologie de FoDoMuST a vocation à automatiser au moins partiellement le processus de photointerprétation sur des images satellites très haute résolution. Elle fournit le vocabulaire avec lequel

¹Pour Fouille de Données Multi-Stratégie, <http://isiit.u-strasbg.fr/afd/fodomust/>.

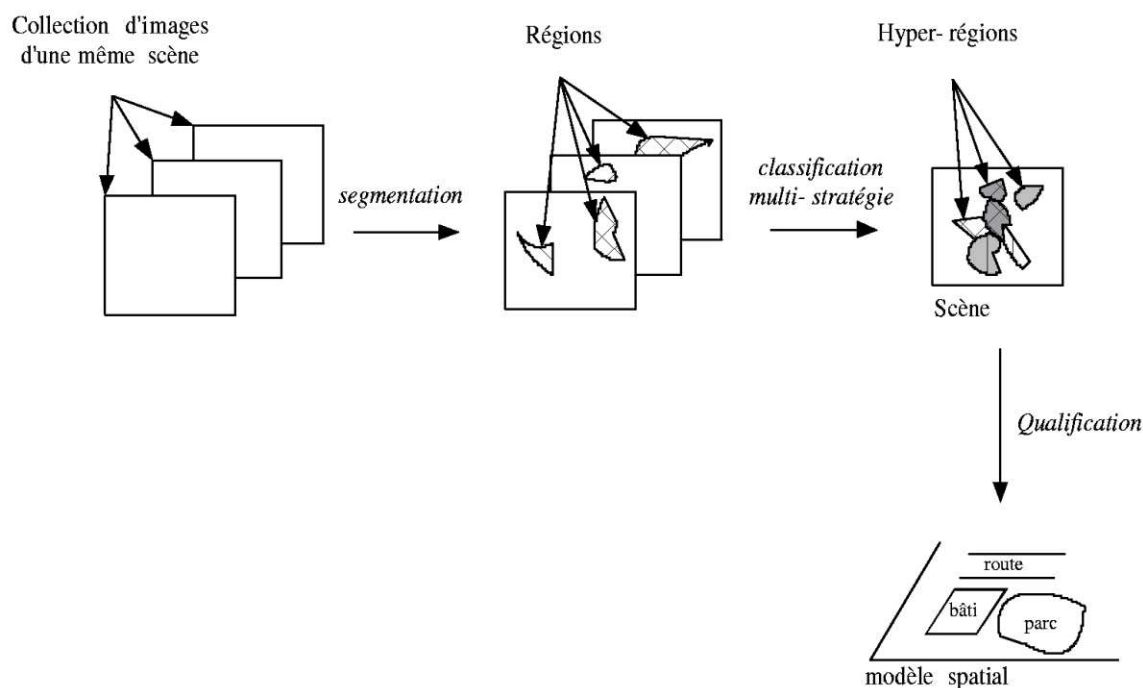


FIG. 1 – La chaîne de traitement des images satellites.

seront étiquetées les différentes régions extraites de ces images. L'ontologie sera parcourue pour proposer à l'expert les objets urbains dont la définition est cohérente avec les caractéristiques de la région identifiée sur les images. Il est donc indispensable de disposer dans l'ontologie non seulement d'une **description des différents objets urbains** que l'on est susceptible de retrouver sur les images, mais aussi d'un **mécanisme d'inférence efficace** sur ces descriptions.

De plus, son usage impose de **conserver les descriptions des objets reconnus** tout au long du processus de reconnaissance. En effet, l'ontologie ne peut être dissociée des instances des objets urbains déjà reconnus sur les images, dans la mesure où les objets géographiques entretiennent de fortes dépendances entre eux (relations spatiales d'adjacence et de composition entre objets, occurrences d'un même objet sous différentes formes à plusieurs résolutions et à différentes dates). Identifier les objets urbains d'un lieu géographique donné impose donc de maintenir une base de connaissances tout au long du processus de reconnaissance. L'analyse des instances déjà reconnues peut en sus permettre l'acquisition de connaissances décisionnelles qui viendraient compléter celles déjà présentes dans l'ontologie (voir à ce sujet la partie IV ci-après).

En outre, il est un certain nombre de qualités que l'on attend de l'ontologie de FoDoMuST :

Posséder un langage de description des objets suffisamment riche

La richesse du langage doit être suffisante pour traduire avec précision la connaissance de l'expert. Elle doit refléter le plus fidèlement possible la conceptualisation du domaine recueillie auprès de celui-ci. Doivent être exprimées à la fois les caractéristiques physiques des concepts (caractéristiques spectrales, résolutions d'observation, indices de forme, de texture, etc.), et les relations sémantiques qu'ils entretiennent entre eux (relations spatiales, de composition, de spécialisation, etc.).

Être extensible

L'ontologie doit pouvoir être aisément modifiée ou étendue par l'ajout de nouvelles connaissances.

Être réutilisable

L'ontologie doit être conçue pour être réutilisable quel que soit le lieu géographique étudié, et fournir des garanties d'interopérabilité avec les systèmes informatiques susceptibles de l'utiliser.

Être compréhensible par l'expert

Enfin il est indispensable que l'ontologie soit intelligible pour l'expert et que celui-ci puisse exercer un contrôle sur elle et sur la base de connaissance, car lui seul est à-même de valider la définition d'un concept ou la reconnaissance d'un objet.

Deuxième partie

Les ontologies

3 Qu'est-ce qu'une ontologie ?

Définir ce qu'est une ontologie n'est pas chose aisée. En effet rares sont les termes dont la définition fasse l'objet de si peu de consensus parmi les spécialistes. De manière générale, on peut dire qu'une ontologie est une *représentation des connaissances d'un domaine*. Seront bien sûr représentées dans FoDoMuST des connaissances du géographe.

3.1 Un artefact informatique

On peut affirmer, en reprenant Chantal Reynaud, qu'une ontologie est avant tout un produit d'ingénierie, composé d'un certain vocabulaire pour décrire une réalité assortie d'hypothèses explicites sur le sens des termes.

Elle reflète une compréhension partagée d'un domaine, et est le produit d'un consensus parmi une communauté de spécialistes.

C'est également une spécification formelle d'un certain domaine. Elle constitue un modèle formel, fait pour être exploité par une machine.

Les ontologies servent principalement à partager la connaissance. Elles sont, en complément du protocole de communication et du format de représentation des données une composante essentielle pour permettre l'interopérabilité des systèmes informatiques (SIG par exemple). Elles sont actuellement un enjeu pour la conception de services Web efficaces.

Les ontologies permettent en outre la réutilisation des connaissances.

3.2 Une représentation...

Le terme ontologie tient ses origines d'une partie de la philosophie, l'Ontologie (avec un grand O par opposition à l'artefact informatique), qui depuis le XVII^e siècle, désigne l'étude de l'Être, en référence à la philosophie aristotélicienne. En Intelligence Artificielle, ce qui est est ce qui peut être représenté.

Le philosophe Pierre Jacob définit une représentation comme un dispositif physique qui a pour fonction de véhiculer une information. Selon Patrick Winston, une *représentation* est également un ensemble de conventions pour décrire un groupe d'objets [Win88]. Elle est le résultat d'une conceptualisation, et s'exprime au moyen d'un langage.

3.3 ...des connaissances

La *connaissance* désigne les rapports entre un sujet et un objet [Ta88][Gan91]. On ne peut agir sur cette connaissance que lorsqu'elle devient objectivable, c'est-à-dire lorsqu'elle peut être exprimée par un langage. Les connaissances géographiques sont facilement accessibles en langage naturel, dans un livre de géographie ou en questionnant un géographe. Mais en intelligence artificielle, on veut de plus que ces connaissances soient exploitables par une machine. Elles doivent donc être formalisées, et l'on doit apprendre à la machine à leur donner un sens et à agir sur elles. On considérera en intelligence artificielle la connaissance comme ce qui fait l'objet d'une représentation.

3.4 Les briques de la maison ontologie

Nous revenons dans cette partie sur les principales entités constitutives, selon nous, d'une ontologie. Parmi celles-ci on trouve principalement des concepts, des relations entre concepts et des axiomes [TK04].

Concepts

D'après Edward E. Smith ([Smi89], repris dans [GK03]), un *concept* est "une représentation mentale d'une classe ou d'un individu et a trait à ce qui est représenté et à la façon dont l'information est typiquement utilisée pour la catégorisation". On distingue en effet ici le concept, qui désigne une idée mentale, d'une catégorie, qui désigne l'ensemble des référents de ce concept. Un *lexique* vient le plus souvent documenter les différents concepts ; à chacun d'eux est associée une définition en langue naturelle qui fixe le sens qu'il a dans le domaine considéré.

Il existe diverses manières de définir un concept. Un des courants de la psychologie des concepts, dite *exemplariste*, considère qu'un concept est défini par l'ensemble des objets qu'il dénote (ici les instances)[Mur02][GK03]. Nous nous placerons implicitement dans ce cadre pour développer notre ontologie, bien qu'une définition intentionnelle empirique soient associée à chaque concept pour guider sa reconnaissance sur les images.

De plus, on veut pouvoir raisonner sur des informations incomplètes [CEM75], et par suite on doit faire l'hypothèse dite du *monde ouvert*, dans laquelle on admet qu'un fait dont on ne dispose pas n'est pas forcément faux. Par exemple, si l'on définit la ville comme une localité de plus de 10000 habitants et que l'on dispose d'une liste des villes de France, constater qu'une ville est absente de notre liste ne prouve pas forcément que cette ville n'est pas une ville de France.

Relations

On peut distinguer, selon E. Tomai et M. Kavouras dans [TK04], trois types de relations :

- Les *relations sémantiques entre concepts*, comme la relation de spécialisation entre concepts, qui traduit par exemple qu'un chien est un animal, ou qu'une automobile est un véhicule. Dans le cas des ontologies géographiques, ce seront souvent des relations spatiales comme la relation de partie à tout, d'adjacence, de localisation.
- Les *propriétés inhérentes à un concept*, qui définissent des attributs pour un concept et indiquent les différentes valeurs que celui-ci peut prendre. Ce sera par exemple un attribut de résolution qui définit pour un objet géographique à quelles résolutions il peut être détecté.
- Les *relations entre relations*, qui permettent de construire une taxonomie de relations. On pourra ainsi par exemple définir une relation générique *relationSpatiale*, qui aurait plusieurs sous-relations comme *relationMereologique* ou *localisation*.

Axiomes

Une ontologie est en outre composée d'axiomes, qui forment des contraintes sémantiques pour le raisonnement et donnent un accompte d'une conceptualisation. Ils prennent la forme d'une théorie logique. Un exemple d'une telle théorie serait la théorie (O) ci-dessus relative aux cubes².

L'ensemble formé par les concepts, leurs instances, les relations entre les concepts et les axiomes les contraignant constitue une *base de connaissance*.

²Le terme axiome désigne souvent par extension toutes les définitions de concepts et de relations entre concepts de la base de connaissances.

3.5 Ontologies et bases de données

La communauté des bases de données s'est beaucoup intéressé aux moyens de concevoir des ontologies en adaptant les méthodes de conception de schémas conceptuels. Un exemple de modèle conceptuel spatio-temporel représentatif de l'état de l'art en matière de conception d'ontologies géographiques est MADS [PSZ99] [PSZ⁺97] et une discussion détaillée des atouts et limitations de l'approche bases de données peut être trouvée dans [CPSV03]. Nous pensons que cette démarche est vouée à l'échec, et ce pour un certain nombre de raisons :

- Il est impossible de raisonner sur les concepts et instances de l'ontologie, ni même d'explorer le schéma. Seule peut être effectuée l'interrogation des instances.
- Un schéma de bases de données n'est pas évolutif. Il est défini une fois pour toute avant la définition des instances. Une ontologie doit au contraire être vivante, et la mise à jour des instances doit pouvoir se faire sans une connaissance complète de l'ontologie. Cela impose de disposer de mécanismes de vérification dynamique de la cohérence du schéma et des instances.
- Les bases de données font l'hypothèse du monde clos.

Ces remarques révèlent la différence de nature qui existe entre une ontologie et une base de données. En effet, comme le remarque F Fonseca dans [FEAC02], un schéma de base de données représente ce qui est stocké dans la base, alors qu'une ontologie représente une vue de ce qui existe (ou peut exister) dans le monde réel.

Néanmoins, les modèles conceptuels offrent une représentation sémantiquement riche, ce qui fait d'eux de bons outils pour exprimer la structure des objets. Ils ont aussi l'avantage d'être particulièrement lisibles. En outre, la communauté bases de données bénéficie d'une expérience historique dans la modélisation du réel, et de nombreux outils sont disponibles à cet égard. C'est, nous semble-t-il, le point le plus intéressant de cette approche. On peut par exemple envisager d'utiliser le langage de modélisation UML pour concevoir visuellement une ontologie [PC99][BVEL04], puis utiliser des outils logiciels (comme par exemple le convertisseur UMLtoOWL³[GDDD04]) pour générer automatiquement du code OWL à partir de ceux-ci. Ce code pourra être donné en entrée à un éditeur comme Protégé.

4 Conception d'une ontologie

4.1 La démarche de conception

La conception d'une ontologie se réalise en trois étapes [Für02] :

– Conceptualisation

Cette étape s'effectue en langage naturel, et est réalisée auprès de l'expert ou par analyse de la littérature. Elle consiste à identifier les connaissances d'un domaine. Celui-ci est d'abord délimité. Puis les termes du domaine sont définis, ainsi que leurs propriétés, les relations qu'ils entretiennent entre eux, les contraintes qui s'appliquent sur eux, etc. et leur sens est précisé. Un modèle conceptuel est bâti à partir de cette étude. Cette étape est de loin la plus longue et la plus délicate puisqu'elle constitue un travail multi-disciplinaire. Elle nécessite en particulier une entente sur les termes employés entre l'ingénieur de la connaissance et le géographe.

– Ontologisation

L'étape d'ontologisation est une étape de spécification de la conceptualisation obtenue. Son objectif est de - partiellement au moins - la formaliser.

D'après T. Gruber dans [Gru93], l'ontologie obtenue doit satisfaire les principes de :

- Cohérence : les concepts doivent pouvoir admettre des instances.

³<http://afrodita.rcub.bg.ac.yu/gasevic/projects/UMLtoOWL/>

- Extensibilité : l'introduction d'un nouveau concept ne doit pas entraver la révision des concepts déjà présents.
- Biais minimal du point de vue syntaxique : la conceptualisation doit se faire au *knowledge level*⁴, et doit être indépendante de toute implémentation.
- Biais minimal du point de vue sémantique : l'engagement ontologique doit être minimal, tout en permettant le partage des connaissances. La théorie logique doit donc être la plus petite possible.
- Opérationnalisation
L'opérationnalisation sert enfin à formaliser complètement l'ontologie, et à la doter de capacités de raisonnement suivant l'utilisation que l'on veut en faire (raisonnement spatio-temporel sur les objets urbains par exemple, test d'instanciation...). Suivant le formalisme adopté et les besoins, cette étape n'est pas toujours nécessaire.

4.2 Conceptualisation

4.2.1 Une spécification d'une conceptualisation

Une *conceptualisation* est, selon Thomas Gruber, une vue abstraite, simplifiée d'une partie du monde que l'on veut représenter. Il définit alors dans [Gru93], et c'est la définition qui est généralement retenue, une ontologie comme une spécification explicite d'une conceptualisation.

Nicola Guarino dans [GG95] raffine cette définition en affirmant qu'une conceptualisation n'est pas tant une vue de ce qu'une partie du monde est, mais de ce que cette partie du monde *peut* être.

Guarino suggère alors plutôt de voir une ontologie comme une théorie logique, composée d'axiomes, donc d'énoncés toujours vrais, qui expriment des règles implicites contraignant la structure d'une partie de la réalité. Cette théorie donnerait un accompte partiel et explicite d'une conceptualisation.

Il donne l'exemple d'un ensemble de cubes posé sur une table (figure 2).

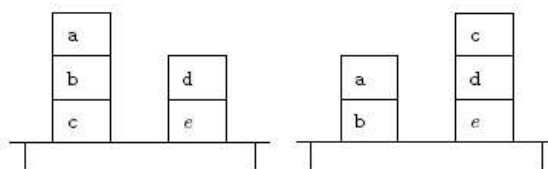


FIG. 2 – Différents arrangements de cubes sur une table.

Ces deux arrangements correspondent à une même conceptualisation, mais traduisent différents mondes possibles. On peut conceptualiser ces arrangements en considérant les cubes et les relations qu'ils peuvent avoir entre eux et avec la table. L'ensemble des objets qui peuvent être représentés, appelé domaine du discours, est ici l'ensemble formé des cubes a, b, c, d et e. Les relations qu'ils entretiennent entre eux sont par exemple la relation sur, qui traduit le fait qu'un cube est posé sur un autre cube, et la relation table, qui traduit le fait qu'un cube est posé sur la table.

4.2.2 Engagement ontologique

Lorsqu'on représente un objet, le choix même des conventions que l'on fait pour le décrire impose un *engagement ontologique*. Celui-ci intervient dès qu'on se pose la question : en quels termes dois-je penser le monde [DSS93]? En effet, toute représentation est un substitut d'un objet réel. Loin

⁴Proposé par A. Newell en 1982, l'hypothèse du *knowledge level* ([New82] présenté dans [Gan91]) suppose que la connaissance a un statut indépendant de toute représentation ou implémentation. Vue comme une compétence ouvrant des possibilités d'action, elle s'exprimerait en termes de buts, d'actions et de stratégies d'un agent rationnel.

de traduire l'immense complexité de l'objet réel auquel il réfère, ce substitut n'est qu'un point de vue sur cet objet, une vision simplifiée qui le réduit à quelques uns de ses aspects, pour les besoins de sa manipulation par une machine (un plan d'eau sera par exemple représenté par sa forme et sa réponse spectrale) Dès lors, choisir un langage de représentation est déjà un parti-pris. Opter alors pour un formalisme logique implique d'appréhender le monde comme un ensemble d'entités ayant des relations entre elles. Adopter un formalisme à base de cadres (*frames*, voir 4.4.1) suppose plutôt de voir le monde en termes de descriptions stéréotypiques, etc.

4.3 Formaliser le modèle conceptuel

4.3.1 Un langage

En accord avec cette conceptualisation, chaque arrangement peut être représenté dans un *langage*, qui consiste en un ensemble de signes (un vocabulaire, ici les variables d'individu a, b, c, d, e, les symboles de prédicats sur et table et le connecteur \wedge), ainsi que des règles pour les assembler (disant par exemple que si deux formules Φ et Ψ sont bien construites, alors $\Phi \wedge \Psi$ l'est aussi).

Dans ce langage l'arrangement de gauche serait exprimé par la formule :

$$\text{sur}(a,b) \wedge \text{sur}(b,c) \wedge \text{table}(c) \wedge \text{sur}(d,e) \wedge \text{table}(e)$$

Une ontologie permet alors de caractériser partiellement une telle conceptualisation par un ensemble d'énoncés vrais dans tous les mondes possibles. Un exemple d'ontologie serait :

$$(O) \begin{cases} \neg \text{sur}(x,x) \\ \text{sur}(x,y) \Rightarrow \neg \text{table}(x) \\ \text{sur}(x,y) \Rightarrow \neg \text{sur}(y,x) \end{cases}$$

4.3.2 Une sémantique

Un formalisme de représentation des connaissances associe à un langage une *sémantique*, qui permet de faire le lien entre les représentations et les objets du monde réel auxquelles elles réfèrent.

La sémantique en intelligence artificielle peut se voir sous deux aspects complémentaires⁵ :

Désambiguation des mots et des phrases

Par elle on associe à un symbole l'idée dont il est le substitut, et ce de manière univoque (On se demandera par exemple : « que désigne-t-on au juste sous le terme ParcJardinSquare ? »).

Inférences autorisées

La sémantique d'un langage formel donne également les faits qu'on a le droit de déduire, en accord avec notre conceptualisation.

Si le travail de désambiguation des termes peut être effectué par la simple constitution d'une terminologie du domaine, il n'en va pas de même pour la spécification des inférences autorisées. De même que la sémantique linguistique nous dit qu'une phrase comme « je mange du vin », bien que syntaxiquement correcte, ne fait aucun sens, la sémantique d'un langage formel fixe, dans un monde réel ou hypothétique, les conditions de vérité et de fausseté d'une proposition. Elle détermine quelles assertions sont correctes. En ce sens elle rejoint la sémantique dans son acception en logique. En logique du premier ordre, en effet, celle-ci est liée à une fonction d'interprétation qui associe à tout symbole de prédicat donné et pour toute affectation donnée une valeur de vérité (par exemple dans l'exemple précédent, $\text{sur}(b,c)$ est vrai dans l'arrangement de gauche mais pas pour celui de droite). Cette interprétation est étendue à toute formule bien construite (par exemple $\text{sur}(x,y) \wedge \text{sur}(u,v)$ est vrai si $\text{sur}(x,y)$ est vrai et $\text{sur}(u,v)$ est vrai).

⁵Voir [Woo75] pour une discussion plus détaillée.

4.3.3 Un raisonnement

Enfin, une représentation de connaissances doit associer à un langage et une sémantique un *raisonnement*. Un raisonnement est un mécanisme d'inférence qui permet d'obtenir de nouvelles connaissances à partir de celles qu'on a déjà. Comme nous l'avons vu, les connaissances que l'on peut inférer sont contraintes par la sémantique du langage, qui se matérialise par un ensemble d'axiomes que doit satisfaire tout fait inféré.

Les bases de données ne sont donc pas des représentations des connaissances, car, excepté peut-être dans le cas des bases de données déductives, aucun fait (ici, un tuple) nouveau ne peut être obtenu à partir de ceux qu'on a déjà dans la base. Une ontologie, au contraire, doit permettre le raisonnement. Elle n'a pas vocation à être purement descriptive, même si l'on parle souvent d'ontologie terminologique pour désigner un simple thésaurus ou une taxonomie de concepts. Par opposition, on qualifie souvent les ontologies de formelles.

Outre l'impératif de correction (toute connaissance inférée doit être sémantiquement valide), il est souhaitable qu'un raisonnement ait les propriétés de calcul suivantes :

Complétude

Toute expression sémantiquement valide doit pouvoir être inférée.

Décidabilité

Le résultat d'une inférence doit pouvoir être obtenu en un temps fini.

Pour toutes ces raisons, la logique s'impose naturellement comme une base de notre formalisme de représentation des connaissances.

4.4 Les systèmes classificatoires

Nous avons vu qu'adopter un formalisme de représentation des connaissances c'est adopter une certaine façon de "voir le monde" à travers lui. Le choix d'un tel formalisme correspond à la recherche d'une adéquation entre cette vision du monde et notre conceptualisation du domaine. Entrent en jeu la richesse de ses constructeurs, ainsi que les différents raisonnements qu'il autorise ou prescrit.

Etant donné la nature hiérarchique des connaissances que nous avons à représenter dans FoDoMuST, nous tournerons ici vers les *systèmes classificatoires* (p.ex. représentation des connaissances par objet, logiques de description...). En suivant Roland Ducournau dans [Duc96], on peut dire qu'un système classificatoire est une hiérarchie de classes décrites en intension, et interprétées en extension.

Ils comportent donc deux composantes :

- les *classes*, qui sont des entités génériques du domaine, définies en intension par des propriétés (par une condition nécessaire et suffisante sur celles-ci par exemple). Elles matérialiseront donc les concepts des ontologies.
- les individus, vérifiant ces propriétés, dont on dit qu'ils sont des *instances* de la classe.

On pourra alors choisir une définition de la classe Ville (représentant le concept "ville") comme étant une localité de plus de 10000 habitants, et une instance de cette classe sera la ville de Lyon.

Parmi les systèmes classificatoires, deux formalismes sont particulièrement importants pour la conception d'ontologies : les cadres et les logiques de description.

4.4.1 Les cadres (*frames*)

La notion de *cadre* a été introduite en 1974 par Marvin Minsky ([Min74]) pour désigner une structure de données représentant une situation stéréotypique. Dans le cadre de ses travaux de vision par ordinateur, il cherchait à expliquer la formidable rapidité d'interprétation de l'homme face à une nouvelle

situation (un nouveau percept visuel, en particulier). Il suggéra qu'on interprète une nouvelle situation à partir d'une situation qu'on s'attend à voir, en se remémorant une structure (un cadre) de connaissances correspondant à une situation similaire. Seule une petite partie de l'information visuelle serait alors exploitée pour comparer cette situation au modèle prédit, adapter le cas échéant ce dernier à la réalité rencontrée et utiliser l'information qu'il contient pour entreprendre des actions. B.Kuipers dans [Kui75] donne l'exemple d'une personne entrant dans une maison qu'il ne connaît pas. Un « cadre de pièce » en tête, il s'attend à y trouver des murs, un sol, quelques meubles, des décorations, etc.

A l'origine, Minsky présentait une approche prototypique des cadres, dans laquelle chaque cadre représenterait un objet typique d'une catégorie (par exemple un pavillon typique serait caractérisé par une surface de 170m² et un indice de Morton de 0,6).

Mais les cadres ont principalement été utilisés dans une approche dite définitionnelle, dans laquelle on distingue les cadres qui représentent les catégories de ceux qui représentent les instances de ces catégories. C'est cette approche qui a donné naissance au paradigme orienté objet.

Un cadre est composé d'attributs (slots), dont chacun est décrit par un certain nombre de facettes. Un cadre 'végétation' comporterait par exemple un attribut 'surface', qui serait décrit par une facette de type spécifiant qu'il est un flottant, une facette spécifiant sa cardinalité (simple ici), une autre la plage de valeurs qu'elle peut prendre, etc. Une facette peut être déclarative ou procédurale. Si elle est déclarative, elle précise des contraintes sur les valeurs que peut prendre l'attribut (ou la valeur elle-même dans le cas des cadres d'instances), ou spécifie une valeur fixe (si c'est une constante commune à toutes les instances) ou une valeur par défaut pour un attribut. Si elle est procédurale, elle peut calculer la valeur d'un attribut à partir d'autres attributs, ou déclencher une action lors de l'accès à l'attribut (facette réflexe).

Les cadres définissent des classes d'appartenance qui sont organisées en taxonomie suivant une relation d'héritage. Cette relation d'héritage entre classes est une relation de spécialisation, de type "est un". Toute instance d'une classe est une instance de toutes ses super-classes, et hérite des caractéristiques de celles-ci. Ainsi un plan d'eau est une instance de la classe "Plan d'eau", mais aussi de ses super-classes "Eaux Continentales", "Surfaces en Eau" et \top (TOP).

Le raisonnement naturellement associé aux cadres est la classification d'instances, qui consiste à déterminer suivant les caractéristiques d'un nouvel individu la classe à laquelle il appartient. La théorie des cadres n'étant néanmoins fondée sur aucun modèle formel, aucune inférence n'est imposée et toute latitude est ici laissée au concepteur.

Ce dernier point proscrit en pratique tout raisonnement sur les cadres. Néanmoins, l'utilisation des cadres pour l'ontologie de FoDoMuST peut être envisagée, dans la mesure où ceux-ci sont ensuite exprimables dans un formalisme logique à des fins d'inférence - nous verrons que c'est le cas par exemple en OWL DL. Ceux-ci peuvent être utiles pour exprimer la variabilité des concepts.

4.4.2 Les logiques de descriptions

Les logiques de descriptions forment une famille de formalismes de représentation de connaissances qui tiennent leurs origines dans la logique du premier ordre, les cadres et les réseaux sémantiques. Dans ces formalismes les connaissances sont décrites par des concepts, des rôles (relations binaires entre concepts) et des instances de ces concepts. Si le document de référence sur ces formalismes est [BCM⁺03], une bonne introduction est donnée dans [Nap97].

Chaque formalisme se caractérise par un ensemble de concepts dits *primitifs* et un ensemble de

constructeurs faisant intervenir des *rôles* associés à ce concept et des *restrictions* sur ces rôles. A partir de ces éléments on peut construire des descriptions conceptuelles structurées, appelés concepts *définis*. Un exemple de concept défini pourra être le concept PARENT de description

$$PARENT \doteq HUMAIN \sqcap (\exists \text{enfant}.HUMAIN)$$

faisant intervenir le concept primitif HUMAIN, le rôle enfant et signifiant qu'un parent est un humain ayant la propriété d'avoir un enfant humain.

Dans les logiques de descriptions, les connaissances sont prises en compte à deux niveaux. Au niveau *terminologique* est défini le vocabulaire d'un domaine donné (*terminologie*, ou TBox), qui consiste en un ensemble fini de définitions de concepts (comme le concept PARENT). Les instances de ces concepts (ou *assertions*, comme PARENT(Marc)) sont quant à elles contenues dans une ABox, qui représente le niveau *factuel*. La réunion d'une TBox et d'une ABox forme une base de connaissances. Une relation de *subsumption* (ou de généralité) organise les concepts (et éventuellement aussi les rôles) hiérarchiquement.

La sémantique associée est celle de la théorie des modèles : un modèle est un couple $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, où $\Delta^{\mathcal{I}}$ est un ensemble appelé domaine d'interprétation et $\cdot^{\mathcal{I}}$ une fonction appelée fonction d'interprétation. Celle-ci fait correspondre :

- à tout concept C un sous-ensemble $C^{\mathcal{I}}$ de $\Delta^{\mathcal{I}}$,
- à tout rôle R un sous-ensemble $R^{\mathcal{I}}$ du produit cartésien $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$,
- et à toute instance a un élément $a^{\mathcal{I}}$ de $\Delta^{\mathcal{I}}$.

Les principaux raisonnements effectués sur la terminologie sont le *test de satisfiabilité* (ou cohérence) d'un concept et le test de *subsumption*.

Un concept est satisfiable s'il admet des instances, i.e. s'il existe une interprétation \mathcal{I} par laquelle ce concept a une extension non vide ($C^{\mathcal{I}} \neq \emptyset$). Un exemple de concept insatisfiable est $C \doteq A \sqcap \neg A$.

Un concept C subsume un concept D (i.e. est plus général que D, noté $D \sqsubseteq C$) si et seulement si son extension contient celle de D ($D^{\mathcal{I}} \subseteq C^{\mathcal{I}}$). Le test de subsumption est à la base du mécanisme de *classification*, qui consiste à déterminer dans la hiérarchie de concepts les subsumés les plus généraux et les subsumants les plus spécifiques d'un concept donné.

Les principaux raisonnements effectués sur le ABox sont le *test de cohérence* d'un ensemble d'assertions et le *test d'instanciation*.

Le test de cohérence consiste à déterminer si un ensemble d'assertions admet un modèle dans une terminologie (par exemple relativement à la description du concept PARENT l'assertion $\neg PARENT(Marc)$ serait incohérente avec les assertions $HUMAIN(Marc)$ et $(\exists \text{enfant}.HUMAIN)(Marc)$).

Le test d'instanciation consiste à déterminer les concepts les plus spécifiques dont un individu est l'instance.

Une des principales différences entre les logiques de descriptions et les formalismes de représentation des connaissances par objets⁶ (RCO, comme les cadres) réside dans l'approche définitionnelle des logiques de descriptions, qui contraste avec l'approche descriptive propre aux RCO. Un concept est défini en intension, par des conditions portant sur ses propriétés.

Pour FoDoMuST, ces formalismes constitueront un outil privilégié pour la réalisation d'inférences saines sur l'ontologie. De plus, ces logiques sont dotés de mécanismes de vérification d'intégrité qui autorisent une extension dynamique de l'ontologie sans risquer d'aboutir à des incohérences.

Néanmoins, la décidabilité des mécanismes d'inférence se fait au détriment de la richesse du langage.

⁶Voir [Duc96] pour une discussion plus détaillée.

comme nous l'avons vu précédemment (4.3.3), les mécanismes d'inférence doivent, pour être utilisables, satisfaire des contraintes de décidabilité. Ces contraintes sont généralement satisfaites par la restriction du nombre de constructeurs du langage.

5 Les ontologies géographiques

5.1 Définition

Une ontologie géographique représente des *concepts géographiques*, qui décrivent l'espace géographique ou des entités ou phénomènes de cet espace géographique.

Ce qui fait la particularité des concepts géographiques, c'est justement leur relation à l'espace, auxquels ils sont intrinsèquement liés [SM98], de telle sorte qu'ils ne peuvent être dissociés du territoire physique qu'ils occupent. En conséquence, décrire des concepts géographiques implique de s'appuyer sur des éléments de représentation spatiale comme les relations mérologique (relation de partie à tout), de topologie (notions de contact et de connexité), de localisation, de distance, d'orientation (alignement, positionnement relatif d'objets), de forme, etc. [CSV98] [BM00]

Suivant le type de concept qu'on souhaite représenter, on peut, en suivant F. Fonseca dans [FDC03], distinguer plusieurs types d'ontologies géographiques :

- Les ontologies de « haut niveau » (*top-level*), qui représentent les concepts les plus généraux liés à l'espace, comme les relations de partie à tout, de topologie, de géométrie. Sont également accessibles de nombreuses ontologies basées sur les différentes normes relatives à de l'information géographique⁷, comme par exemple la norme ISO 19111 :2003 qui définit le référencement par coordonnées spatiales.
- Les ontologies spécifiques à certains domaines, qui peuvent être comme dans notre cas des ontologies d'objets urbains, mais aussi des ontologies hydrologiques⁸ comme hu⁹ décrivant des unités hydrologiques (rivière, lac, etc.) ou des ontologies de phénomènes physiques¹⁰ comme celles développées par les chercheurs de la NASA au cours du projet SWEET¹¹.
- Les ontologies spécifiques à certaines tâches, comme l'interprétation d'images ou l'évaluation de la pollution sonore.

5.2 Régions de l'espace ou objets dans l'espace ?

Pour mieux appréhender la nature des concepts géographiques, il est utile de considérer les deux points de vues sur la réalité géographique que constitue la dichotomie objet / terrain ([Goo92]) :

Le modèle objet

Une première façon de conceptualiser l'espace est de le voir comme un contenant, c'est à dire une surface sur laquelle sont disposés un nombre fini d'objets. Chaque objet peut alors être caractérisé par un certain nombre d'attributs, et la localisation par un ensemble de tuples $\langle x, y,$

⁷Une liste de ces ontologies peut être trouvée à l'url <http://loki.cae.drexel.edu/wbs/ontology/list.htm>.

⁸Des exemples d'ontologies hydrologiques peuvent être trouvées sur la page de Luis Bermudez à l'url <http://loki.cae.drexel.edu/how/>.

⁹Une version OWL Lite de cette ontologie peut être accédée depuis l'espace de nommage <http://loki.cae.drexel.edu/how/HydrologicUnits/2003/09/hu#>.

¹⁰Voir par exemple l'ontologie <http://sweet.jpl.nasa.gov/ontology/phenomena.owl>.

¹¹*Semantic Web for Earth and Environmental Technology*. Une description du projet et un grand nombre d'ontologies sont accessibles à l'url <http://sweet.jpl.nasa.gov/index.htm>.

o_1, o_2, \dots, o_n où les o_i sont des booléens indiquant si la présence ou non de l'objet $n^{\circ}i$ au point de coordonnées (x, y) dans le plan.

Le modèle terrain (*field model*)

Une autre façon de conceptualiser l'espace est de le voir comme un ensemble de distributions spatiales continues sur lesquelles on peut agir en les délimitant en régions. Un terrain est alors assimilable à une fonction de correspondance $\phi_\lambda : R \rightarrow V$ entre une région R de la surface terrestre et un ensemble V de variables géographiques. Le modèle terrain est souvent employé pour modéliser des phénomènes comme le climat, ou des densités de population, de végétation.

5.3 Objets microphysiques, objets mésoscopiques

Une autre distinction très utile est celle évoquée dans [SM98] entre objets microphysiques et objets mésoscopiques.

objets microphysiques

Un objet microphysique est en quelque sorte un quantum de réalité physique, un "édifice complexe de molécules", comme de l'eau, du béton, de la végétation. Il est — en principe — purement physique, donc indépendant de tout jugement cognitif.

objets mésoscopiques

Un objet mésoscopique, au contraire, est un assemblage d'objets microphysiques qui n'existe qu'en vertu de l'interprétation que le géographe en fait. Le concept d'emprise hospitalière, par exemple, ne correspond pas à une réalité physique particulière, mais plutôt à une composition de différentes réalités physiques dont l'unité est due au fait que le géographe a reconnu à un endroit donné la projection de ce concept géographique sur le plan spatial.

On distingue plusieurs types d'objets mésoscopiques suivant le type de bordure que le géographe est susceptible de lui attribuer. Lorsque cette bordure correspond à une discontinuité physique, comme c'est le cas pour un lac ou une route, on parle d'objet *Bona Fide*. Lorsqu'au contraire la bordure est indépendante de toute discontinuité physique, et est fixée arbitrairement par l'homme comme peut l'être la frontière d'un pays, alors on parle plutôt d'objet *Fiat* ([SV00]).

La nature même des objets Fiat rend leur localisation et leur modélisation délicate. En effet si l'homme est capable de reconnaître une forêt lorsqu'il en rencontre une et de lui attribuer un contour, une machine aura toutes les difficultés à déterminer si une distribution d'arbres peut être considérée comme une forêt et, le cas échéant, à délimiter celle-ci (voir par exemple le papier de B. Bennett, [Ben01], sur l'imprécision des concepts géographiques).

5.4 Revue de presse

Depuis quelques années, les acteurs de l'information géographique montrent un intérêt croissant pour les ontologies géographiques et leur utilisation dans les SIG.

Réaliser un état de l'art en matière d'ontologies géographiques était le but de l'atelier dédié aux géo-ontologies (Geo-Ontology Workshop¹², voir le document [Har02]) organisé à l'automne 2002 par Brandon Bennett (Université de Leeds) et Jenny Harding (*Ordnance Survey*), et qui fut suivi en automne 2003 d'un autre atelier intitulé "Fundamental Issues in Spatial and Geographic Ontologies"¹³

¹²<http://www.comp.leeds.ac.uk/brandon/geo-ontology/>

¹³<http://www.comp.leeds.ac.uk/brandon/cosit03ontology/>

organisé en collaboration avec le COSIT¹⁴.

En 2000, le Consortium Universitaire américain pour les Sciences de l'Information Géographique¹⁵, qui regroupe les universités du Maine, de Californie et de New York, a fait de la recherche sur les ontologies géographiques une de ses priorités. L'UCGIS prend en charge la coordination d'un axe de recherche, appelé "Fondations Ontologiques de l'Information Géographique" [MEHS00], qui se donne pour objectif à terme de développer une ontologie pour chaque phénomène géographique. L'idée est de concevoir d'abord une ontologie de 'haut niveau' (upper-level) regroupant les concepts les plus généraux de la géographie, puis de développer conjointement une ontologie des concepts de la géographie naïve, une ontologie de l'imprécision (*vagueness*) et une ontologie d'échelle, qui prenne notamment en compte les aspects liés à la multi-résolution.

Les activités de recherche du groupe OntoGeo¹⁶ de l'université technique nationale d'Athènes sont également exclusivement dédiées aux ontologies géographiques. Les aspects théoriques sont abordés (géographie naïve, granularité et imprécision), ainsi que les aspects conception et intégration d'ontologies.

Plusieurs revues ont édité des numéros spéciaux consacrés aux ontologies géographiques. Edité par le philosophe Achille C. Varzi, le volume 20 :2 de septembre 2001 de la revue *Topoi*¹⁷, intitulé "Philosophie de la Géographie", questionne la nature des objets géographiques. Dans le même temps l'*International Journal of Geographical Information Science*¹⁸ édite un volume (le 15 :7 d'octobre 2001) pour tenter de clarifier le rôle que celles-ci devront occuper dans les sciences de l'information géographique.

5.5 Le modèle conceptuel de SPRING

SPRING [CSF96] est un logiciel développé à l'institut national brésilien de recherche spatiale dans le but de suivre l'évolution de la déforestation de la forêt amazonienne. Bien qu'il suive une approche base de données, SPRING propose une architecture pour intégrer des données satellitaires dans un système d'information géographique dont il peut être bon de s'inspirer.

En effet, SPRING reprend et fait coexister dans son modèle conceptuel les deux classes de modèles de l'information géographique que sont le modèle objet et le modèle terrain. Dans ce modèle, une classe terrain représente une variable géographique définie sur une région de la surface de la Terre. Cette classe est déclinée en deux sous-classes suivant la nature de cette variable. L'une d'elles représente les variables thématiques, nominales, et l'autre les variables numériques des capteurs. Le lien est alors fait avec les objets géographiques par l'introduction d'objets "cadastraux", qui sont en fait des fonctions de correspondance associées à une région donnée, à une résolution donnée et à un mode de représentation donné (raster ou vecteur), et qui associent à tout objet géographique une représentation sur cette région.

¹⁴*Conference On Spatial Information Theory*. Cette conférence biannuelle pluridisciplinaire autour de la représentation de l'espace, et en particulier de l'espace géographique.

¹⁵UCGIS, <http://www.ucgis.org/>

¹⁶<http://ontogeo.ntua.gr/>

¹⁷<http://springerlink.metapress.com/>

¹⁸<http://taylorandfrancis.metapress.com/>

5.6 Etat de l'art : bilan

De nombreux travaux ont donc été réalisés sur les ontologies géographiques. Mais ces travaux, émanant tant de spécialistes des sciences de l'information géographique que de l'informatique ou de la philosophie sont relativement récents et restent en général très théoriques. Peu d'ontologies ont effectivement été développées, et encore moins ont été effectivement opérationnalisées. La plupart des ontologies géographiques mises au point ont une vocation purement terminologique. Et parmi celles celles qui ont été intégrées dans un SIG, seules ont été proposées des ontologies suivant des approches issues des bases de données comme celles de MADS ou de SPRING.

Par ailleurs, les choix de conception d'une ontologie sont très liés à l'utilisation que l'on compte en faire. On peut développer des ontologies pour décrire les concepts les plus généraux d'un domaine, comme l'espace et le temps. Mais on peut aussi, et c'est ce qui nous intéresse ici, créer une ontologie pour chaque tâche à effectuer. Lors du recueillement des connaissances de l'expert, on privilégie dans ce cas dans la représentation des concepts l'un ou l'autre de leurs aspects, auquel on les réduit (engagement ontologique). Puis on choisit pour les représenter un langage qui permette d'appliquer sur eux et sur leurs instances les raisonnements souhaités.

Pour FoDoMuST, nous avons besoin de représenter les objets urbains sous l'angle de leur perception par un satellite, c'est à dire de les réduire à certaines de leurs propriétés comme leur réponse spectrale (telle qu'un capteur peut la mesurer) ou leur localisation à la surface terrestre. Par ailleurs, l'ontologie doit être conçue pour être opérationnalisée car l'objectif est de s'en servir pour reconnaître ces objets sur les images satellites.

Troisième partie

L'ontologie de FoDoMuST

6 Conceptualisation

Notre point de départ pour la conception de l'ontologie de FoDoMuST a été une ébauche d'ontologie réalisée sous Protégé 2000 par des étudiantes et qui était inspirée de la nomenclature Corine Land Cover. Ce travail, s'il introduisait l'éditeur et un bon nombre de classes thématiques d'objets géographiques, a néanmoins été complètement remanié au cours de l'étape de conceptualisation.

Nos travaux se sont appuyés sur la thèse d'Anne Puissant [Pui03] et sur un certain nombre de documents textuels fournis par les géographes comme des tableaux ou des définitions terminologiques. De nombreuses rencontres ont également été organisées avec les experts géographes et de télédétection, qui avaient pour but d'identifier les connaissances devant figurer dans l'ontologie, de préciser l'articulation sémantique des concepts du domaine (relations d'héritage ou de composition par exemple, différence entre occupation et utilisation du sol, etc.) et à définir le rôle de l'ontologie dans le projet FoDoMuST. Tout ce matériel a alors servi de support pour la proposition de plusieurs modèles conceptuels, que les nombreuses remarques et corrections de l'expert ont permis à chaque fois d'affiner. Le modèle conceptuel présenté ici est donc le résultat de ce processus.

6.1 Les images

Pour bâtir le modèle conceptuel, une première réflexion doit s'articuler autour de la nature des différentes images sur lesquelles on souhaite extraire les objets géographiques. Comme nous l'avons vu plus haut, Câmara et al. considèrent dans SPRING les images satellites comme des sous-classes de terrains. Dans ce modèle l'image est réduite à sa fonction d'échantillonnage par le capteur de l'énergie reflétée.

Mais, comme ce même G. Câmara le remarque dans [CEFM01], on a souvent besoin de dépasser ce postulat, étant donné la complexité de la structure des images. De ce fait, on en vient plutôt à travailler sur des objets extraits de ces images. En effet, le satellite fournit pour chaque *scène*, c'est à dire pour chaque prise de vue relevée à un instant donné, plusieurs images (dites *images brutes*), de résolutions différentes, et correspondant chacune à un canal spectral. De plus, on peut à partir de ces images brutes produire pour une même scène, par des processus d'analyse d'image ou par l'application de transformations, des *images calculées*, qui sont des indices tel l'indice de végétation. Nous ferons donc référence à ces mesures brutes ou calculées sous le terme de *paramètres de mesure des capteurs*. Avant d'être exploitées par l'expert en télédétection qui effectue le travail de photo-interprétation, les scènes sont donc segmentées par des méthodes d'analyse d'images, pour en extraire des ensembles de pixels homogènes du point de vue des paramètres de mesure des capteurs. Ces régions de la scène sont ensuite classifiées, et de cette classification sont extraits des objets, généralement représentés par des polygones (représentation vectorielle).

Néanmoins, ces objets ne correspondent encore à aucun objet géographique. Ils n'ont aucune existence réelle mais correspondent seulement à des valeurs homogènes de paramètres de mesure relevés à une résolution donnée sur une certaine zone de la surface terrestre.

6.2 Différents types d'objets géographiques

Examinons maintenant les différents objets géographiques dont nous avons une représentation sur les images. Ils peuvent être de trois types :

Les objets micro

Les objets micro constituent des zones de la surface terrestre de caractéristiques physiques homogènes. Ils identifient des matières comme l'eau, l'asphalte, la végétation ou la tuile sur les toits. Suivant qu'on ait ou non déterminé l'appartenance fonctionnelle d'un objet micro, on parlera d'*objet microfonctionnel* ou d'*objet microphysique*. On pourra par exemple dire de l'objet microphysique bâtiment que c'est un pavillon s'il appartient à un tissu urbain pavillonnaire, d'une parcelle de végétation qu'elle correspond à un jardin ouvrier si elle est située en centre ville. L'objet microphysique décrit l'occupation du sol, alors que l'objet microfonctionnel décrit son utilisation.

Il est nécessaire de se donner une échelle de référence pour l'extraction des objets microphysiques, de façon à obtenir des zones délimitées et de caractéristiques homogènes. En effet il n'est pas à priori d'échantillon irréductible de matière (sinon peut-être l'atome). Nous suivrons ici A. Puissant, qui introduit dans [Pui03] un certain nombre d'*objets élémentaires* révélateurs de la forme urbaine. Ces objets, qui joueront le rôle d'objets microphysiques 'atomiques', sont identifiables à une résolution de l'ordre du mètre. Ce sont le réseau viaire (voirie et chemin de fer), le bâti, la végétation intra-urbaine, le réseau hydrographique et les espaces libres.

Les objets méso

Les objets méso sont des objets mésoscopiques qui résultent d'une organisation spatiale des objets micro. Ce peuvent être par exemple un alignement d'arbres, ou un groupement d'arbres.

Les objets macro

Les objets macro sont des objets mésoscopiques qui résultent d'une organisation spatiale d'objets méso. Un exemple d'objet macro est un parc, qui est entre autres composé de plusieurs alignements d'arbres et de groupements d'arbres.

Les régularités extraites des images correspondent chacune à un de ces trois types d'objets géographiques du monde réel. En raison de la problématique inhérente au projet FoDoMuST, nous n'envisagerons dans l'ontologie que des objets urbains, c'est à dire des objets géographiques constitutifs de l'objet géographique qu'est la ville. Ce peuvent être un type d'habitat, un plan d'eau, un parc, une route, une gare [Pui03].

Chaque objet géographique est défini par un certain nombre de caractéristiques, qui peuvent être des valeurs de paramètres de mesure à une résolution donnée ou des indicateurs comme la densité de végétation ou de bâti calculés au sein d'un objet.

6.3 La photointerprétation

Dans ce contexte, on peut décomposer le processus d'interprétation des régions des images réalisé par l'expert en deux phases, comme le montre la figure 3 : d'abord la reconstitution de *zones géographiques*, dont la localisation est évaluée, puis une analyse spatiale sur ces zones qui détermine les objets géographiques que ces zones composent, et élucide leur appartenance fonctionnelle.

Traduire la mesure satellitaire en zones géographiques

On peut considérer, en s'inspirant de T. Bittner et S. Winter dans [BW99], que la première phase met en jeu une fonction de correspondance $\phi_{localisation}$, qui associe à une région de la scène, et pour une résolution donnée, une zone de la surface terrestre. Cette zone géographique peut alors être considérée comme un terrain dans la mesure où elle est homogène du point de vue des paramètres de mesure considérés.

Considérons en effet la zone terrestre représentée en bleu sur la figure 4, empruntée à [BW99].

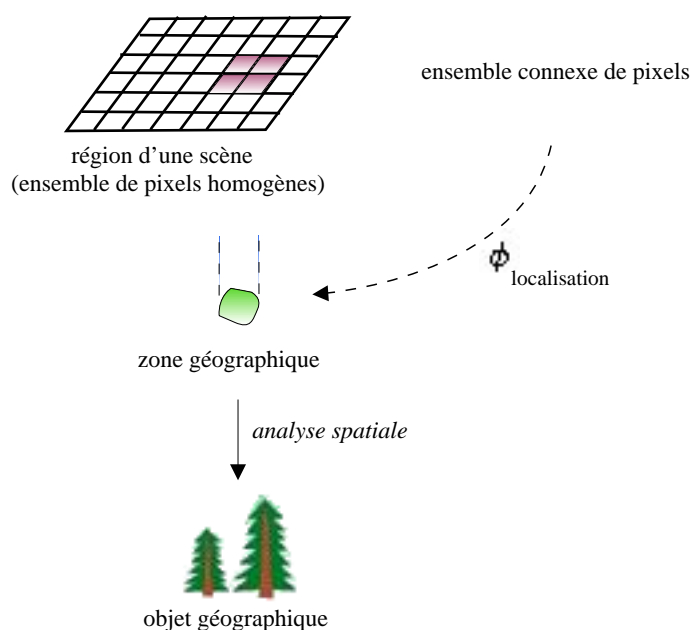


FIG. 3 – Le processus de photo-interprétation des images satellites.

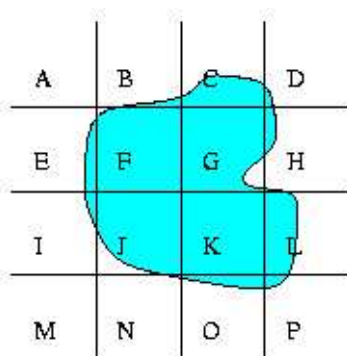


FIG. 4 – Localisation à partir des images raster (d’après [BW99]).

Sur cette figure, le quadrillage matérialise les zones de mesure des différents pixels du capteur. Le pixel K est ici le seul qui réalise des mesures sur une zone de la surface terrestre totalement incluse dans la zone bleue. La mesure effectuée par ce pixel renverra donc des valeurs correspondant aux caractéristiques physiques de la zone bleue (moyennant bien sûr la variation due aux conditions météorologiques et à la marge d’erreur inhérente au matériel de mesure). Par contre, les mesures effectuées sur tous les autres pixels ne traduiront que partiellement ces caractéristiques, et les valeurs obtenues ne seront qu’une moyenne des réponses de la zone bleue et des zones environnantes. Il est donc délicat d’évaluer la localisation précise de la zone bleue à partir des seules réponses des différents pixels, d’autant qu’il a été observé une variabilité des réponses spectrales d’une même zone terrestre suivant la nature des zones qui l’environnent (un même carré d’herbe renvoie des mesures différentes selon qu’il est au milieu d’un pré ou qu’il est une clairière au milieu d’un bois).

Reconstituer les objets géographiques

A une zone géographique ne correspond pas nécessairement un objet géographique. Une zone

géographique détectée peut correspondre soit à une partie d'un objet, soit à un ou plusieurs objets géographiques d'un des trois types cités ci-dessus. En effet :

- Il se peut que la réponse spectrale d'un même objet microphysique varie fortement au sein même de la zone occupée par cet objet, entraînant la segmentation de sa réponse sur la scène en deux ou plusieurs régions. C'est le cas des zones d'ombres, qui apparaissent pour certaines caractéristiques de prise de vue comme l'heure ou l'angle de visée [Pui03]. Ces zones doivent alors être fusionnées.
- Plusieurs objets géographiques contigus et de nature très différentes peuvent également dans certains cas constituer une seule et même zone géographique. C'est le cas par exemple des sols très humides et des bâtis, qui ont une signature spectrale similaire, ou des champs cultivés qui à certaines périodes de l'année apparaissent tous, quel que soit le végétal cultivé, comme des sols nus.

Une fois que l'expert a localisé différentes zones représentées sur la scène, il utilise sa connaissance du milieu urbain pour reconstituer les différents objets géographiques à partir de ces zones. Par exemple, s'il obtient une succession d'alignements de sol nu et de végétation à une résolution de moins d'un mètre, il pourra déduire la présence d'une vigne.

Cette étape est une étape d'*analyse spatiale*, qui met en jeu des connaissances de **morphologie urbaine**.

6.4 Variabilité des concepts

Une des notions clés dans la conception d'une ontologie géographique est celui d'*identité* [Har02]. En effet, suivant le contexte, un même concept géographique peut être identifié sous différentes formes. Pour construire une ontologie d'objets urbains, nous aurons donc à prendre en compte une grande variabilité des concepts en fonction de la résolution, du lieu et de la date d'observation.

Variabilité avec la résolution

La définition des objets géographiques varie avec la résolution. Par exemple l'objet macro Parc-JardinSquare peut être détecté à une résolution de 30 mètres sur une zone de signature spectrale uniforme. A cette résolution les objets méso (groupements d'arbres, alignement d'arbres) et micro (étang, terrain de boule) qui le composent ne pourront pas être détectés. Ceux-ci au contraire, pourront être identifiés à une résolution de l'ordre du mètre, où leurs signatures spectrales seront très différentes.

A. Puissant et C. Weber dans [PW02] introduisent donc pour un même objet les notions de résolution *minimale* et *fonctionnelle*. La résolution minimale est le seuil de résolution permettant d'identifier globalement la forme et la bordure de l'objet, tandis que la résolution fonctionnelle définit le seuil de résolution à partir duquel il est possible d'analyser l'objet, c'est à dire d'en déterminer les différents composants.

Variabilité spatiale

La nature des objets géographiques varie aussi considérablement avec la localisation. En effet, la forme de la ville et de ses espaces verts est modelée par de très nombreux facteurs, comme l'activité humaine passée et présente, la culture, le climat, le relief. On s'attendra à trouver des vergers d'oliviers ou d'orangers en pays méditerranéen, des cités ouvrières dans les bassins industriels plutôt qu'en haute ou moyenne montagne, un sol sec et peu ou pas d'autoroutes au Mali, etc. Si les objets micro restent quasiment inchangés lorsque d'un lieu à l'autre, on n'y retrouve pas les mêmes objets méso et macro. Et lorsqu'on les retrouve leur forme est souvent très différente.

Variabilité temporelle

De la même façon, l'aspect des objets géographiques observés peut varier beaucoup dans le temps. C'est le cas des cultures agricoles au cours de l'année, qui passent suivant la saison du sol nu à un espace de végétation. De même la météo peut avoir un fort impact sur la réponse captée par le satellite, en modifiant en particulier l'humidité du sol. A une échelle de temps plus longue, la modification des pratiques humaines fait aussi évoluer le type d'objets présents, leur nombre et leur forme.

6.5 Le modèle conceptuel proposé

En tenant compte des remarques faites plus haut ont été créés à la racine de l'ontologie (figure 5) :

- un concept *ZoneGéographique*, qui représente toutes les zones terrestres qu'on peut associer à une région de la scène. A chacune de ces zones est associée un jeu de paramètres de mesures du capteur et un attribut *resolution*, qui représente la plage de résolution à laquelle est identifiée la zone en question. Les paramètres de mesure dont on dispose pour une région de la scène sont représentés par des attributs, qui à une instance du concept associe une valeur réelle. On a par exemple les valeurs pour chaque bande spectrale B0 à B4, l'indice de Morton I_H , l'indice de végétation. A chaque zone géographique détectée est également associé un lien *localisation* vers une instance du concept *Localisation* et une relation *contient* vers des instances d'objets géographiques¹⁹.
- un concept *Localisation*, qui représente la délimitation à la surface de la Terre d'une zone géographique. Il est le résultat de la fonction de correspondance $\phi_{localisation}$ appliquée à la région de la scène (figure 3).
- trois concepts *ObjetMicro*, *ObjetMéso* et *ObjetMacro*, qui représentent respectivement les objets micro, méso et macro qui ont été reconnus sur cette zone géographique.

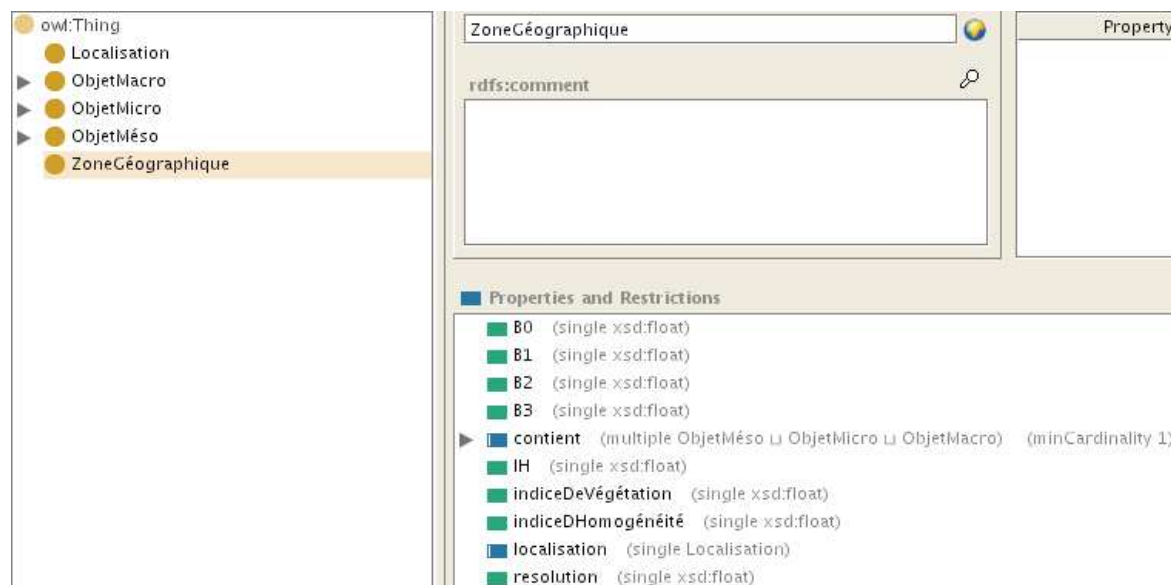


Fig. 5 – La hiérarchie générale de l'ontologie, et les propriétés de la classe *ZoneGéographique* (copie d'écran depuis le plugin OWL de l'éditeur Protégé).

¹⁹Pour retrouver la ou les zone(s) qu'occupe un objet on définit même un lien bidirectionnel entre zones et objets en introduisant avec la relation *contient* une relation *faitPartieDe*, qu'on définit comme la relation inverse de *contient* (i.e. a faitPartieDe b ssi b contient a).

Les objets méso et macro étant définis comme des regroupement d'objets, nous avons également introduit une relation *estComposéDe*, qui relie un objet aux objets qui le composent. Ainsi la définition du concept *ParcJardinSquare* (figure 6) stipulera qu'il est composé d'au moins un objet micro, et que les objets qui le composent sont de types *Eau*, *Bâtiment*, *SolNu* ou *Végétation*.

The image shows a software interface for defining an ontology concept. On the left, a tree view shows the hierarchy of classes under `owl:Thing`. The classes are: `Localisation`, `ObjetMacro`, `ObjetMicro`, `ObjetMéso`, `AlignementD'Arbres`, `Camping`, `GroupeD'Arbres`, `JardinDeL'HabitatDiscontinu`, `JardinOuvrierOuFamilial`, `ParcDeLoisirsOuAttractions`, `ParcJardinSquare` (highlighted), `TerrainDeSportArtificialisé`, `TerrainDeSportNaturel`, and `ZoneGéographique`. On the right, the definition for `ParcJardinSquare` is shown. It includes an `rdfs:comment` field and a section titled "Properties and Restrictions". The restrictions are: `compose` (allValuesFrom `ObjetMacro`), `estComposéDe` (allValuesFrom `Eau ∪ Bâtiment ∪ SolNu ∪ Végétation`), `densitéDeSurfacesArtificialisées` (single `xsd:float`), `densitéDeVégétation` (single `xsd:float`), `faitPartieDe` (multiple `ZoneGéographique`), and `localisation` (single `Localisation`).

Fig. 6 – La définition du concept *ParcJardinSquare* dans l'ontologie.

7 Ontologisation

7.1 Le langage retenu : OWL

Le langage OWL²⁰ est une recommandation du W3C depuis le 10 février 2004. Il a été conçu pour intégrer les ontologies avec les standards du web, et bénéficie largement des travaux réalisés dans les domaines de la représentation des connaissances par objets et des logiques de descriptions. Ceci est la meilleure garantie d'**interopérabilité** de la base de connaissances de FoDoMuST avec d'autres bases de connaissances existantes via internet.

7.2 Un langage riche

OWL s'appuie sur le langage de description de ressources RDF/S

RDF²¹ est le langage d'annotation du web. Il permet d'affirmer des relations entre les ressources, et représente l'information sous forme de triplets < sujet, prédicat, objet >, dont les éléments peuvent être des URIs, des variables ou des littéraux. Les triplets RDF sont généralement représentés par des multi-graphes étiquetés orientés (voir l'exemple de la figure 11 donné en annexe, extraite du document de recommandation du W3C).

OWL reprend donc la capacité d'annotation de RDF, mais aussi la possibilité offerte par son extension RDF Schema²² (RDFS) de définir des hiérarchies de classes et de propriétés, ainsi que de spécifier les types de classes auxquels doivent appartenir les individus source et cible des propriétés (rdfs:domain et rdfs:range).

Il y ajoute notamment la possibilité de déclarer qu'une propriété est transitive, symétrique, fonctionnelle ou encore qu'elle est l'inverse d'une autre propriété. Mais le plus grand apport d'OWL par rapport à RDFS est sans doute de pouvoir définir des restrictions locales sur les propriétés [HHPs03].

La liste des constructeurs du langage OWL utiles pour le projet FoDoMuST est donnée en annexe (figure 12).

OWL reprend l'approche par cadres

Toute classe ou propriété définie dans OWL peut également être vue comme un cadre. Les slots correspondent alors aux différentes sous-propriétés et les facettes aux restrictions déclarées sur chacune d'elles (le type de données dans lequel un attribut prend ses valeurs par exemple).

Une approche à base de cadres est utile dans FoDoMuST dans la mesure où les objets et zones géographiques sont définis par la donnée d'un certain nombre d'attributs (comme des indices spectraux ou de forme) et de domaines dans lesquels ils peuvent prendre leurs valeurs. Il est alors intéressant de regrouper en un cadre l'information concernant chaque classe.

De plus les cadres peuvent être une structure intéressante pour traduire les différentes spécialisations d'un concept. En effet comme nous l'avons vu au paragraphe 6.4, les composants d'un objet ne sont détectés qu'à partir du seuil de résolution fonctionnelle, et la définition de l'objet varie beaucoup suivant la résolution d'observation. Cette réalité peut être traduite par la déclinaison de chaque concept en plusieurs sous-concepts contenant chacun la définition de l'objet pour une plage de résolution donnée. Cette solution, quoique peu optimale car inapte à traduire l'effet linéaire de la résolution sur la variation de la définition des concepts, augmenterait néanmoins considérablement l'intelligibilité de l'ontologie.

²⁰Web Ontology Language. <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>

²¹<http://www.w3.org/RDF/>

²²<http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>

OWL s'appuie sur les logiques de descriptions

Les recherches sur les logiques de descriptions ont beaucoup influencé la conception d'OWL. Parmi les trois sous-langages d'expressivité croissante qu'il offre, deux d'entre eux (OWL Lite et OWL DL) peuvent même être vus comme des logiques de descriptions très expressives. Il tire en particulier des logiques de description sa sémantique, l'intégration de types de données et de valeurs, ainsi que des constructeurs permettant par exemple de définir des concepts par l'union ou l'intersection d'autres concepts.

7.3 Un mécanisme d'inférence efficace

Les capacités de raisonnement d'OWL viennent des logiques de description sur lesquelles il est fondé. Parmi ces langages nous nous intéresserons à OWL DL qui est l'espèce la plus riche sémantiquement mais tout en préservant la propriété de décidabilité. Une ontologie OWL DL correspond à une terminologie (TBox) de la logique *SHOIN(D)*. Celle-ci est fondée sur la logique de description *ALCQHI_{R+}* (aussi appelée *SHIQ* [HST00]), qu'elle étend pour permettre de définir une classe par l'énumération de ses instances et par l'introduction de types de données comme les entiers, les réels et les chaînes de caractères.

Le langage SHIQ

SHIQ étend la logique *ALC*, qui permet de définir un concept par l'intersection, l'union et la négation d'autres concepts et de restreindre le domaine d'application des rôles (application de quantifieurs universels et existentiels). Il est basé sur la logique *S*, qui y inclut des rôles transitifs (*R₊*). Il supporte de plus la définition de rôles inverses (*I*), de hiérarchies de rôles (*H*), et permet de restreindre la cardinalité des rôles (*Q*) pour chacun de leurs domaines d'application. Néanmoins, ces restrictions de cardinalité ne sont autorisées que pour des rôles dits simples, c'est à dire qui ne sont pas transitifs et dont aucun sous-rôle n'est transitif. Un résumé de ces constructeurs est donné en annexe dans le tableau de la figure 7.

syntaxe	sémantique
concepts C et D, rôles R, rôles simples S domaine d'interprétation Δ^I entier $n \in \mathbb{N}$	
\top	$\top^I = \Delta^I$
\perp	$\perp^I = \emptyset$
C	$C^I \subseteq \Delta^I$
$\neg C$	$\neg C^I = \Delta^I \setminus C^I$
$C \sqcap D$	$(C \sqcap D)^I = C^I \cap D^I$
$C \sqcup D$	$(C \sqcup D)^I = C^I \cup D^I$
R^-	$\{(a,b) \in \Delta_I \times \Delta_I / (b,a) \in R^I\}$
$\exists R.C$	$\{a \in \Delta^I / \exists b \in \Delta^I : (a,b) \in R^I, b \in C^I\}$
$\forall R.C$	$\{a \in \Delta^I / \forall b \in \Delta^I : (a,b) \in R^I \implies b \in C^I\}$
$\geq n \text{ S.C}$	$\{a \in \Delta^I / \parallel b \in C^I \text{ et } (a,b) \in S^I \parallel \geq n\}$
$\leq m \text{ S.C}$	$\{a \in \Delta^I / \parallel b \in C^I \text{ et } (a,b) \in S^I \parallel \leq m\}$

FIG. 7 – Syntaxe et sémantique de *SHIQ*

Enumerations et types de données

SHOIN(D) permet de plus de définir un concept par l'énumération de ses instances (*O*). Un concept *Couleur* peut ainsi être défini par trois instances possibles : Bleu, Rouge ou Vert. La déclaration de types de données est aussi rendue possible (*D*). Les types de données supportées sont les types de données XMLSchema, comme *xsd:decimal* ou *xsd:string*, ainsi que

rdfs:Literal.

Pour des raisons de décidabilité, l'interprétation des types de données a été séparée de l'interprétation des autres classes [HHPs03]. Ceci a été réalisé par l'introduction d'un domaine d'interprétation Δ_D^I , disjoint de Δ^I , qui représente le domaine des valeurs de données. Un type de données est alors interprété comme un sous-ensemble de Δ_D^I . La séparation entre concepts et types de données est accentuée par la distinction dans le langage entre deux types de propriétés, *ObjectProperty* et *DatatypeProperty*, suivant qu'elles relient un individu à un autre individu (de Δ^I) ou à une donnée (de Δ_D^I).

7.4 Des limitations concernant les domaines de définition des attributs

La principale limitation d'OWL DL réside en son incapacité à décrire les domaines de définition des attributs. En effet, aucun de ses constructeurs ne permet de définir un concept par un intervalle réel. En conséquence, il est impossible de définir un concept comme le concept C suivant :

$$C = \text{"contient au moins 2 bâtis d'indice de Morton } > 0.4 \text{ et densitéDeVégétation } < 0.2\text{"} \quad (1)$$

Nous verrons au paragraphe 9.2 qu'une extension d'OWL est à l'étude pour pallier à ces lacunes.

8 Opérationnalisation

8.1 Utilisation d'un raisonneur

Une fois le modèle conceptuel retenu exprimé dans un langage adéquat, il reste à mettre en place un raisonneur (figure 8). Celui-ci fournit les mécanismes de raisonnement sur l'ontologie utiles au processus de reconnaissance des objets urbains en exploitant les mécanismes d'inférence autorisés par le langage. Les principaux mécanismes dont on a besoin dans le cadre de FoDoMuST sont les suivants :

Classification de concepts

La recherche du ou des concepts de l'ontologie dont la définition correspond à une description donnée s'effectuera en exploitant le mécanisme de classification de concepts. On introduira pour ce faire le concept correspondant à chaque description de zone identifiée, puis on le classifera dans la hiérarchie de concepts de l'ontologie définissant les objets et on renverra les plus petits subsumants dans la hiérarchie.

Vérification de la cohérence des instances

Réaliser un appariement entre objets détectés à des résolutions différentes est alors un bon moyen de valider des hypothèses de reconnaissance. On peut utiliser le test de cohérence sur les instances pour vérifier la cohérence entre un objet reconnu sur une zone géographique à basse résolution et ceux que l'on a reconnus sur la même zone géographique à une résolution plus fine. Par exemple, figure 6 section 6.5, si l'on reconnaît un parc à une résolution de 30 mètres, on vérifiera qu'à une résolution supérieure à sa résolution fonctionnelle n'apparaissent sur la même zone que des objets de type *Eau*, *Bâtiment*, *SolNu* ou *Végétation*. Une reconnaissance d'objet *Tronçon* sur cette zone entraînerait une incohérence. De plus, confronter l'expression d'un même objet sur plusieurs scènes permet également d'évaluer plus précisément son contour.

8.2 Déroulement du processus de reconnaissance

La reconnaissance s'effectuera par un peuplement progressif de la base de connaissances. Seront d'abord instanciées des zones géographiques, qui seront le résultat de l'application de la fonction

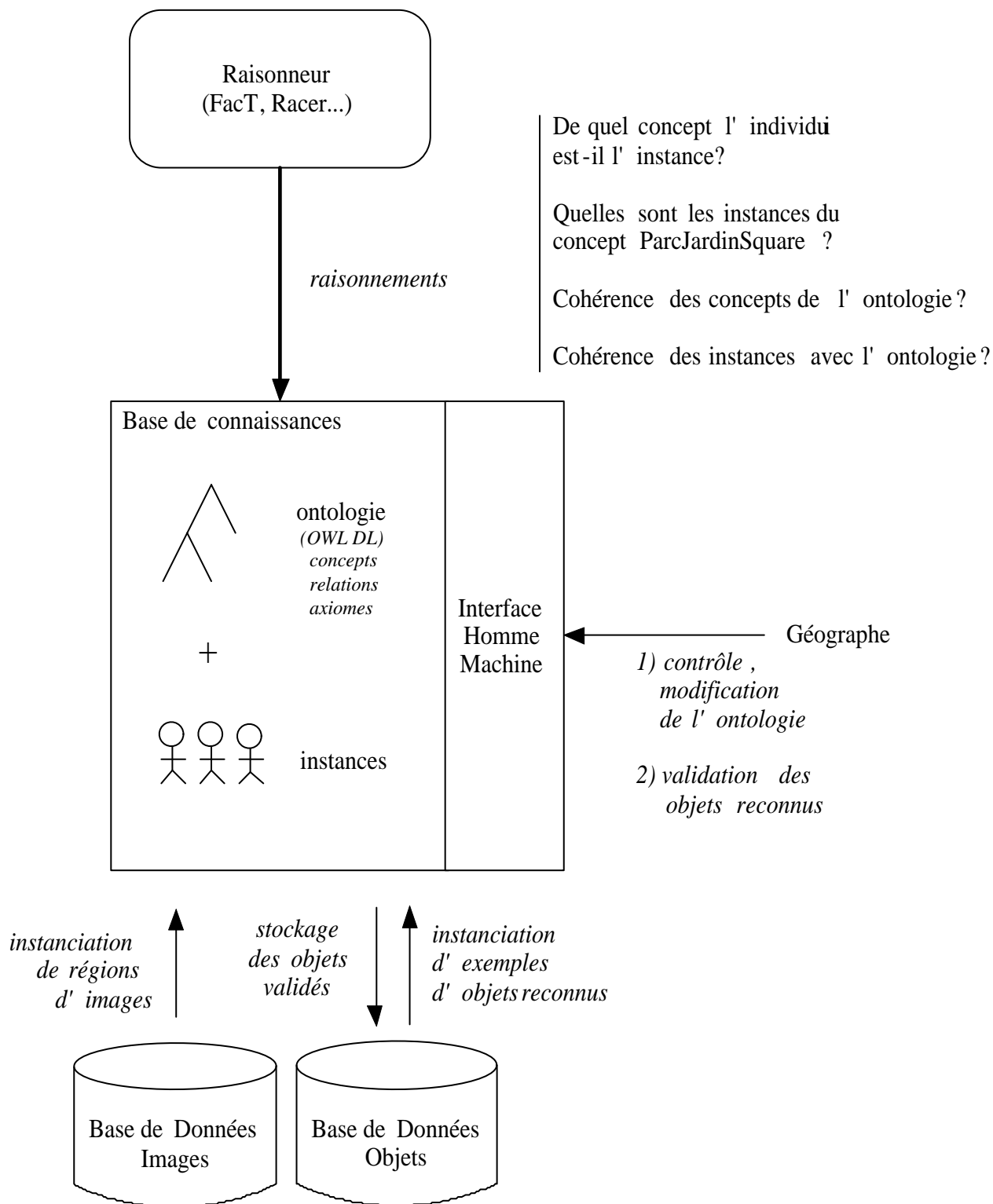


FIG. 8 – Opérationnalisation de l'ontologie.

$\phi_{localisation}$ sur les différentes régions de la scène. Puis sera appliquée une **stratégie de reconnaissance des objets géographiques**, qui aboutira à l’instanciation dans l’ontologie des objets urbains.

Un modèle spatial du lieu géographique considéré pourra alors être constitué. Il représentera à chaque instant les objets dont l’étiquetage a été validé par l’expert. La validation se fera sur la base de propositions, formulées à partir de la classification des descriptions des zones dans l’ontologie. Le modèle spatial sera ainsi affiné au fur et à mesure (et la base de connaissances révisée) soit par l’acquisition de nouvelles images pour le même lieu géographique, soit par l’intervention directe de l’expert. Une fois le modèle spatial entièrement construit, c’est à dire une fois que tous les objets géographiques présents sur le lieu géographique considéré ont été reconnus, les instances des zones géographiques n’auront en principe plus d’utilité.

Les objets identifiés ont vocation à être stockés, une fois le processus de reconnaissance achevé, dans une base de données géographique pour être utilisés par un système d’information géographique. Néanmoins pour être utilisables dans des applications comme l’urbanisme ou la gestion des risques leur description devra être enrichie par des connaissances plus larges que celles servant strictement à leur identification sur les images.

9 Implémentation

L’implémentation de l’ontologie a été effectuée sous l’éditeur Protégé²³. Dans Protégé une ontologie est composée de classes, de slots, de facettes et d’axiomes, et forme avec les instances correspondantes aux classes une base de connaissances [MNF02]. Nous avons alors utilisé le plugin OWL de Protégé, qui fournit une interface graphique permettant de développer directement du code OWL. Dans la mesure où l’ontologie est destinée à être parcourue, nous avons ensuite étudié les différentes solutions techniques qui nous permettraient d’exploiter les mécanismes d’inférence associés à OWL pour raisonner sur la base de connaissances.

9.1 Choix d’un raisonneur : Racer

Racer²⁴ [MH01] est un serveur de logiques de descriptions développé à l’université de Hamburg par Ralph Möller et Volker Haarslev, qui fournit de nombreuses possibilités de raisonnement tant sur les concepts (TBox) que sur les instances (ABox) d’une base de connaissances. Son point fort est sa capacité à raisonner sur des domaines de définition d’attributs entiers, chaînes de caractères, mais aussi réels voire même complexes. Il implémente la logique $\mathcal{ALCQHI}_{\mathcal{R}^+}(\mathcal{D})^-$, ou $\mathcal{SHIQ}(\mathcal{D})^-$.

La logique $\mathcal{SHIQ}(\mathcal{D})^-$

La logique $\mathcal{SHIQ}(\mathcal{D})^-$, étend le langage \mathcal{SHIQ} par l’ajout de définitions de *domaines concrets*. Un domaine concret \mathcal{D} est un couple $(\Delta_{\mathcal{D}}, \Phi_{\mathcal{D}})$ composé d’un domaine d’interprétation $\Delta_{\mathcal{D}}$ et d’un ensemble de prédicats n-aires $\Phi_{\mathcal{D}}$. Un tel prédicat est par exemple le prédicat unaire \leq_{10} sur les entiers. Racer raisonne en particulier sur les domaines concrets définissant des restrictions (min/max) sur les valeurs entières et des équations et inéquations sur les réels.

²³Protégé est un éditeur de systèmes à base de connaissances *open-source* développé au Knowledge Systems Laboratory à l’université de Stanford. Il est compatible avec l’*Open Knowledge-Based Connectivity protocol* (OKBC), qui est une API développée par l’*Open Knowledge Base Connectivity Workgroup* (<http://www.ai.sri.com/~okbc/>) pour assurer l’interopérabilité entre systèmes de représentation des connaissances. Le modèle de connaissances de l’OKBC définit le cadre comme objet primitif représentant une entité du domaine du discours [CFF⁺98].

²⁴*Renamed Abox and Concept Expression Reasoner*

Donc pour peu qu'on parvienne à contourner les limitations d'OWL DL concernant la déclaration de domaines de définition des attributs, on pourra utiliser Racer pour effectuer les raisonnements requis, y compris sur les données et les types de données.

Précautions

Néanmoins, il faudra prendre soin pour utiliser Racer de ne pas définir de concept OWL par énumération de ses instances, étant donné qu'un raisonnement sur un tel concept n'est pas supporté (cette contrainte reste néanmoins mineure pour notre usage).

De plus, notre besoin de déclarer des intervalles de réels proscrit d'attaquer Racer via son interface DIG²⁵, car dans sa version actuelle (1.1, décrite dans [Bec03]), ce format de fichier ne les supporte pas. L'ontologie devra donc lui être fournie en entrée par son interface fichier, et donc convertie dans son format natif, dont la syntaxe est très proche de la syntaxe KRSS²⁶.

9.2 Solutions techniques pour la déclaration de domaines de définition des attributs

Plusieurs solutions techniques peuvent être envisagées pour contourner les limitations d'OWL DL et définir quand même les concepts de l'ontologie par des domaines de définition d'attributs.

L'une d'elles consiste à conserver le langage OWL DL et à rajouter les déclarations manquantes manuellement. Le plus simple serait alors de les spécifier en commentaire, dans la balise `rdf:comment` par exemple. Afin d'être utilisées par Racer, ces déclarations devront être parsées, puis incorporées à la traduction du code OWL vers un format Racer natif. Cette solution est assez lourde et peu praticable si l'ontologie est composée en grande partie de ces définitions de domaines de définition d'attributs (ce qui est malheureusement le cas).

Une autre solution serait d'écrire directement l'ontologie en format Racer natif. Malheureusement celle-ci deviendrait vite illisible, même pour l'ingénieur de la connaissance, et on perdrait la clarté apportée par l'approche à base de cadres que suit OWL. Il ne serait en outre plus possible d'utiliser un éditeur d'ontologie comme Protégé. Le développement d'une interface graphique serait nécessaire.

Une troisième solution intermédiaire est d'utiliser le plugin RDF de Protégé. Celui-ci permet de développer l'ontologie sous Protégé en conservant l'approche par cadres. Il génère un code RDFS augmenté de balises spécifiques à Protégé qui remplacent une partie des constructeurs OWL manquants, ainsi que les domaines de définition des attributs. Resterait à convertir ce format hybride en un format Racer natif, ce qui peut être raisonnablement envisagé en implémentant par exemple une feuille de style XSLT et un parseur XML (le format Racer natif est un pseudo-Lisp).

Ces astuces ne sont néanmoins que des solutions temporaires, dans la mesure où les lacunes d'OWL concernant les domaines de définition des attributs seront très vraisemblablement comblées rapidement. Dans sa thèse soutenue en 2004 [Pan04], Jeff Z. Pan propose à ce sujet une extension d'OWL, appelée OWL-E, qui comble ce manque en introduisant de nouveaux constructeurs. Il met en évidence l'existence d'un domaine concret par lequel on puisse étendre OWL tout en conservant sa propriété de décidabilité. En OWL-E, la définition du concept C de l'équation (1) serait :

$$C \doteq \geq 2\text{contient}(\text{indiceDeMorton} \geq_{[0.4]}^{\text{float}}) \sqcap \text{densitéDeVégétation} \leq_{[0.2]}^{\text{float}}$$

Néanmoins pour l'instant aucun raisonneur ni aucun éditeur d'ontologies ne supporte l'extension OWL-E.

²⁵DIG est un format de fichier développé par le *Description Logics Implementation Group* (<http://dl.kr.org/dig/>) dans le but de fournir une interface unifiée à tous les serveurs de logiques de description.

²⁶*Knowledge Representation Systems Specification* [Ps93]. Cette spécification du groupe KRSS de l'ARPA vise à définir les propriétés que doit satisfaire toute implantation d'un système de représentation de connaissances basé sur les logiques de description, et ce notamment en terme de syntaxe supportée en entrée pour la base de connaissances et d'inférences minimales requises sur cette base.

Quatrième partie

Enrichissement de l'ontologie

10 Motivations

10.1 Ajouter des connaissances décisionnelles à l'ontologie

Enrichir l'ontologie consiste à rajouter de nouvelles connaissances à celles qu'on a déjà. L'enrichissement peut prendre deux formes :

Une révision de l'ontologie

On peut vouloir modifier directement l'ontologie, pour préciser la définition de certains concepts, de certaines relations, ajouter/supprimer certains concepts ou certaines relations, bref modifier partiellement le modèle conceptuel sous-jacent. Toute intervention sur l'ontologie doit alors être avalisée par l'expert.

Les nouvelles connaissances peuvent venir d'un avis de l'expert, être extraites de documents textuels (documents d'urbanisme par exemple) ou provenir de l'analyse des objets urbains déjà étiquetés.

L'adjonction de connaissances décisionnelles

On peut également vouloir extraire des objets urbains déjà étiquetés des connaissances décisionnelles. Celles-ci viendront en complément des connaissances de l'ontologie, et permettront d'améliorer le processus de reconnaissance.

Dans le cadre de FoDoMuST, nous nous sommes intéressés à la production de connaissances décisionnelles. Nous sommes partis du constat énoncé dans [Pui03] selon lequel les critères de regroupement donnés par le géographe pour reconstituer les objets méso à partir des objets micro ne permettent pas à eux seuls d'identifier tous les objets méso demandés par les utilisateurs. Ceci se traduit dans l'ontologie par des définitions d'objets méso qui ne sont pas assez précises pour assigner de manière univoque une étiquette à chaque regroupement d'objets micro rencontré. L'extraction de connaissances visera donc à préciser localement ces définitions, en apprenant des connaissances décisionnelles à partir des objets méso déjà reconnus.

En effet, comme nous l'avons vu en 6.4, les critères de regroupement doivent être redéfinis pour chaque lieu géographique étudié du fait de la grande variabilité spatiale des objets géographiques. Comme il semble néanmoins raisonnable de supposer que pour un lieu géographique donné (l'agglomération strasbourgeoise par exemple) les différents exemples d'un même objet urbain sont de constitutions relativement homogènes (en termes de matériaux utilisés, d'agencement spatial, d'environnement, de taille, de forme ou de composition), nous proposons d'extraire les caractéristiques communes aux descriptions de ces exemples et de s'en servir pour affiner localement la définition des concepts.

La mise au point d'une méthode d'apprentissage automatique sur ces descriptions supposera bien sûr que ces exemples d'apprentissage sont en nombre suffisant, c'est à dire que l'expert a déjà validé sur le lieu géographique considéré un nombre suffisant d'objets méso.

10.2 Quelles sont les données d'apprentissage ?

Les exemples d'apprentissage sont des descriptions d'objets méso déjà reconnus, et auxquels le géographe a assigné une étiquette qui correspond à un concept de l'ontologie. Ce qui caractérise ces descriptions c'est essentiellement d'être des objets structurés et de pouvoir être instances simultanément de plusieurs concepts de l'ontologie.

Des objets structurés

Un objet *structuré* [Bou96] est un objet constitué de plusieurs objets ayant des relations entre eux. Ce qui fait la spécificité des objets structurés, c'est de pouvoir, d'un exemple à l'autre, être constitué d'un nombre variable d'objets, et varier dans sa structure relationnelle [Bis92].

Les descriptions d'objets méso que nous utiliserons comme exemples d'apprentissage sont des objets structurés comme celui de la figure 9 suivante :

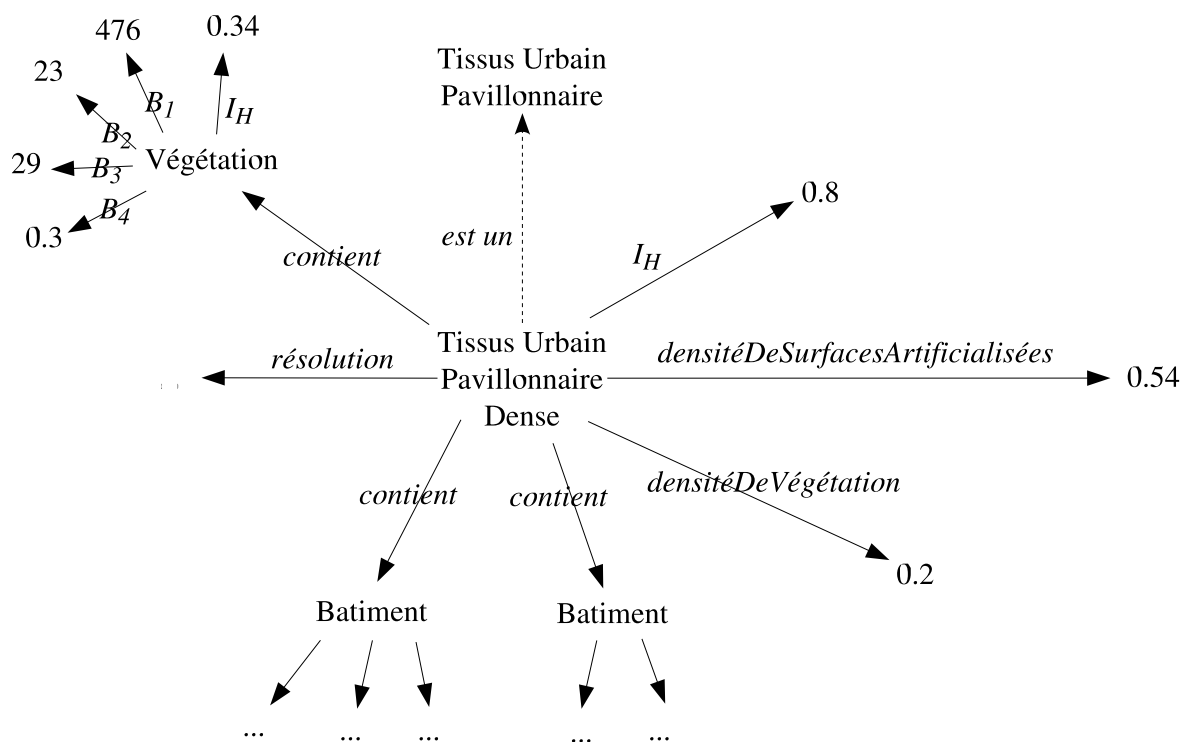


FIG. 9 – Exemple de Description d'un Objet Méso étiqueté Tissu Urbain Pavillonnaire Dense

Chaque exemple peut être instance de plusieurs concepts

Il est important de remarquer ici que si chaque exemple est bien, de par sa description, une instance du concept par lequel il a été étiqueté, il peut très bien être également instance d'un nombre quelconque d'autres concepts de l'ontologie. Ainsi un exemple étiqueté *EmpriseHospitalière* peut avoir une description également conforme à la définition du concept *EmpriseUniversitaire*.

Pour illustrer cette idée, nous avons représenté sur la figure 10 le domaine du discours Δ^I (i.e. toutes les descriptions susceptibles d'être rencontrées) et les exemples de ce domaine qui ont déjà été étiquetés par les concepts C ou D. Parmi ces exemples, nous avons colorié en rouge

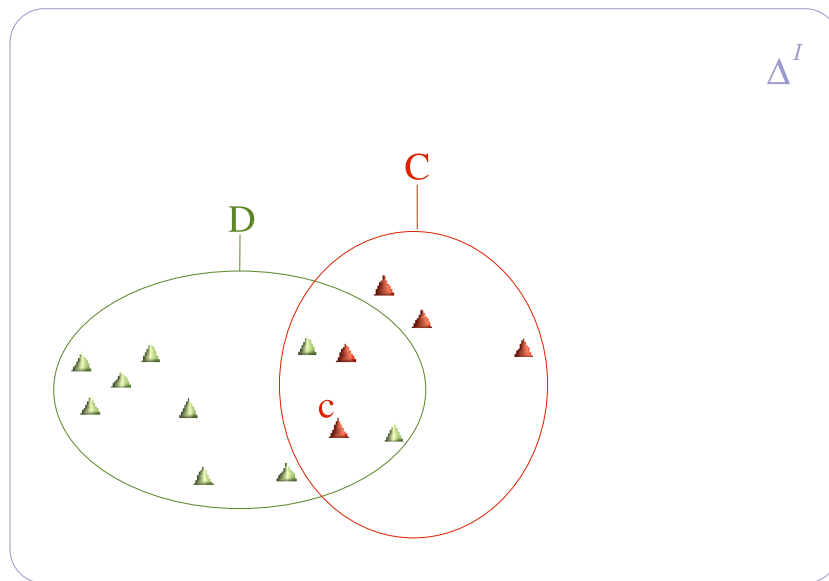


FIG. 10 – L'interprétation de deux concepts C et D dans le domaine du discours

les exemples qui ont été étiquetés par le concept C et en vert ceux étiquetés par D. On voit bien que l'individu c , s'il est rouge car étiqueté par le concept C, a une description qui correspond simultanément aux deux concepts C et D, et est par conséquent instance dans la base de connaissances de ces deux concepts.

10.3 Quelle stratégie adopter ?

Plusieurs solutions s'offrent à nous pour découvrir des connaissances décisionnelles à partir des objets méso déjà étiquetés.

On peut envisager de classifier ces objets par une méthode non-supervisée. Les classes pourraient être déterminées par une méthode de réallocation (inspirée des k-moyennes par exemple) ou une méthode ascendante hiérarchique (CAH). Mais alors il faut s'assurer que les définitions des classes obtenues soient bien des sous-concepts des classes de l'ontologie, sinon les deux sources de connaissances (définitions dans l'ontologie et connaissances décisionnelles) entreraient en contradiction.

Une autre approche serait de réaliser une classification supervisée sur ces objets, en remarquant qu'une ontologie instanciée est déjà une classification. On ferait alors pour chaque objet méso de l'apprentissage de concept à partir des exemples positifs et négatifs que l'on a. On s'inspirerait des méthodes de programmation logique inductive ou d'arbres de décision pour réapprendre pour chaque objet méso une définition qui 'colle' aux exemples déjà étiquetés. Mais alors il faut prendre en compte le fait que nous recherchons moins une définition précise de chaque concept d'objet méso que des éléments **discriminants**, c'est à dire des critères qui permettent de décider pour toute nouvelle composition d'objets méso par quel objet méso il faut l'étiqueter.

En somme soit on applique une méthode discriminante et on risque de perdre la définition de l'objet méso, soit on "réapprend" une définition plus précise de chaque objet méso, et alors on risque de perdre l'objectif d'aide à la décision.

Conclusion et perspectives

Nous avons montré au cours de cette étude l'intérêt d'utiliser une ontologie d'objets urbains dans FoDoMuST pour automatiser au moins partiellement le processus de photointerprétation. Celle-ci intervient à la fois comme formalisme de représentation des connaissances et comme outil informatique pour effectuer des raisonnements. Sa conception se nourrit d'un dialogue constant avec l'expert qui aboutit à la création d'un modèle conceptuel. Ce modèle doit alors être formalisé, et nous avons montré que le choix du langage dans lequel le formaliser est déterminant car il traduit un parti-pris de conceptualisation et conditionne les inférences qui pourront être effectuées sur l'ontologie.

L'étude de la littérature a révélé un grand nombre de travaux théoriques sur les ontologies géographiques et la représentation spatiale, mais la quasi-inexistence de réalisations pratiques d'ontologies géographiques réellement opérationnelles. Seules quelques équipes ont tenté d'intégrer une ontologie à un système d'information géographique mais ces expériences se sont en pratique réduites à l'adaptation d'un schéma de bases de données.

Nous avons néanmoins entrepris de dégager un modèle conceptuel et de le formaliser en optant, à la suite d'une analyse des besoins exprimés au sein du projet, pour le langage OWL. Nous avons alors proposé une implémentation de l'ontologie, qui a été réalisée sous l'éditeur Protégé, et montré comment un raisonneur comme Racer pouvait exploiter les mécanismes d'inférences du langage OWL pour effectuer sur l'ontologie les tâches de reconnaissance d'objets urbains. Ont ensuite été discutés les limites des outils informatiques existants en vue de réaliser son opérationnalisation et les solutions à envisager pour y remédier. Enfin, dans une dernière partie, nous avons ouvert des perspectives sur l'enrichissement de cette ontologie par des processus d'extraction de connaissances à partir de données.

Il ressort de cette étude que le développement d'une ontologie est un processus très long du fait de la première étape de conceptualisation qui consiste à recueillir et formaliser des connaissances de l'expert. Si cette étape a donné lieu, au terme de nombreuses rencontres avec des spécialistes de géographie urbaine et de télédétection, à la proposition d'un modèle conceptuel de l'ontologie, elle reste néanmoins toujours en cours car celui-ci nécessite d'être affiné.

De plus, l'absence d'interface homme-machine rend l'ontologie peu accessible par l'expert, ce qui freine beaucoup les échanges et limite d'autant l'avancement de la conceptualisation. Faute d'interface satisfaisante, la communication s'est faite via l'éditeur Protégé ou un langage graphique proche d'UML.

Enfin, nous avons été confrontés à la très grande complexité des objets à représenter. En effet la nature de ces objets varie aussi bien dans l'espace que dans le temps, et la réponse captée pour ces objets par le satellite varie considérablement — et de manière non linéaire — en fonction de la résolution de prise de vue. La prise en compte de cette complexité reste un véritable défi à relever dans l'optique d'intégrer à un SIG une ontologie réellement opérationnelle.

Cinquième partie

Annexes

Références

- [BCM⁺03] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.
- [Bec03] S. Bechhofer. The DIG description logic interface : DIG/1.1. Technical report, University of Manchester, February 2003.
- [Ben01] B. Bennett. What is a forest ? on the vagueness of certain geographic concepts. *Topoi*, 20(2) :189–201, 2001.
- [Bis92] G. Bisson. Learning in FOL with a similarity measure. In *AAAI92*. MK, 1992.
- [BM00] P. Balbiani and P. Muller. Le raisonnement spatial. In *Le temps, l'espace et l'évolutif en sciences du traitement de l'information*. Cepadues Editions, 2000.
- [Bou96] I. Bournaud. Regroupement conceptuel pour l'organisation de connaissances. Thèse de Doctorat, Spécialité Informatique Université Paris VI, 1996.
- [BVEL04] Sara Brockmans, Raphael Volz, Andreas Eberhart, and Peter Löffler. Visual modeling of owl dl ontologies using uml. In *International Semantic Web Conference*, pages 198–213, 2004.
- [BW99] T. Bittner and S. Winter. On ontology in image analysis. In *ISD '99 : Selected Papers from the International Workshop on Integrated Spatial Databases, Digital Images and GIS*, pages 168–191, London, UK, 1999. Springer-Verlag.
- [CEFM01] G. Camara, M.J. Egenhofer, F. T. Fonseca, and A.M. Vieira Monteiro. What's in an image ? In *COSIT 2001 : Proceedings of the International Conference on Spatial Information Theory*, pages 474–488, London, UK, 2001. Springer-Verlag.
- [CEM75] A. Collins, E.H. Warnock, N. Aiello, and M. L. Miller. Reasoning from incomplete knowledge. In A. Collins and D. G. Bobrow, editors, *Representation and Understanding*, Language Thought and Culture : Advances in the Study of Cognition, chapter 13. Academic Press, New York, 1975.
- [CFF⁺98] V.K. Chaudhri, A. Farquhar, R. Fikes, P.D. Karp, and J.P. Rice. Open knowledge base connectivity 2.0.3, 1998.
- [CPSV03] N. Cullot, C. Parent, S. Spaccapietra, and C. Vangenot. Des sig aux ontologies géographiques. *Revue Internationale de Géomatique*, pages 285–306, 2003.
- [CSF96] G. Câmara, Ricardo C. M. Souza, and Juan Garrido Ubirajara Moura Freitas. Spring : integrating remote sensing and GIS by object-oriented data modelling. *Computers & Graphics*, 20(3) :395–403, May 1996. ISSN 0097-8493.
- [CSV98] R. Casati, B. Smith, and A.C. Varzi. Ontological tools for geographic representation. In N. Guarino, editor, *Formal Ontology in Information Systems*, pages 77–85. IOS Press, Amsterdam, 1998.
- [DSS93] Randall Davis, Howard E Shrobe, and Peter Szlovits. What is knowledge representation ? *AI Magazine*, Spring 1993, 14(1) :17–33, 1993.
- [Duc96] R. Ducournau. Des langages à objets aux logiques terminologiques : les systèmes classificatoires. Rapport de Recherche 96-030, LIRMM, Montpellier, 1996.
ftp ://ftp.lirmm.fr/pub/LIRMM/papers/1996/RRI-Ducournau-96.ps.

- [FDC03] F. T. Fonseca, C. A. Davis, and G. Câmara. Bridging ontologies and conceptual schemas in geographic information integration. *GeoInformatica*, 7(4) :355–378, 2003.
- [FEAC02] F.T. Fonseca, M.J. Egenhofer, P. Agouris, and G. Câmara. Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3), 2002.
- [Für02] F. Fürst. L'ingénierie ontologique. Technical report, Rapport de Recherche, IRIN, Université de Nantes, 2002.
- [Gan91] J.-G. Ganascia. L'hypothèse du knowledge level : Théorie et pratique. *Les Sciences Cognitives en Débat.*, 1991.
- [GDDD04] D. Gasevic, D. Djuric, V. Devedzic, and V. Damjanovi. Converting uml to owl ontologies. In *WWW Alt. '04 : Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 488–489. ACM Press, 2004.
- [GG95] N. Guarino and P. Giaretta. Ontologies and knowledge bases : Towards a terminological clarification. In N. J. I. Mars, editor, *Towards Very Large Knowledge Bases*, pages 25–32. IOS Press, Amsterdam, 1995.
- [GK03] R.L. Goldstone and A. Kersten. Concepts and categorization. In *Comprehensive Handbook of Psychology Volume 4 : Experimental Psychology*, pages 599 – 621. Proctor Eds, 2003.
- [Goo92] M.F. Goodchild. Geographical data modeling. *Comput. Geosci.*, 18(4) :401–408, 1992.
- [Gru93] T. R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands, 1993. Kluwer Academic Publishers.
- [Har02] J. Harding. Geo-ontology concepts and issues. Workshop Report, Ordnance Survey, 2002.
- [HHPs03] Frank Van Harmelen, Ian Horrocks, and Peter F. Patel-schneider. From *SHIQ* and RDF to OWL : The making of a web ontology language, June 18 2003.
- [HST00] I. Horrocks, U. Sattler, and S. Tobies. Reasoning with individuals for the description logic *SHIQ*. In *CADE-17 : Proceedings of the 17th International Conference on Automated Deduction*, pages 482–496, London, UK, 2000. Springer-Verlag.
- [Kui75] B. J. Kuipers. A frame for frames. In D. G. Bobrow and A. Collins, editors, *Representation and Understanding : Studies in Cognitive Science*, pages 151–185. Academic Press, New York (NY), USA, 1975.
- [MD93] O. Marino Drews. *Raisonnement Classificatoire dans une Représentation à Objets Multi-Points de Vue*. PhD thesis, Université Joseph Fourier - Grenoble 1, 1993.
- [MEHS00] D. Mark, M. Egenhofer, S. Hirtle, and B. Smith. Ontological foundations for geographic information science, 2000.
- [MH01] R. Moller and V. Haarslev. Description of the RACER system and its applications, June 26 2001.
- [Min74] M. Minsky. A framework for representing knowledge. Technical report, Massachusetts Institute of Technology, 1974.
- [MNF02] Mark A. Musen, Natalya Fridman Noy, and Ray W. Ferguson. The knowledge model of protege-2000 : combining interoperability and flexibility. Technical report, July 26 2002.
- [Mur02] G.L. Murphy. *The Big Book of Concepts*. MIT Press, Cambridge, 2002.

- [Nap97] A. Napoli. Une introduction aux logiques de descriptions. Rapport de Recherche RR-3314, INRIA, 1997.
- [New82] A. Newell. The knowledge level. *Artificial Intelligence*, 18 :87–127, 1982.
- [Pan04] J.Z. Pan. *Description Logics : Reasoning Support for the Semantic Web*. PhD thesis, School of Computer Science, The University of Manchester, Oxford Rd, Manchester M13 9PL, UK, 2004.
- [PC99] Martin Purvis and Stephen Crane. UML as an ontology modelling language, May 21 1999.
- [Ps93] Peter F. Patel-schneider. Description-logic knowledge representation system specification from the KRSS group of the ARPA knowledge sharing effort, January 15 1993.
- [PSZ+97] C. Parent, S. Spaccapietra, E. Zimanyi, P. Donini, C. Plazanet, C. Vangenot, N. Rognon, J. Pouliot, and P-A. Crausaz. Mads : un modèle conceptuel spatio-temporel. *Revue Internationale de Géomatique*, pages 317–352, 1997.
- [PSZ99] C Parent, S. Spaccapietra, and E. Zimanyi. Spatio-temporal conceptual models : data structures + space + time. In *GIS '99 : Proceedings of the 7th ACM international symposium on Advances in geographic information systems*, pages 26–33. ACM Press, 1999.
- [Pui03] A. Puissant. *Information Géographique et Images à Très Haute Résolution : Utilité et Applications en Milieu Urbain*. PhD thesis, Université Louis Pasteur, Strasbourg I, France, 2003.
- [PW02] A. Puissant and C. Weber. The utility of very high spatial resolution images to identify urban objects. *Geocarto International*, 17-1 :31–41, 2002.
- [SM98] B Smith and D M Mark. Ontology and geographic kinds. January 01 1998.
- [Smi89] E.E. Smith. Concepts and induction. In *Foundations of Cognitive Science*, pages 501–526. Cambridge, MIT Press, 1989.
- [SV00] B. Smith and A.C. Varzi. Fiat and bona fide boundaries. *Philosophy and phenomenological research*, 60(2) :401–420, 2000.
- [Ta88] A. Thayse and al. *Approche Logique de l'Intelligence Artificielle*, volume 1. Dunod Editions, 1988.
- [TK04] E. Tomai and M. Kavouras. From "onto-geonoesis" to "onto-genesis" : The design of geographic ontologies. *GeoInformatica*, 8(3) :285–302, 2004.
- [Win88] P.H. Winston. *Intelligence Artificielle*. Addison-Wesley Europe, 1988.
- [Woo75] W. A. Woods. What's in a link ? foundations in semantic networks. In D. G. Bobrow and A. M. Collins, editors, *Representations and Understanding : Studies in Computer Science*, pages 32–82. Academic Press, New York, 1975.

Annexe 1 : Un graphe RDF décrivant l'adresse d'un employé

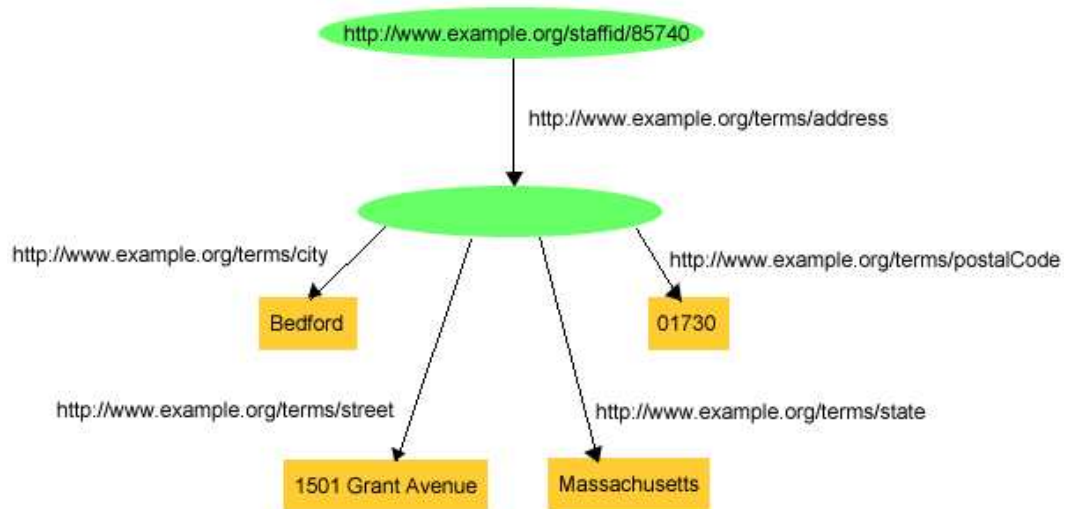


FIG. 11 – Un graphe RDF décrivant l'adresse d'un employé.

Annexe 2 : Constructeurs OWL utilisés pour construire l'ontologie

constructeur	observations
définitions de concepts	
owl :Class rdfs :subClassOf owl :intersectionOf owl :unionOf owl :complementOf	Définition d'une classe Relation d'héritage entre classes $C = D \cap E$ $C = D \cup E$ $C = \neg D$
définitions de propriétés	
owl :ObjectProperty owl :DatatypeProperty rdfs :subPropertyOf	propriété d'objet (relie des individus à des individus) propriété de type de donnée (relie des individus à des types de données) relation de spécialisation entre propriétés
caractéristiques (globales) des propriétés	
rdfs :range rdfs :domain owl :TransitiveProperty owl :SymmetricProperty owl :FunctionalProperty owl :InverseFunctionalProperty owl :inverseOf	$a \xrightarrow{R} b \implies a \in C$ $a \xrightarrow{R} b \implies b \in C$ $a \xrightarrow{R} b$ et $b \xrightarrow{R} c \implies a \xrightarrow{R} c$ $a \xrightarrow{R} b \implies b \xrightarrow{R} a$ $a \xrightarrow{R} b$ et $a \xrightarrow{R} c \implies b = c$ $b \xrightarrow{R} a$ et $c \xrightarrow{R} a \implies b = c$ $a \xrightarrow{R} b \implies b \xrightarrow{R^{-1}} a$
restrictions (locales) sur les propriétés	
owl :allValuesFrom owl :someValuesFrom owl :hasValue owl :cardinality owl :minCardinality owl :maxCardinality owl :oneOf owl :disjointWith	$a \xrightarrow{R} b \implies b \in C$ $\exists b / a \xrightarrow{R} b$ et $b \in C$ existence de valeurs de propriétés particulières pour une classe contraintes de cardinalité énumération directe des membres d'une classe classes disjointes

FIG. 12 – Constructeurs OWL utilisés pour construire l'ontologie.