
Ontologies et extraction de connaissances : une expérience dans le domaine des bases de données image pour la biologie

Arnaud DA COSTA* — **Éric LECLERCQ*** — **Arnaud GAUDIN****
— **Jean GASCUEL**** — **Marinette SAVONNET*** — **Marie-Noëlle
TERRASSE***

* *LE2I UMR CNRS 5158*

Faculté des Sciences Mirande

Aile de l'Ingénieur

BP 47870

21078 DIJON Cedex

{arnaud.da-costa,eric.leclercq,

marinette.savonnet,marie-noelle.terrasse}@u-bourgogne.fr

** *Centre Européen des Sciences du Goût - UMR 5170*

15, Rue Hughes Picardet

21000 DIJON

{gaudin,gascuel}@cesg.cnrs.fr

RÉSUMÉ. Il n'est pas toujours aisé de mettre au point des outils automatiques d'annotation d'images, notamment lorsque les images sont complexes et hétérogènes comme dans la base d'images biologique qui nous sert d'exemple. Nous proposons une approche basée sur une ontologie de domaine qui permet d'adapter l'interface d'annotation à des utilisateurs non experts du domaine.

ABSTRACT. It is not always easy to develop automatic tools to annotate images, especially when those images are heterogenous like the ones in the biology image collection we use as an example. We propose an approach based on a domain ontology that allows to adapt the annotation interface to users who are not expert in this domain.

MOTS-CLÉS : ontologie, modèle, coût d'annotation d'images

KEYWORDS: ontology, model, images annotation cost

1. Problématique

L'objectif des biologistes avec lesquels nous travaillons est de comprendre la réorganisation des réseaux de neurones cérébraux au cours de la vie post natale. Ils utilisent pour cela le modèle biologique du Xénope. Cet amphibien est très utilisé dans le cadre des recherches dans le domaine de la biologie du développement. De manière remarquable cet animal possède à l'état adulte deux systèmes olfactifs, l'un impliqué dans la détection des molécules odorantes dans l'eau et l'autre dans la détection des molécules odorantes dans l'air [GAU 04, GAU 05]. Le système olfactif aérien apparaît au cours de la métamorphose et induit une profonde réorganisation du système olfactif du têtard qui est uniquement doté d'un système olfactif aquatique. Afin d'analyser comment ces deux différents systèmes olfactifs évoluent au cours de cette réorganisation, nos partenaires utilisent de manière intensive les techniques d'imagerie confocale¹. Le système olfactif périphérique (nez) et central (cerveau) ont été principalement étudiés par des techniques de neuroanatomie mettant en œuvre successivement dissections, coupes d'organes, coloration, et acquisition d'images. Différents protocoles de coloration ont été utilisés tels que l'immunocytochimie, les lectines, les carbocyanines chacun étant spécifique des différents composants du système étudié. Les cibles de ces différents marqueurs peuvent être des cellules gliales, ou des compartiments des cellules nerveuses tels que les axones, les dendrites.

Le résultat de ce processus expérimental résulte en un nombre conséquent d'images ayant une diversité importante (comme par exemple les images des figures 1 et 2). Cette hétérogénéité trouve son origine dans :

- 1) les différentes parties du système observé ;
- 2) les différents types cellulaires (neurones, axones, etc.) ;
- 3) les différents protocoles (type de marquage de cellules, préparation des individus, etc. qui produisent des images très différentes) ;
- 4) les différentes techniques d'acquisition.

Les biologistes ne disposent pas à l'heure actuelle de protocole d'annotation, de classification ni d'archivage. Les images sont stockées sur divers formats (DVD, disque dur amovible, micro ordinateur, etc.). De fait, les seules façons d'accéder à une image particulière sont de se rappeler où elle a été stockée ou de parcourir tous les supports. L'absence de classification entraîne une impossibilité d'effectuer une recherche automatique parmi les images (par exemple pour exhiber les images ayant un ensemble précis de caractéristiques). L'information contenue dans ces images n'est donc pas ou peu exploitable.

Afin de construire une base de données images à partir de cette collection, il est nécessaire d'annoter les images. Pour cela, nous avons envisagé des approches utilisant

1. Les techniques d'imagerie confocale permettent de choisir le niveau de profondeur du plan focal utilisé pour obtenir des images d'un échantillon. Cela permet notamment de disposer d'images d'un même échantillon à différents niveaux sans nécessiter de coupe.

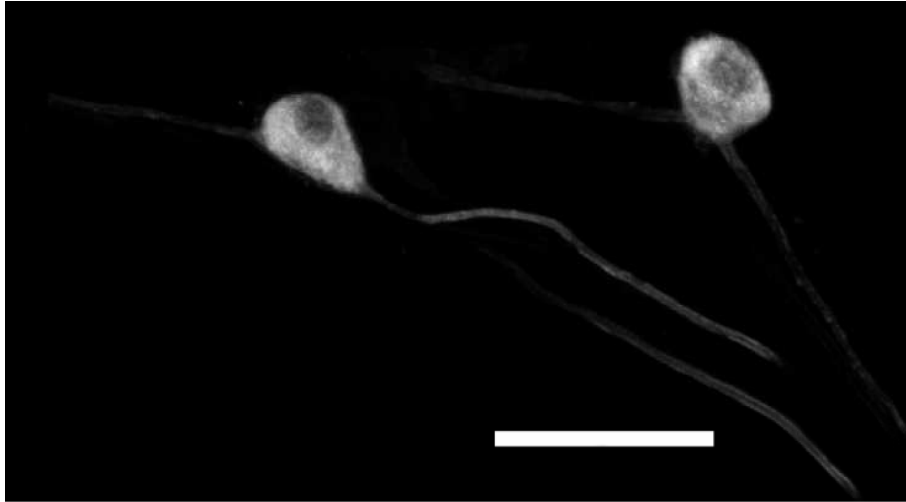


Figure 1. Morphologie de deux neurones bipolaires dans le système olfactif trigéminal. Les neurones ont été marqués avec le carbocyanine DiI. Les cellules du corps de deux neurones ainsi que leur axones et dendrites sont distinguables [GAU 04, GAU 05].

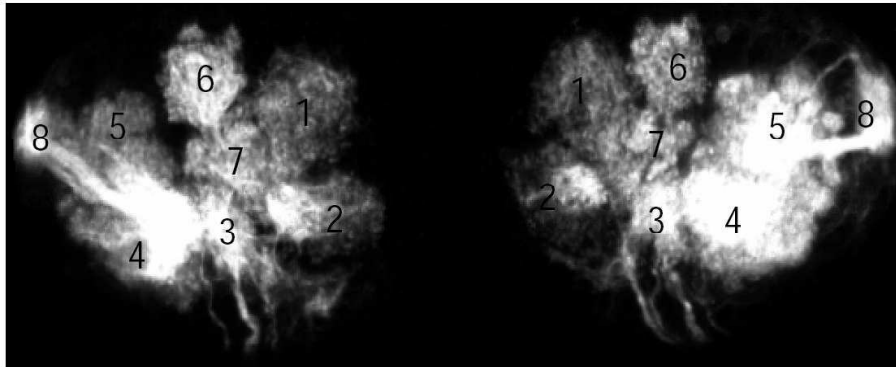


Figure 2. Vue frontale de la projection axonale dans le bulbe olfactif au stade 47 (marqueur DiI) [GAU 04, GAU 05].

des algorithmes automatiques de segmentation pour classifier les images ou reconnaître des zones spécifiques [LEM 07]. À cause de la proximité de certaines zones et de leur enchevêtrement (comme par exemple les zones 3 et 4 de la figure 2), les spécialistes en traitement du signal ont estimé que la résolution du problème de la reconnaissance automatique de ce type d'images pouvait prendre énormément de temps ; de la même façon, la reconnaissance de certains types de zones dans les images est un

problème à part entière et ne peut pas être généralisé à la collection d'images du fait des problèmes d'hétérogénéité.

Nous avons par conséquent orienté notre recherche vers l'utilisation de toutes les informations disponibles pour permettre la construction à moindre coût de cette base d'images très spécialisée. Pour diminuer le coût d'annotation, nous proposons que l'annotation ne soit pas effectuée par des experts du domaine, mais par des techniciens. Cela nécessite de pouvoir contrôler les annotations pour garantir qu'elles sont conformes à la connaissance des experts du domaine. Ceci nous a conduit à définir une classification précise des rôles des utilisateurs du système d'annotation que nous souhaitons développer. Cette classification permet ensuite de proposer une méthodologie de construction de la base d'images spécialisée qui met en action ces différents rôles autour d'une ontologie dédiée :

les techniciens du domaine sont les personnes qui seront en charge de l'annotation des images. Les techniciens du domaine ont une connaissance limitée de la sémantique du domaine : ils peuvent être entraînés pour reconnaître certaines configurations visuelles sans connaître tous les concepts biologiques sous-jacents ;

les experts du domaine sont les personnes qui maîtrisent la connaissance du domaine. Ils peuvent lever les ambiguïtés potentielles ; ce sont eux qui apportent des modifications à la connaissance du domaine ;

les spécialistes en représentation des connaissances servent d'interface entre les experts du domaine et les techniques qui serviront à représenter la connaissance du domaine. Ces spécialistes ont une expérience significative dans les bonnes pratiques pour la représentation d'un domaine de connaissance (création de la représentation, modification et suppression d'éléments de la représentation, etc.).

Cet article est organisé de la façon suivante. La section 2 présente un panorama sur la représentation des connaissances puis précise nos choix de représentation. Les sections 3 à 4 traitent plus spécifiquement des ontologies : notre expérience de construction d'ontologie ainsi que les avantages de l'utilisation des ontologies dans le cas de la base d'images de Xénopes. La section 5 présente les perspectives de nos travaux à moyen terme.

2. Éléments de représentation des connaissances

Les Systèmes d'Information (SI) sont souvent construits en ciblant uniquement des besoins spécifiques [WAN 00, USC 96], ce qui conduit à l'existence d'un grand nombre de SI, utilisant chacun sa propre méthode pour manipuler des concepts similaires pour représentant des connaissances d'un même domaine. Cette représentation des connaissances dans un SI peut être considérée selon les deux dimensions introduites par [SPE 06] pour les ontologies :

– la dimension horizontale (ou pertinence) a pour objectif de déterminer l'étendue de l'information qui devra être incluse dans la représentation de connaissance. Par exemple, si l'on représente la connaissance du domaine des *Xénopes*, la pertinence

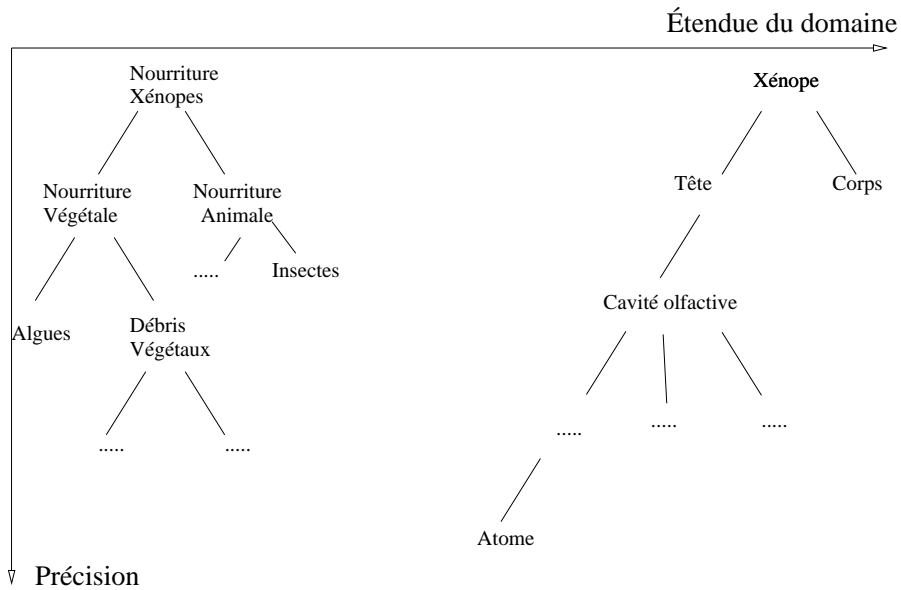


Figure 3. Dimensions horizontale et verticale de la représentation de connaissances

est le choix d’inclure ou non la connaissance de l’environnement aquatique dont le *Xénope* tire sa nourriture ;

– la dimension verticale (ou granularité) a pour objectif de déterminer le degré de finesse de la représentation des connaissances. Dans le domaine des *Xénopes*, la granularité est le choix d’inclure ou non la description du domaine jusqu’au niveau de l’atome.

La figure 3 présente un exemple de représentation de ces deux axes qui nous servent de base pour comparer les modèles et ontologies dans le cadre de la représentation des connaissances.

2.1. Représentation des connaissances : points communs entre modèles et ontologies

Les ontologies et les modèles sont similaires pour la représentation des connaissances dans la mesure où ils proposent tous deux des modes de description à base de concepts et de relations entre ces concepts. Les ontologies et modèles mettent en avant la réutilisation. Dans les deux cas la réutilisation est rendue possible en publiant sous format ouvert des spécifications qui sont le plus souvent complétées par des guides de bonne pratique (voire par des outils). Les ontologies et modèles peuvent être intégrés pour faciliter l’interopérabilité des applications. Dans les deux cas, la connaissance intégrée représente le consensus reconnu par les applications qui interagissent. Les

deux formalismes sont utilisés pour la définition de standards et permettent plusieurs types d'interopérabilité :

L'interopérabilité technique a pour objectif de faire en sorte qu'il soit possible, à partir de différents outils, d'accéder à la connaissance représentée. Chacune des méthodes de représentation des connaissances permet plusieurs moyens d'accès à la connaissance qu'elle contient :

- pour les modèles, l'interopérabilité est possible une fois que le modèle et sa représentation technique sont connus. Par exemple des classes Java dont la persistance est assurée en utilisant HIBERNATE, JPOX [HAT , JPO], ou encore des données stockées dans des fichiers XML dont la structure est définie dans un fichier au format XML-schema ;

- pour les ontologies, l'interopérabilité est possible une fois que le format de stockage est choisi pour l'ontologie (par exemple OWL[MCG 04], DAML+OIL[HOR], etc.).

L'utilisation d'un format ouvert (par exemple la licence *W3C Document Notice/License* [W3C 04b] proposée par le W3C) de représentation des connaissances peut être un atout supplémentaire dans un objectif d'interopérabilité. Cela permet à tous de savoir comment accéder aux connaissances représentées.

L'interopérabilité sémantique a pour objectif de pouvoir utiliser des sources de connaissance sans qu'il y ait d'ambiguïté sur la signification des éléments qu'elles contiennent. Par exemple dans le web sémantique, l'enjeu est de s'assurer que la sémantique récupérée depuis chaque source est correctement interprétée. Dans le cas des ontologies, l'interopérabilité sémantique est obtenue au cours de discussions entre les experts du domaine, puis en réalisant une ontologie qui soit consensuelle pour les experts et qui servira de référence de connaissance pour les travaux dans le domaine, comme l'est par exemple Gene Ontology [GEN 07]. Dans le cas des modèles, l'interopérabilité sémantique est favorisée en utilisant des modèles qui sont conformes à un même métamodèle.

La réutilisation de la modélisation : l'ouverture du format et la pérennité d'un format de stockage d'une ontologie permet de réutiliser ses propres composants (tels qu'ils sont considérés dans l'orienté objet) d'une application à l'autre, ou de réutiliser des composants développés par d'autres. Cette réutilisation diminue le temps nécessaire au développement et le coût d'une application.

2.2. Représentation des connaissances : différence entre modèles et ontologies

L'architecture de métamodélisation proposée par l'OMG [UML00] assigne aux modèles et métamodèles des niveaux d'abstraction bien délimités (voir figure 4). Il est difficile de faire apparaître dans un même modèle des éléments de description généraux et des éléments relevant de détails très fins, sauf à prendre le risque de proposer

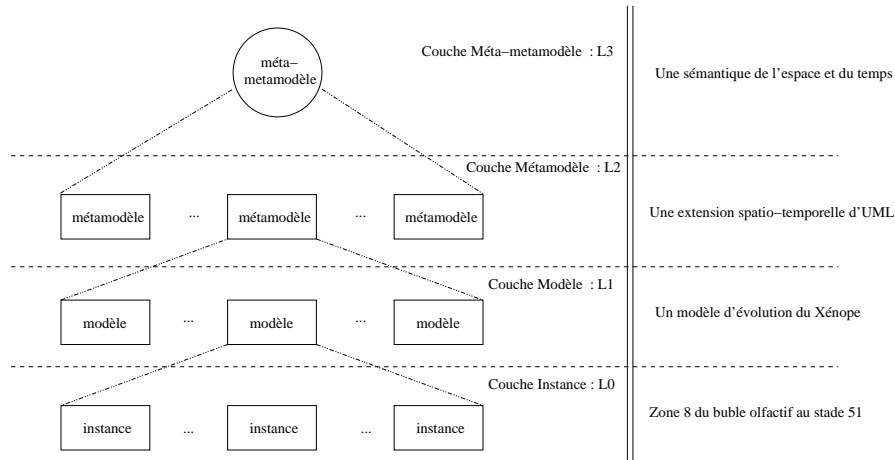


Figure 4. Les quatre couches de l'architecture des métamodèles basé sur UML, d'après [UML00]

un modèle difficilement lisible. Un modèle peut par contre utiliser plusieurs sources de représentation des connaissances, et permet donc d'ajuster l'étendue de la connaissance qu'il couvre. Dans l'exemple de la figure 3, un modèle pourra par exemple combiner la connaissance du domaine du Xénope et la connaissance de son environnement aquatique. De nombreux outils exploitent les modèles, comme les modeleurs UML tel que ArgoUML [COL], et il existe des environnements permettant de gérer les modèles et leurs transformations, comme par exemple Eclipse [ECL].

GRUBER définit le terme ontologie comme "*la spécification d'une conceptualisation : une ontologie est une description (au sens description formelle d'un programme) des concepts et de leurs relations*" [GRU 93]. Les ontologies fournissent un moyen unique de représentation de la connaissance d'un domaine, en faisant référence à un ensemble de concepts qui peuvent être utilisés pour décrire une partie de la connaissance ou en construire une représentation [SWA 99]. Les ontologies offrent une grande liberté dans la gestion de la granularité (dimension verticale) de la représentation des connaissances, cependant chaque ontologie couvre un domaine limité : sur la figure 3, chacune des deux arborescences peut être considérée comme une ontologie d'un domaine particulier. Nous appelons branche d'une ontologie une hiérarchie *is_a* avec une seule racine. Pour pouvoir couvrir un plus vaste domaine de connaissance, il faut envisager des relations entre plusieurs branches. C'est ce que proposent Grenon et al. [GRE 04] avec la définition des trois types de relations suivantes :

- intra-ontologie : relation entre deux concepts d'une même branche d'une ontologie ;
- trans-ontologie : relation entre un concept d'une branche et un concept d'une autre branche ;

– méta-ontologie : relation entre un concept d’une ontologie et une autre ontologie (considérée dans son ensemble).

Au delà de leur structure qui permet de proposer un vocabulaire structuré dans un domaine, les ontologies permettent de définir des règles logiques représentant des contraintes du domaine. Cette couche logique peut être manipulée par des outils qui vont permettre de vérifier la sémantique contenue dans l’ontologie. Dans cette optique, le W3C propose le langage OWL [MCG 04], basé sur XML. L’utilisation de règles OWL pour représenter la connaissance d’un domaine peut garantir un niveau minimum de cohérence sémantique au sein de l’ontologie et pour l’utilisation qui est faite de cette ontologie. Ce niveau minimum de cohérence sémantique permet de disposer d’un avantage pour la navigabilité et la recherche : la nature hiérarchique des ontologies peut être une aide précieuse lors de la navigation entre les concepts d’une ontologie (qui vont du plus général au plus particulier). Lors d’une recherche, si la réponse ne contient pas de résultat ou trop peu, il peut être intéressant d’utiliser le concept parent de l’un (ou plusieurs) des concepts apparaissant dans la requête, ou sous certaines conditions, un concept frère (ayant même parent) pour trouver des résultats approximatifs [MEN 99].

Dans le domaine de la représentation des connaissances, les modèles et les ontologies présentent donc chacun des avantages selon des axes qui sont orthogonaux : l’étendue du domaine pour les modèles, la granularité pour les ontologies. De plus, si les modèles comme les ontologies permettent une représentation à base de concepts et de relations, leur organisation en terme de relations est différente. Les ontologies privilégient les relations de spécialisation et contrôlent strictement les autres relations pouvant être mises en jeu (comme par exemple les treize relations de l’ontologie “OBO Relation Ontology” [MUN]). Les modèles pour leur part laissent une grande liberté dans le choix du jeu des relations à utiliser.

2.3. Choix de représentation des connaissances

Nous avons souhaité tirer parti des avantages des ontologies et des modèles et nous proposons donc une approche combinant ontologie et modèle pour l’annotation d’une base d’images très spécialisée. Nous présentons dans cet article la partie de notre travail qui traite des ontologies (travail réalisé avec OWL).

OWL a été créé pour être utilisé pour des applications qui ont besoin de manipuler l’information dans le web sémantique [W3C 04a, MCG 04] (par opposition avec celles qui se contentent de présenter l’information). Le format OWL permet de représenter des ontologies et de les stocker dans un fichier s’appuyant sur la syntaxe de RDF qui est lui même au format XML. OWL fournit trois sous-langages [BEC 04] avec un niveau d’expressivité croissant, créés pour être utilisés par des communautés spécifiques d’implémenteurs et d’utilisateurs. Nous travaillons avec OWL-DL avec lequel il est toujours possible d’utiliser des outils formels pour effectuer des vérifications et des inférences en fonction des relations de l’ontologie (et ce que ce soit des

Nom	Technologie	Licence
CEL	LISP	gratuit (usage non-commercial)
Cerebra Engine	C++	commerciale
FaCT++	C++	gratuit et open-source
fuzzyDL	Java/C++	libre
KAON2	Java	libre (gratuit pour un usage non-commercial)
MSPASS	C	gratuit open-source
Pellet	Java	libre et open-source
QuOnto	Java	gratuit ((gratuit pour un usage non-commercial)
RacerPro	lisp	commerciale (versions d'essais et licence recherche disponibles)

Tableau 1. *Un échantillon de raisonneurs (d'après [SAT 06])*

relations intra-ontologie, trans-ontologie ou méta-ontologie [GRE 04]). Le Tableau 1 liste quelques raisonneurs et leur type de licence, d'après [SAT 06].

Une fois le choix effectué pour représenter les connaissances du domaine, la construction de l'ontologie a pu être envisagée ; la section suivante présente plusieurs méthodologies possibles pour cette construction et détaille celle que nous avons retenue.

3. Méthodes de construction d'ontologies

Un des problèmes majeurs qui peut survenir lorsque l'on souhaite représenter la connaissance d'un domaine est que cette connaissance peut ne pas être la même entre différents experts d'un domaine. Par exemple, deux experts différents peuvent l'un considérer que la barbille d'un Xénope doit être complètement formée pour qu'elle soit considérée comme existante, l'autre considérer que dès qu'elle est discernable, elle existe.

Dans une optique de partage des connaissances, cela introduit une contrainte sérieuse dans la représentation des connaissances : pour être pertinente, l'ontologie ne peut pas être une représentation exhaustive de la connaissance spécifique d'un domaine. Pour que les raisonnements automatiques s'appuyant sur l'ontologie soient pertinents, il est nécessaire que cette ontologie ne représente que la partie du domaine qui fait consensus [FÜ 02]. À partir de ce consensus, la construction d'une ontologie de domaine peut alors être mise en œuvre en deux étapes distinctes :

l'amorçage détermine la méthodologie initiale qui va permettre d'assembler les différents concepts de l'ontologie ;

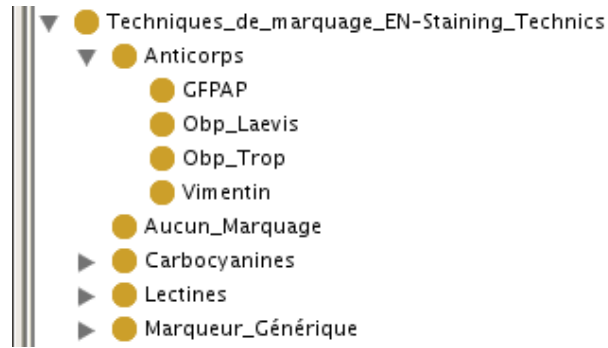


Figure 5. Le concept *Anticorps* peut être raffiné.

le raffinement détermine la méthodologie qui sera utilisée pour obtenir une ontologie qui sera considérée comme un consensus à la fois entre les experts du domaine, mais aussi entre les experts du domaine et les experts en représentation de connaissance.

Nous détaillons par la suite ces deux étapes de constructions d'une ontologie.

3.1. Amorçage de la construction d'une ontologie

Il existe trois familles d'approches présentées dans [NOY 01] pour amorcer la construction d'une ontologie :

les méthodes descendantes partent des concepts les plus généraux du domaine de discours pour les raffiner. Dans le domaine des Xénopes, une telle approche revient à partir de concepts de haut niveau tels que "techniques de marquage", qui sont raffinés en concepts fils (tel que *Anticorps*), ces concepts fils pouvant eux même être raffinés, voir figure 5 ;

les méthodes ascendantes partent des concepts les plus spécifiques et ont pour principe de les regrouper sous des concepts plus généraux. Dans le domaine des Xénopes, cette approche revient à lister des concepts tels que *Ebop*, *Nerf Trigéminal*, *Nerf Olfactif*, *Buble Olfactif Principal*, puis à regrouper ces concepts sous des concepts plus généraux (les nerfs, le bulbe olfactif, etc.)

les méthodes combinées avec lesquelles, les concepts saillants du domaine de discours sont exhibés puis ils sont généralisés ou spécialisés selon les principes des méthodes ascendantes ou descendantes.

3.2. De l'amorçage à une ontologie opérationnelle : le raffinement

L'ontologie telle qu'elle existe après une phase d'amorçage est perfectible. En effet, il est possible que la connaissance du domaine exprimée par les experts du do-

Les barbilles sont présentes sur les têtards mais pas à tous les stades. (Les barbilles arrivent au stade 44 et disparaissent au stade 61)
à partir du Stade 47 : Cavité olfactive formée d'une seule cavité. Organe voméronasal observable Au niveau du bulbe on distingue 7 zones dans le bulbe olfactif principal tandis que la zone 8 constitue le bulbe olfactif accessoire. Les zones 1,5,6, sont en position dorsales, les zones 2,3,4 sont en position ventrales. La zone 8 est visible à la fois en position ventrale et dorsale. Toutes les zones ne sont visibles que selon l'axe d'observation antero-postérieur. La zone 7 n'est visible que dans cet axe d'observation. La zone 7 peut être subdivisée en 3 sous zones 7a, 7b, 7c, et la zone 3 en 2 sous zones 3e et 3i. Les zones SIAAlpha, SIBeta, SIGama, et SIDelta sont également visibles. Les tailles des glomérules sont comprises entre 13 micromètre et 17 micromètre à partir du Stade 51 : Apparition de la cavité médiane (système olfactif aquatique)
à partir du Stade 52 : Les neurones de la cavité médiane atteignent le bulbe olfactif
à partir du Stade 59 : les axones de la cavité principale disparaissent du bulbe ventral
Adulte : stades 66 et ultérieurs

Tableau 2. Première ébauche de règles du domaine par l'expert du domaine. Le découpage selon les stades exhibe déjà l'importance de délimiter chaque règle à un ensemble précis de stades. À ce stade, les règles sont encore exprimées avec des complexités hétérogènes.

maine ne soit pas reflétée de façon fidèle ni exhaustive (dans la limite du consensus sur le domaine). Il est donc nécessaire de raffiner cette ontologie, et de vérifier par des discussions entre les experts du domaine et les spécialistes en représentation des connaissances que l'ontologie décrit correctement le domaine. Cette vérification nécessite la mise en place de cycles de discussion entre les experts du domaine et les spécialistes en représentation de connaissances [SUR 02] ; l'ontologie initiale puis les versions modifiées servent de base à ces discussions.

3.3. Construction de l'ontologie pour la base de données images Xénope

Les méthodes ascendantes et descendantes de construction d'ontologie présentent un inconvénient : elles limitent le travail à des opérations de généralisation ou de spécialisation. Nous avons souhaité conserver une certaine souplesse et pouvoir généraliser et spécialiser des termes durant la construction de l'ontologie ; c'est pourquoi notre approche pour le début de la construction de l'ontologie a été une approche combinée. Le premier pas de notre approche a été un travail avec les biologistes pour tenter de dégager les concepts saillants dans le domaine des Xénopes. À partir des informations recueillies et dont une partie est présentée dans le tableau 2, une liste d'une trentaine de concepts a été mise au point, ce qui a permis de proposer un premier niveau à la racine de l'ontologie. Pour ce premier niveau, dans une logique similaire à celle de [SCH 01], nous avons dissocié les aspects de l'ontologie utiles pour les annotations

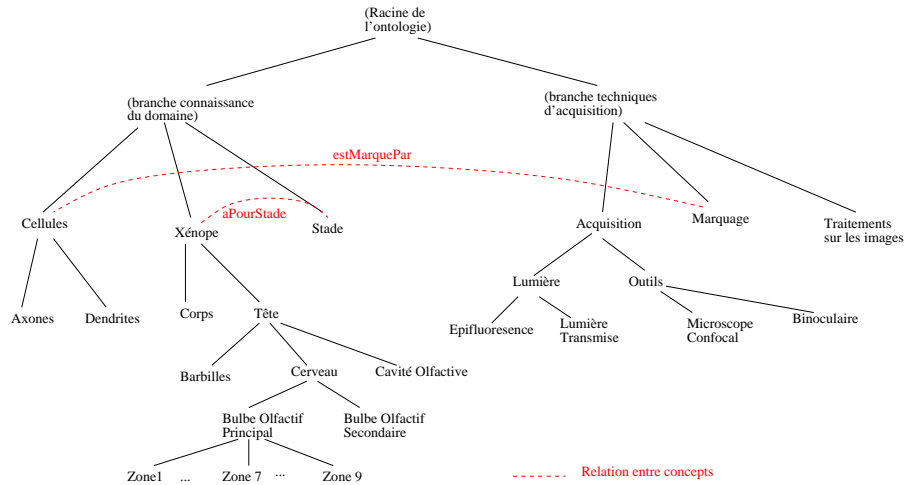


Figure 6. Hiérarchie entre quelques concepts de l'ontologie et relations que nous avons mises en place entre les différents concepts.

des aspects représentant la connaissance du domaine. Après avoir cerné les concepts importants du domaine, l'expert en représentation des connaissances établit une grille que l'expert du domaine va remplir en indiquant quelles sont les implications des métadonnées des images, et dont un extrait est présenté dans le tableau 4. la construction de l'ontologie commence alors, avec deux types de branches. Les branches suivantes contiennent la connaissance du domaine :

- Cellules : différents types de cellules connus dans le domaine des Xénopes (Axones, Cellules gliales, ...);
- Stades d'évolution : stades d'évolution des Xénopes (Stade 1, Stade 2, ...);
- Xénopes : structure biologique des Xénopes (Corps, Tête, Barbilles, Cerveau, Nerfs, ...);

Les branches suivantes contiennent la connaissance technique :

- Techniques d'acquisition :
 - différentes lumières (épifluorescence, lumière directe, etc.);
 - différents outils (microscope confocal, loupe binoculaire, etc.).
- Techniques de marquage (Anticorps, Lectines, etc.);
- Traitements sur les images (Rotation, Reconstructions 3D, etc.);

Les concepts qui n'avaient pas leur place à la racine ont été placés dans les branches de l'ontologie qui les concernent; nous avons ensuite obtenu une première ossature en raffinant et généralisant les concepts.

	EXISTE	NEXISTE PAS	VISIBLE	NON VISIBLE
EXISTE		X		
NEXISTE PAS	X		X	
VISIBLE		X		X
NON VISIBLE			X	

Tableau 3. *Incompatibilité entre les classes des différents concepts.*

Nous avons mis en place des relations entre les branches de l'ontologie relatives aux techniques d'acquisition et les branches de l'ontologie relatives à la connaissance du domaine. Par exemple la relation *estMarquePar* a pour domaine la branche *Marquage* et pour champ la branche *Cellules*. La figure 6 présente un extrait de l'ontologie Xénope et des relations qui nous ont permis de définir ensuite des classes (l'équivalent dans notre langage d'implémentation - Java - des concepts de l'ontologie) qui seront utilisées lors de l'annotation : *Existe*, *NExistePas*, *Visible* et *NonVisible*.

Nous avons décidé de ne pas nous appuyer sur la théorie du monde clos [MAZ 05]. Nous nous plaçons donc dans la théorie du monde ouvert, ce que induit que si une assertion ne peut pas être prouvée, elle est considérée comme *inconnue* (au lieu de faux dans le *monde clos*). Cela évite qu'un élément qui n'est pas considéré comme *Visible* ne soit considéré comme *NonVisible* : il est possible que la connaissance du domaine ne permette pas de déterminer le caractère *Visible* ou *NonVisible* d'un élément. Avec la théorie du *monde ouvert*, chaque appartenance d'un élément à l'une des classes *Existe*, *NExistePas*, *Visible* et *NonVisible* ne peut alors pas être remise en question puisqu'elle dépend directement de la connaissance exprimée dans l'ontologie et des propriétés de l'élément en question.

En fonction des informations connues sur les différentes instances, nous souhaitons définir quand cela est possible l'appartenance à la classe *Visible* ou à la classe *NonVisible*. Comme un élément ne peut être considéré à la fois *visible* et *non visible*, nous avons défini ces deux classes comme étant disjointes l'une de l'autre. De la même façon, les classes *Existe*, *NExistePas* sont disjointes. Le tableau 3 récapitule les incompatibilités d'appartenance d'un élément aux classes correspondant aux différents concepts. À partir des concepts du domaine retenus et des règles de compatibilité, le spécialiste en représentation de connaissance a établi un tableau afin que l'expert du domaine puisse représenter la connaissance de son domaine d'une manière plus formelle. Le tableau 4 présente quelques règles du domaine ; en l'état actuel, l'ontologie comporte deux cent quatre-vingt-dix règles. Sur l'ensemble de ces règles seul un sous ensemble limité ne contient que des règles atomiques (i.e. : ne comportant qu'un seul critère) :

- les règles atomiques comme par exemple "au stade 47, la zone 9 EXISTE" sont syntaxiquement et sémantiquement correctes, mais n'ont qu'un intérêt limité dans le cadre de l'annotation : elle peuvent au mieux servir à vérifier qu'aucun élément n'a été annoté de façon incohérente ;

– les règles plus complexes comme par exemple "au stade 51, avec le marqueur SBA et en observation confocale, la zone 5 est visible" présentent un intérêt plus fort que les règles atomiques dans le cadre de l'annotation car elles véhiculent plus de sémantique.

Le spécialiste en représentation des connaissances a donc guidé l'expert du domaine dans la rédaction des règles pour obtenir des règles ayant à la fois une réalité visuelle sur les images et une sémantique forte.

4. Une expérience dans le domaine de la biologie : exemple d'utilisation de l'ontologie

Cette section traite l'apport de l'ontologie pour la classification, puis de l'utilisation que nous en faisons dans le cadre de l'annotation d'images.

4.1. Utilisation de l'ontologie pour la classification

Avec la sémantique représentée dans l'ontologie, il est intéressant d'utiliser un raisonneur pour effectuer des déductions et vérifications automatiques [BEN 01] :

- vérifier qu'une classe donnée généralise une autre classe, en se basant sur les propriétés de chacune d'elles. Cette vérification est appelée *classification de la taxonomie* dans Protégé ;
- définir à quelle classe(s) appartient un élément donné. Cette étape porte le nom d'*inférence des types* dans Protégé.

En utilisant les relations entre les branches (*aPourStade*, *aPourTypeDeVue*, *est-MarquePar*, etc.), nous avons pu inclure la connaissance du domaine exprimée par les experts (tableau 4) dans la définition des classes sous Protégé. Dans son second bloc "*Necessary & Sufficient*", la figure 7 montre un exemple de définition de la classe *Existe* conformément à la connaissance du domaine exprimée dans la ligne 9 du tableau 4. En créant deux instances "Image001" et "Image002" (figures 8 et 9), puis en utilisant le raisonneur, il est possible de déterminer automatiquement que seul "Image001" satisfait à la définition de la classe *Existe* (figure 10). Par contre les informations relatives à "Image002" ne correspondent à aucune règle donnée par l'expert : cela ne permet pas de déterminer automatiquement l'appartenance de "Image002" à l'une des classes *Existe*, *NExistePas*, *Visible* et *NonVisible*

4.2. Utilisation de l'ontologie pour l'annotation

En s'appuyant sur la connaissance du domaine et sur les outils de raisonnement, il est possible d'effectuer des déductions automatiques grâce aux raisonneurs. La théorie du monde ouvert [MAZ 05] nous impose d'être exhaustifs dans la représentation des

	Concept			Implication			
	Stade	marqueur	moyen d'observation	EXISTE	NEXISTE PAS	VISIBLE	NON VISIBLE
1	47				zone 9		
2	47				glomérules dans zones 1,3,7		
3	47			zone 8			
4	47			zone 7a			
5	47			zone 7b			
6	47			zone 7c			
7	47			Si alpha			
8	51-57	SBA	confocal	zone 9			zone 9
9	57-59	SBA	confocal	zone 5		zone 5	
10	57-59	SBA	confocal	zone 3e		zone 3e	
11	57-59	SBA	confocal	EBOP		EBOP	

Tableau 4. Certaines règles véhiculent une sémantique simple : La zone 9 n'existe pas au stade 47 ; d'autres sont plus complexes : dans les stades 51 à 57, même si la zone 9 existe, elle ne sera pas exhibée par un marqueur de type SBA en imagerie confocale. Voir la figure 7 pour une implémentation dans Protégé de la règle qui concerne la zone 5 aux stades 57 à 59.

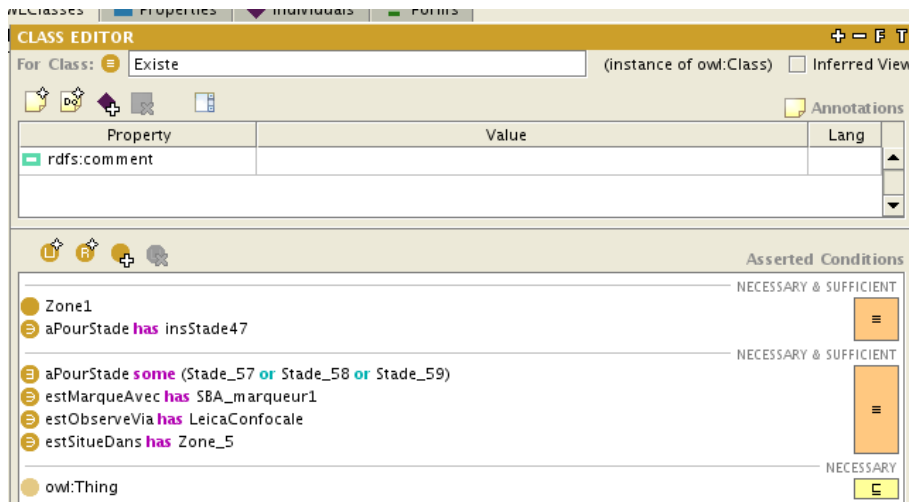


Figure 7. Définition de la classe “Existe”. Le premier bloc “Necessary & Sufficient” indique qu’un élément de la Zone 1 existe au stade 47. Le second bloc indique qu’un élément de la Zone 5, de stade 57, 58 ou 59, marqué avec le marqueur SBA1 (un marqueur de type SBA) et observé avec une LeicaConfocale existe.

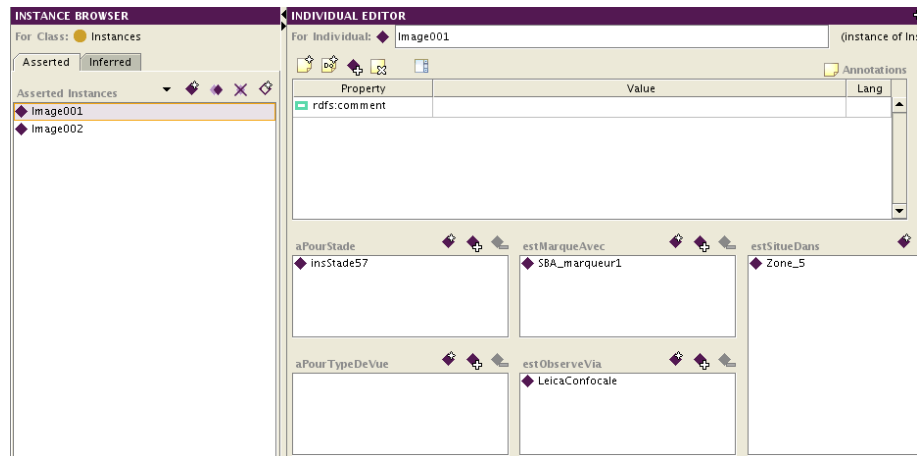


Figure 8. Définition d’une instance “Image001” dans Protégé, et saisie des relations connues pour cette instance : son stade, le marqueur utilisé et le moyen d’observation. L’instance “Image001” est créé pour être conforme aux règles du domaine du tableau 4, ligne 9.

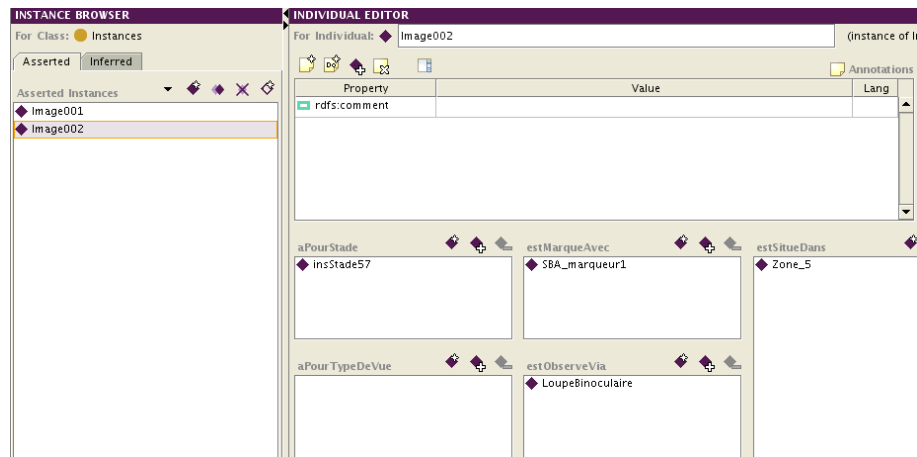


Figure 9. Définition d’une instance “Image002” qui diffère de “Image001” uniquement par la méthode d’observation.

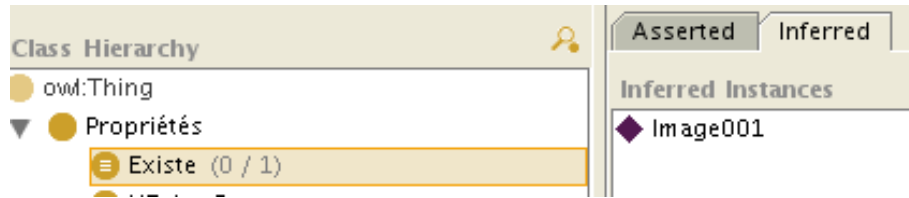


Figure 10. L'indication (0/1) à côté de la classe *Existe* indique qu'aucun élément n'a été créé comme étant visible, mais que le raisonneur a déterminé qu'un élément appartient à cette classe. Les définitions de la classe et l'usage du raisonneur ont permis d'inférer que "Image001" existe.

connaissances, mais elle présente l'avantage de n'aboutir qu'à des résultats de calcul conformes aux règles exprimées dans le domaine. Au fil de l'annotation d'une image, il est possible d'utiliser un raisonneur avec les annotations en cours pour pouvoir déterminer un certain nombre de propriétés :

Liste des concepts pertinents : l'utilisation des règles de l'ontologie et du raisonneur RacerPro [GMB 04] (licence académique, voir tableau 1) permet de :

- vérifier la cohérence de l'ontologie au format OWL-DL ;
- effectuer une classification automatique des classes en fonction de leur définitions ;
- inférer les types des différentes instances en fonction des relations et règles.

Il est alors possible de déterminer au fil de l'annotation d'une image quels sont les autres éléments qui existent. Le choix des concepts proposés pour la suite de l'annotation de cette image sera ainsi limité aux seuls concepts qui restent possible. Par exemple si un élément est annoté avec les concepts de la ligne 1 du tableau 4, il ne sera pas possible d'annoter cet élément avec le concept *zone 9* qui est considéré comme n'existant pas. Au fil de l'enrichissement des annotations sur un élément, seuls les concepts qui ne sont pas considérés comme *impossibles* restent disponibles pour poursuivre l'annotation de cet élément ;

Topologie des images : par exemple le stade avec lequel une image est annotée déterminera si une zone existe ; l'ajout d'annotations sur le type de vue et le type de marquage pourra permettre de déterminer si une zone existante est visible. En fonction de ces informations, il sera possible de proposer une topologie par défaut des différentes zones. Une telle topologie doit être déterminée par un expert du domaine et le technicien pourra adapter cette topologie par défaut à la réalité de l'image. Par exemple, sur la figure 11, les zones visibles et à détecter sont schématisées par des ellipses ; le technicien peut alors ajuster le positionnement des différentes ellipses sous réserve de contraintes exprimées par l'expert du domaine, et la topologie des différentes zones après annotation sera alors stockée dans une base de données.

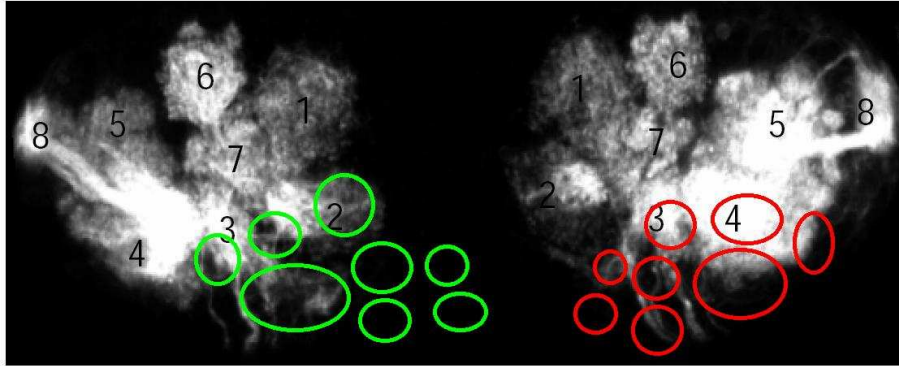


Figure 11. *En fonction des annotations, il a été possible de proposer à l'utilisateur une topologie des zones du bulbe olfactif qui existent (membre de la classe Existe) et qui sont visibles (membre de la classe Visible).*

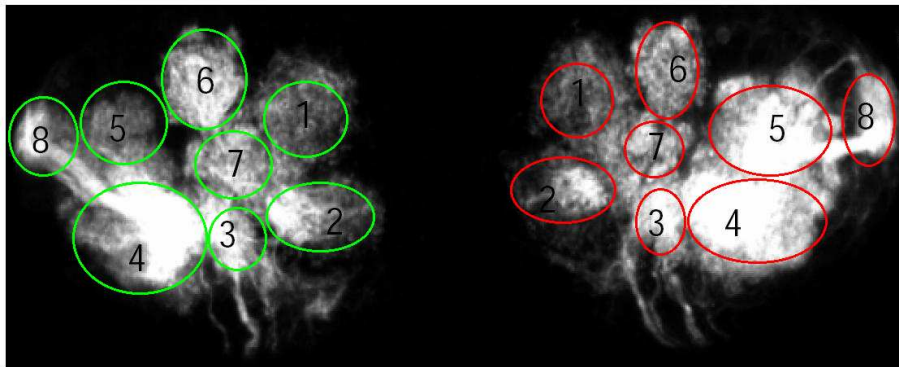


Figure 12. *La topologie proposée est ensuite ajustée à la réalité de l'image.*

5. Conclusion

Nous avons présenté dans cet article les points forts et points faibles respectifs des ontologies et des représentations basées sur les modèles pour la représentation des connaissances. La combinaison des deux types de représentation permet d'obtenir une granularité fine et d'avoir une marge intéressante pour la représentation d'un domaine particulier de connaissance : le Xénope. À partir de cette combinaison, nous avons proposé une aide pour permettre l'annotation d'images d'un domaine particulier par des techniciens de ce domaine. Ceci permet une réduction du coût de l'annotation d'une base d'images tout en proposant un mécanisme de contrôle de la sémantique des annotations. L'ontologie que nous avons présentée dans cet article termine son amorçage [SUR 02], et va être modifiée au fil de discussions avec les experts du do-

maine. Ce raffinement sera l'occasion pour nous de réutiliser des ontologies formelles telles que BFO [SPE 06], et de proposer une traduction en anglais de l'ontologie. La réalisation en Java d'un prototype d'annotation est en cours, elle utilise l'API Jena [JEN 04] et une base de données PostgreSQL [POS] pour le stockage.

6. Bibliographie

- [BEC 04] BECHHOFFER S., VAN HARMELEN F., « OWL Web Ontology Language Reference », 2004, <http://www.w3.org/TR/owl-ref/>.
- [BEN 01] BENCH-CAPON T. J. M., « The role of ontologies in the verification and validation of knowledge-based systems », *Int. J. Intell. Syst.*, vol. 16, n° 3, 2001, p. 377-390.
- [COL] COLLABNET, « ArgoUML », <http://argouml.tigris.org/>.
- [ECL] ECLIPSE FOUNDATION, « Eclipse - an open development platform », <http://www.eclipse.org/>.
- [FÜ 02] FÜRST F., « L'Ingénierie Ontologique », rapport, 2002, Institut de recherche en Informatique de Nantes.
- [GAU 04] GAUDIN A., « Analyse des réorganisations neuroanatomiques des projections olfactives au cours de la métamorphose chez le xénope et la recherche des mécanisme moléculaires sous-jacents », PhD thesis, Centre Européen des Sciences du Goût, 2004.
- [GAU 05] GAUDIN A., GASCUEL J., « 3D Atlas Describing the Ontogenic Evolution of the Primary Olfactory Projections in the Olfactory Bulb of *Xenopus Laevis* », *The Journal Of Comparative Neurology*, , 2005.
- [GEN 07] GENE ONTOLOGY CONSORTIUM, « An Introduction to the Gene Ontology », 2007, <http://www.geneontology.org/G0.doc.shtml>.
- [GMB 04] GMBH R. S., « RacerPro : OWL reasoner and inference server for the Semantic Web », 2004, <http://www.racer-systems.com/>.
- [GRE 04] GRENON P., SMITH B., « SNAP and SPAN : Towards dynamic spatial ontology », *Spatial Cognition and Computation*, vol. 1, 2004, p. 69-104, Kluwer Academic Publishers.
- [GRU 93] GRUBER T., « What is an Ontology ? », 1993, <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
- [HAT] HAT R., « Hibernate : Relational Persistence for Java and .NET », <http://www.hibernate.org/>.
- [HOR] HORROCKS I., VAN HARMELEN F., BERNERS-LEE T., « Open Biomedical Ontologies (OBO) », <http://www.daml.org/2000/12/daml+oil-index>.
- [JEN 04] JENA, « Jena – A Semantic Web Framework for Java », 2004, <http://jena.sourceforge.net/>.
- [JPO] JPOX, « Java Persistent Objects », <http://www.jpox.org/>.
- [LEM 07] LEMAITRE C., MITERAN J., MATAS J., « Definition of a model-based detector of curvilinear regions », *International Conference on Computer Analysis of Images and Patterns (CAIP)*, 2007.
- [MAZ 05] MAZZOCCHI S., « Closed World vs. Open World : the First Semantic Web Battle », 2005, <http://www.betaversion.org/~stefano/linotype/news/91/>.

- [MCG 04] MCGUINNESS D. L., VAN HARMELEN F., « OWL Web Ontology Language », 2004, <http://www.w3.org/TR/owl-features/>.
- [MEN 99] MENZIES T., « Cost benefits of ontologies », *Intelligence*, vol. 10, n° 3, 1999, p. 26–32, ACM Press.
- [MUN] MUNGALL C., « OBO Relation Ontology », <http://www.obofoundry.org/ro/>.
- [NOY 01] NOY N. F., MCGUINNESS D. L., « Ontology Development 101 : A Guide to Creating Your First Ontology », 2001.
- [POS] POSTGRESQL, « PostgreSQL the object-relational database system », <http://www.postgresql.org/>.
- [SAT 06] SATTLER U., « Description Logic Reasoners », 2006, <http://www.cs.man.ac.uk/~sattler/reasoners.html>.
- [SCH 01] SCHREIBER A. T. G., DUBBELDAM B., WIELEMAKER J., WIELINGA B., « Ontology-Based Photo Annotation », *IEEE Intelligent Systems*, vol. 16, n° 3, 2001, p. 66–74, IEEE Educational Activities Department.
- [SPE 06] SPEAR A. D., « Ontology for the Twenty First Century : An Introduction with Recommendations », 2006.
- [SUR 02] SURE Y., STAAB S., STUDER R., « Methodology for development and employment of ontology based knowledge management applications », *SIGMOD Rec.*, vol. 31, n° 4, 2002, p. 18–23, ACM Press.
- [SWA 99] SWARTOUT W., TATE A., « Guest Editors' Introduction : Ontologies », *IEEE Intelligent Systems*, vol. 14, n° 1, 1999, p. 18–19, IEEE Educational Activities Department.
- [UML00] « OMG Unified Modeling Language Specification », March 2000, Version 1.3, Available at URL <http://www.omg.org>.
- [USC 96] USCHOLD M., GRÜNINGER M., « Ontologies : principles, methods, and applications », *Knowledge Engineering Review*, vol. 11, n° 2, 1996, p. 93-155.
- [W3C 04a] W3C, « Semantic Web », 2004, <http://www.w3.org/2001/sw/>.
- [W3C 04b] W3C C., « Intellectual Rights FAQ », 2004, <http://www.w3.org/Consortium/Legal/IPR-FAQ-20000620>.
- [WAN 00] WANG N., XU X., « A Method to Build Ontology », *HPC '00 : Proceedings of The Fourth International Conference on High-Performance Computing in the Asia-Pacific Region-Volume 2*, Washington, DC, USA, 2000, IEEE Computer Society, page 672.