
Représentation de connaissances descriptives et classificatoires : le modèle CoDesc

David Grosser* — Noël Conruyt — Yannick Geynet

Laboratoire IREMIA, Université de la Réunion,
15 avenue René Cassin, BP 7151,
97715 St-Denis Msg, Cedex 9,
La Réunion - France

{grosser, conruyt, geynet}@univ-reunion.fr

<http://www.univ-reunion.fr/coraux/>

* actuellement affilié au DIRO, Université de Montréal
grosserd@iro.umontreal.ca

RÉSUMÉ. Pour répondre aux besoins des biologistes souhaitant disposer de modèles de représentation et d'outils de gestion des connaissances adaptés à leur démarche scientifique, nous proposons un modèle orienté-objets, CoDesc, leur permettant de représenter des connaissances descriptives et classificatoires. Une entité de plus haut niveau que la classe, le modèle descriptif, est utilisée pour représenter les concepts du monde réel. Les propriétés des modèles descriptifs peuvent être de deux types, nécessaires ou contingentes, ce qui permet de considérer ces modèles à la fois comme une représentation de l'extension et de l'intension des concepts. Nous étudions l'impact de cette spécificité du modèle pour maintenir la cohérence de la relation de composition et de subsomption. Le modèle est implanté au sein d'une plate-forme Java, TKBS. Le système a été utilisé pour plusieurs applications, en systématique et en médecine.

ABSTRACT. To meet the needs of biologists wishing to have representation models and knowledge management tools adapted to their scientific method, we propose an object-oriented model, CoDesc, allowing them to represent descriptive and classifying knowledge. An entity of higher level than the class, the descriptive model, is used to represent the concepts of the real world. The properties of the descriptive models can be of two types, necessary or contingent, which makes it possible to consider these models at the same time as a representation of the extension and of the intension of the concepts. We study the impact of this specificity of the model to maintain the coherence of the relation of composition and of subsomption. The model is implemented within a Java platform, TKBS. The system was used for several applications in systematics and in medicine.

MOTS-CLÉS: Représentation de connaissances, concepts, descriptions, classification, Systématique

KEYWORDS: knowledge representation, concepts, descriptions, classification, Systematics

1. Introduction

A l'heure actuelle, se développent de nombreuses bases de données scientifiques en Systématique disponibles sur Internet ou sur CD-ROM, telles que World Biodiversity database de l'ETI en Hollande, Reefbase et Fishbase de l'ICLARM aux Philippines, Hawaii Biological Survey databases et Coral Id à l'AIMS en Australie. Les modèles relationnels ou hiérarchiques sont souvent utilisés [BEA 93], ainsi que le format de données semi-structurées DELTA [DAL 95], standard de l'OMG pour la représentation des données descriptives qui est développé et maintenu au CSIRO en Australie. Le modèle objet ne s'est que trop rarement imposé dans ce domaine [VIG 00], et les bases construites s'apparentent davantage à des systèmes d'informations taxonomiques et biogéographiques qui offrent des clefs d'identification aux biologistes [PAN 98], qu'à de véritables bases de connaissances. En fait la plupart du temps, ces systèmes reproduisent ce qui existait déjà dans la littérature : des monographies [VER 00] et des clefs d'identification, sous une forme électronique aisément accessible. Ils sont appropriés à la transmission d'un savoir encyclopédique lorsque les taxons sont bien connus et stables, mais de nombreuses difficultés apparaissent dès lors que la connaissance du domaine évolue et qu'il faut maintenir la cohérence de la base [PUL 00]. De plus, les systèmes utilisés représentent les concepts du domaine (taxons) mais ne permettent pas de représenter les individus (spécimens types), dont les descriptions sont fondamentales à la définition des concepts et à la classification [BER 97]. En fait, comme le souligne [LEB 96], il existe un réel besoin de mise en oeuvre de modèles de représentation de type classe/instance, de méthodologies d'acquisition des connaissances, ainsi que d'outils informatiques pour supporter le travail des experts et diffuser leur savoir.

Nous présentons dans cet article un nouveau modèle de représentation des connaissances orienté-objet appelé CoDesc pour CONnaissances DESCriptives et classifica-toires (§ 2), développé initialement pour la Systématique dans le cadre du projet "base de connaissances sur les coraux des Mascareignes" [CON 97a]. Dans le même esprit que le système TROEPS [SHE 95], la notion de classe et de concept est distinguée. Les concepts sont décrits par une entité de plus haut niveau que la classe appelée *modèle descriptif* (§ 2.1), qui représente l'ensemble des *connaissances de fond* relatives à un large groupe d'individus (taxon de haut niveau : Famille ou Genre par exemple) appartenant au domaine modélisé par l'expert. A chaque concept est associé un ensemble de classes qui décrivent des sous-ensembles significatifs d'instances d'un concept (taxon de bas niveau : espèce, sous-espèce ou écomorphe).

CoDesc présente cependant certaines spécificités par rapport aux modèles de représentation des connaissances par objets qui favorisent la mise en oeuvre de l'approche scientifique des biologistes de nature essentiellement inductive. En effet, la plupart de ces modèles adoptent essentiellement une **approche descendante** pour la construction d'une taxonomie de classes [VAL 99] : les classes sont formées par spécialisations successives, par ajout de nouveaux descripteurs dans les sous-classes. Le but essentiellement recherché est une caractérisation intensionnelle des classes, c'est-à-dire un ensemble de conditions nécessaires d'appartenance d'un individu à la

classe. CoDesc considère deux types de descripteurs pour définir les concepts : ceux qui peuvent être absents des descriptions sont dits *contingents*¹, car ils ne peuvent remettre en cause l'appartenance de l'individu au concept. Les autres, qualifiés de *nécessaires*, forment la partie intensionnelle du concept et doivent obligatoirement être renseignés dans les descriptions. Cette particularité du modèle offre l'avantage d'introduire des modalités (est ou n'est pas présente, parfois présente, etc.) pour représenter des concepts dans les domaines de connaissances liées à l'observation du monde réel pour lesquels une logique purement booléenne est rarement satisfaisante.

Le modèle CoDesc offre un cadre d'expression très riche pour représenter les concepts à l'aide d'un ensemble de descripteurs de différents types : schémas, composants (§ 2.2) et attributs (§ 2.3), auxquels sont attachés des relations, des règles et des données contextuelles. La sémantique originale associée à la spécialisation et à la généralisation des concepts et des composants (§ 2.4) offre une assistance aux experts pour construire une taxonomie par une **approche ascendante** : les descriptions d'individus appelées *cas* (§ 2.5) sont exprimées par des *valeurs complexes* : multiple, taxinomique (ou hiérarchisée), conjonctive, disjonctive, inconnue, absente, et sont utilisées pour la conception de nouveaux concepts plus généraux, ce qui favorise la mise au point des hiérarchies de concepts par une approche *inductive*. Il est en effet très difficile en Systématique d'élaborer *a priori* des modèles généraux, sans l'aide d'un ensemble de descriptions fines et détaillées d'individus (en particulier les spécimens types), qui expriment la variabilité et la richesse des objets de la Nature.

2. Le modèle CoDesc

CoDesc est un modèle de représentation des connaissances orienté-objet, élaboré pour la conception de bases de connaissances en Systématique. Les connaissances considérées sont de type *descriptives* et *classificatoires*. Les connaissances descriptives sont constituées des concepts modélisés du domaine étudié, ainsi que par des descriptions d'individus appartenant à l'extension de ces concepts. Les connaissances classificatoires sont représentées par des relations établies entre concepts (spécialisation et généralisation) et entre individus et concepts. Trois entités principales apparaissent dans une base de connaissances : les concepts, les cas et les relations. Chaque cas est une description d'un individu attaché à un unique concept du domaine par une relation de type *est-un*. Un cas est donc une instance de ce concept. Ceux-ci sont organisés en une hiérarchie d'héritage simple par la relation de spécialisation de type *sorte-de*. Des données contextuelles peuvent être attachées à chaque entités du domaine. Il peut s'agir de documents de type texte, d'articles, de liens hypertextuels, d'annotations : toute sorte de données informatisées utiles à la compréhension du domaine. Un ensemble de concepts et de cas, annotés et illustrés, mis en relation dans une base de connaissances, constituent une "ontologie" [GRU 93] d'un domaine de la systématique [VIG 00, ROU 01].

1. **Contingent**,e adj. Qui peut se produire ou non, être ou ne pas être (par opposition à nécessaire). *Le Petit Larousse, éd. 1998*

2.1. Les concepts

Un concept est le représentant d'une famille d'individus. Sa représentation informatique est une structure de données arborescente appelée **modèle descriptif** qui peut être interprétée comme le plan d'organisation de la description des individus appartenant au concept. Un concept est muni de prérogatives ontologiques sur l'ensemble de ses instances, car il contient l'ensemble des caractères potentiellement utiles à la description des individus et définit leur structure.

Certaines caractéristiques sont *nécessaires* : une instance doit posséder obligatoirement ces caractéristiques pour appartenir à l'extension du concept. Les autres sont uniquement descriptives, elles sont utilisées pour affiner la description de sous-groupes d'individus particuliers et sont qualifiées de *contingentes*. Elles peuvent être absentes des descriptions sans remettre en cause l'appartenance de l'individu au concept. Cette distinction est fondamentale dans le modèle proposé, elle permet de considérer les concepts selon deux points de vue :

– **Point de vue intensionnel.** La restriction du modèle descriptif dénotant un concept à l'ensemble des *caractéristiques nécessaires* permet une caractérisation de l'*intension généralisée*² du concept.

– **Point de vue extensionnel.** Un modèle descriptif possède la totalité des *descripteurs observables* pour décrire les individus. Il offre ainsi une vision synthétique complète de l'univers possible de description. Il peut être interprété comme un *représentant maximal* du concept, couverture totale de l'extension du concept.

En Systématique, les concepts considérés sont essentiellement les taxons d'un groupe taxonomique particulier : espèce < genre < famille < sous-ordre < ordre. Le choix de réifier ou non un concept par un modèle descriptif dépend de plusieurs facteurs pris en compte par les experts, comme le nombre d'espèces et la variabilité des individus, qui conditionneront la complexité du modèle développé lié au nombre de caractères nécessaires à la description de tous les individus. La figure 1 illustre le modèle descriptif dénotant la famille des *Pocilloporidae*³. Le niveau de complexité de ce modèle est acceptable pour les experts, il présente un peu plus de cent attributs qui suffisent à décrire une vingtaine d'espèces différentes appartenant à trois genres.

Dans cet exemple, les éléments qui composent ce modèle descriptif sont : la racine de l'arbre de description, qui est un composant particulier appelé schéma (1), les composants de type "point de vue" (2), les composants réels (3), les attributs (4) et les liens de composition (5). Un composant peut également être "absent possible" (7) (présence du signe - devant le composant) et caractériser la racine d'un sous-arbre possiblement absent des descriptions.

2. L'**intension généralisée** donne des conditions nécessaires d'appartenance d'un individu à une classe ou un concept, par opposition à l'**intension stricte** qui exprime des conditions *nécessaires et suffisantes*.

3. Les *Pocilloporidae* sont une des seize familles de coraux (Ordre des *Scléactiniaires*) qui comptent environ 800 espèces [VER 00].

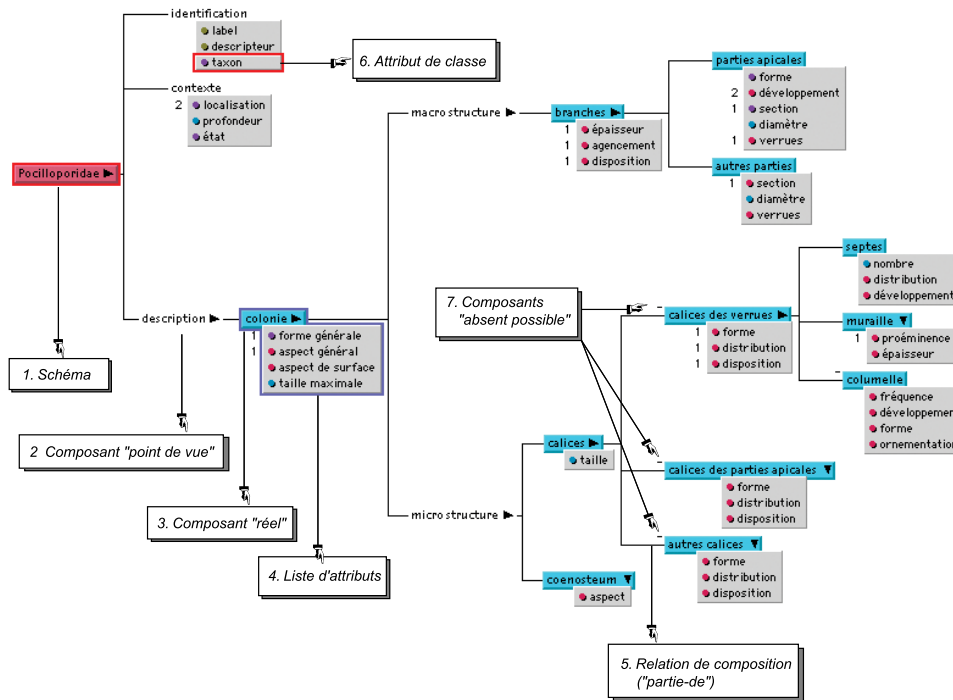


Figure 1. Le modèle descriptif de la famille des *Pocilloporidae*, vue partielle. Extrait de la “Bases de connaissances sur les coraux des Mascareignes”.

L’attribut Taxon (6) joue un rôle particulier dans le modèle. Il est appelé *attribut de classe* et le domaine de valeurs qui lui est attaché définit l’ensemble des classes significatives d’individus : les différents genres et espèces de la famille *Pocilloporidae* dans l’exemple. Cet attribut est muni d’une relation d’ordre partiel entre les valeurs du domaine (type hiérarchisé cf. 2.3). Les relations de spécialisation entre classes sont ainsi exprimées par la hiérarchie des valeurs du domaine.

La classe est considérée dans notre modèle dans un sens équivalent à celui de l’apprentissage supervisé ou de l’analyse de données : comme une étiquette attachée à chaque objet de la base. L’attribut de classe est utilisé comme variable cible par les différentes méthodes d’analyse implantées dans *TKBS* : pour la classification ou l’identification de nouveaux spécimens [CON 97b, GRO 00]. Un attribut alternatif peut être sélectionné afin de conduire une classification selon un autre point de vue ou bien spécifier une autre variable à déterminer.

Dans la version actuelle de *CoDesc*, une classe n’est donc pas réifiée comme une entité munie de propriétés et d’une structure à la manière des modèles objets classiques. Il n’est donc pas possible d’associer des propriétés spécifiques aux classes autrement que par la définition de règles associées à l’attribut de classe (cf. § 2.3).

2.2. Les composants

Un modèle descriptif est constitué d'un ensemble de composants mis en relation. Les composants du modèle CoDesc sont très semblables aux *objets composites* des modèles objets [KIM 87, NAP 92, MAR 93]. Il n'existe cependant qu'un seul type de relation de composition dans CoDesc, dont la sémantique est précisée par la nature des objets mis en relation et non explicitement, à l'inverse des modèles où sont réifiées les relations, tel que AROM [PAG 00]. En effet, la nature d'un composant peut être précisée par quatre statuts distincts : *fictif*, *absent possible*, *multiple* et/ou *spécialisable* qui peuvent être combinés [CON 94]. Par défaut, un composant ne possède aucune de ces propriétés et correspond à une entité réelle observable.

Un composant *fictif* caractérise une entité non-observable ou abstraite du monde réel. Il définit la racine d'un nouveau point de vue du modèle descriptif (composant (2) sur la figure 1), ce qui est utile pour représenter différents aspects d'un concept. Dans l'exemple proposé, le point de vue *description* définit le sous-arbre regroupant l'ensemble des caractéristiques morphologiques, *macrostructure*, celles qui sont observables à l'oeil nu, *contexte*, les paramètres du lieu de récolte des individus, etc.

Le type *absent possible* (7) permet d'exprimer des relations de *concomitance* ou d'*exclusion* particulièrement fréquentes en biologie, qui traduisent respectivement une condition de présence ou d'absence d'un caractère en fonction du "contexte" formé par d'autres caractères. Un composant absent possible caractérise un sous-arbre possiblement absent des descriptions et donc un ensemble de propriétés contingentes. Une propriété peut donc être contingente sans être explicitement *absente possible* dans la mesure où lors de la description d'un individu, l'absence d'un composant est propagée à l'ensemble des composants dépendants⁴ ainsi qu'aux attributs qui leur sont attachés. Ainsi, le composant *septes* (figure 1) n'est pas explicitement absent possible dans ce modèle développé par les experts, car la présence de *calices des verrues* entraîne systématiquement la présence de *septes* (règle de présence déduite). Par contre, certains spécimens ne possèdent pas de *calices des verrues* et par conséquent ni *septes*, ni *muraille*, ni *columelle* (absence déduite).

La *multiplicité* d'un composant est exprimée sous la forme d'un intervalle d'entiers [*inf sup*], ce qui signifie qu'il peut être instancié plusieurs fois dans la description des cas. La relation qu'il entretient avec son composant père est alors une association *n*-aire. La multiplicité offre un moyen d'exprimer la variabilité dans les observations pour un objet d'étude particulier, de renseigner plusieurs états différents observés simultanément pour un même objet.

De même qu'un concept, un composant peut être *spécialisable* en différents sous-composants, au sens de la relation de spécialisation. Nous détaillerons les mécanismes spécifiques mis en place dans le modèle CoDesc pour gérer la spécialisation et la relation duale de généralisation à la section 2.4.

4. L'ensemble des composants dépendants d'un composant donné C est constitué des composants pour lesquels il existe un chemin de la relation *est-composé-de* (transitive) d'origine C.

2.3. Les attributs

Un attribut traduit une propriété atomique du modèle. Un type unique lui est associé : symbolique, numérique, taxinomique (hiérarchisé) ou textuel. Il est défini par un identificateur unique, un domaine de définition et une valeur par défaut. Des contraintes peuvent lui être ajoutées par l'intermédiaire de règles définies par l'expert, équivalentes aux facettes procédurales des frames et des modèles objets [MAS 89]. En particulier, associées à l'attribut de classe Taxon, les règles permettent d'exprimer des **conditions suffisantes** d'appartenance d'un individu à un sous-taxon particulier, lorsqu'une condition remarquable est vérifiée. Par exemple, très peu d'espèces vivent au delà de 50 mètres de profondeur. L'activation d'une règle exprimant une propriété suffisante permettra de déduire l'appartenance de l'individu à une espèce particulière.

Les attributs sont tous multivalués sans que la définition d'une facette de type ensemble (*set-of*) ou de type liste (*list-of*) soit précisée, tel que dans SHIRKA [REC 90] ou TROEPS. Tout attribut est considéré comme une application de son domaine vers l'ensemble des parties du domaine (le co-domaine) qui constitue l'ensemble des valeurs admissibles par les instances. Les valeurs admissibles sont organisées en un treillis de généralisation de valeurs, où l'extremum supérieur est l'ensemble des valeurs possibles qui correspond à l'*inconnu* et l'extremum inférieur, l'ensemble vide (absence de valeur). Les valeurs non renseignées, l'inconnu, les singletons, les intervalles de valeurs (pour le type numérique), les valeurs disjonctives et conjonctives et les ensembles peuvent être utilisés dans les descriptions [GRO 02].

2.4. Spécialisation et généralisation

Les composants peuvent également être spécialisés pour distinguer différentes sortes d'objets, du plus général au plus spécialisé. Le sens accordé à la spécialisation de composants est cependant différent de celui de la spécialisation des classes dans les modèles objets. En effet, trois règles internes au modèle CoDesc permettent de maintenir la cohérence de la spécialisation et de la généralisation dans CoDesc.

2.4.1. Règles de spécialisation

Ces règles relient le statut d'une propriété d'un concept au statut de la même propriété héritée dans les sous-concepts.

- R1. <nécessaire> → <nécessaire>
- R2. <contingente> → <contingente> | <nécessaire> | <absente>
- R3. <absente> → <absente>

Au sens de la relation de spécialisation, un composant hérite de toutes les caractéristiques nécessaires du composant père (R1). Les caractéristiques contingentes peuvent ne pas être héritées (à l'inverse des classes), rester contingentes ou devenir nécessaires (R2). La règle (R3) complète le système de règles.

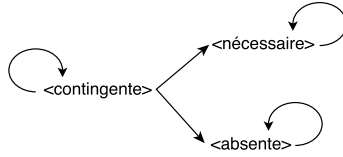


Figure 2. Trois états possibles des caractéristiques et de leurs transformations par spécialisation.

Interprétation des règles de spécialisation : (R1) Un concept possède obligatoirement les propriétés nécessaires du concept qui le subsume. En effet, si l'on dit que "tous les oiseaux ont des plumes" est une condition nécessaire du concept *Oiseau*, on ne peut définir de sous-concept d'oiseaux qui n'ont pas de plumes, cela reviendrait à dire que ce "drôle d'oiseau" n'est pas vraiment un oiseau (comme le grand pingouin). Dans ce cas, il est préférable de préciser que cette propriété est contingente : *la plupart des oiseaux ont des plumes*. La règle (R2) peut s'illustrer de la manière suivante : la propriété contingente *la plupart des oiseaux ont des plumes* (composant plumage absent possible) peut se spécialiser en "cette sorte d'oiseaux a nécessairement des plumes" (regroupe la plupart des oiseaux, présence du plumage nécessaire) et "cette sorte d'oiseau n'a pas de plume" (les pingouins, absence de plumage). Elle peut également rester contingente pour définir un sous-type d'oiseaux avec ou sans plumes. Enfin, une propriété absente est nécessairement absente des sous-concepts (R3).

2.4.2. Règles de généralisation

La figure 2 peut être interprétée pour la généralisation en inversant le sens des relations. Nous obtenons les règles suivantes qui relient le statut d'une propriété définie dans un concept fils à la propriété généralisée dans le concept père :

- R4. <nécessaire> → <contingente> | <nécessaire>
- R5. <contingente> → <contingente>
- R6. <absente> → <absente> | <contingente>

Interprétation des règles de généralisation : Un composant possède, par union généralisée, toutes les propriétés nécessaires du concept qui le subsume, c'est-à-dire qu'une propriété nécessaire dans un sous-concept peut devenir contingente dans un super-concept (R4). En effet, si deux sous-concepts C_1 et C_2 possèdent la même propriété *cette sorte d'oiseau vole*, nécessaire dans C_1 et contingente dans C_2 , qui s'interprète : *la plupart des oiseaux de C_2 volent* (certains d'entre eux ne volent pas). Cette propriété est également contingente dans le concept C_0 qui les subsume (R5), c'est-à-dire que *la plupart des oiseaux de C_0 volent* puisqu'en particulier certains oiseaux de C_2 ne volent pas. Par contre, si la propriété est nécessaire dans tout sous-concept, elle l'est aussi dans le super-concept. (R6) Une propriété absente ne peut pas être nécessaire dans le super-concept à cause de la règle 1. Elle est donc soit absente, soit contingente.

2.4.3. Construction d'une hiérarchie de concepts

La construction d'une hiérarchie de concepts cohérente par rapport à cet ensemble de règles peut paraître déroutante, car la plupart des modèles objets procèdent par ajout de propriétés dans les sous-classes ou par insertion de nouvelle classe dans la hiérarchie existante, ce qui n'entraîne aucune modification des classes plus générales.

A l'inverse, lors de la construction d'un modèle descriptif, les propriétés de celui-ci sont automatiquement propagées aux concepts plus généraux par application des règles 4, 5 et 6, ce qui est rendu possible par l'héritage simple et statique. Ce mécanisme permet de faciliter la construction d'une taxonomie de manière ascendante par une approche inductive. En effet, l'expert développe un ensemble de modèles spécifiques pour la description de groupe restreints d'individus (modèles des différents genres) et peut induire de nouveaux modèles par synthèse de description généralisante à partir d'individus ou de modèles [GRO 02], la cohérence de la taxonomie étant maintenue par l'activation des règles de cohérence. Il existe ainsi une relation de causalité forte entre la description des individus et la définition des modèles, ainsi qu'entre modèles spécifiques et modèles généraux. Cette relation est au coeur même de la démarche inductive des naturalistes : les concepts existent et sont définis sur la base d'un ensemble de spécimens de référence, les échantillons types mis en collections dans les muséums.

2.5. Les cas

Les cas sont des représentations descriptives des individus du monde réel, le résultat d'un travail de description d'un échantillon effectué par l'expert. Un cas est instance d'un concept, représenté par un modèle descriptif et hérite les caractéristiques définies par celui-ci, au moment de l'instanciation. Le travail de l'expert consiste donc à renseigner ces caractéristiques : les valeurs des attributs, l'absence de certains composants, leur multiplicité, etc. Une fois complété, le cas est ajouté à la base de cas attachée au concept. Celle-ci est ainsi considérée comme une représentation explicite de l'extension du concept.

La figure 3 offre un aperçu de la structure globale d'un cas modélisé dans *IKBS*, issu de la base de connaissances sur les coraux des Mascareignes [FAU 99]. Le concept dont il est instance est celui de la famille des *Pocilloporidae* (cf. fig. 1). Il illustre une description de l'échantillon n° **Reu 560** du Muséum National d'Histoire Naturelle, décrit par **Gérard Faure**, les caractères essentiels de la description sont donnés par le point de vue *description*. La valeur associée à l'attribut de classe *Taxon* (5) : *Pocillopora verrucosa*, caractérise le nom d'espèce de cet échantillon. Ce cas est donc une description d'un individu de l'espèce *verrucosa*, de genre *Pocillopora* de la famille des *Pocilloporidae*. Cet exemple illustre également l'étendue des types de valeurs possibles lors de la description d'un cas : (1) désigne une valeur inconnue, (2) un composant absent, (3) un intervalle de valeurs numériques, (4) un ensemble de valeurs symboliques observées simultanément exprimant une conjonction de variation.

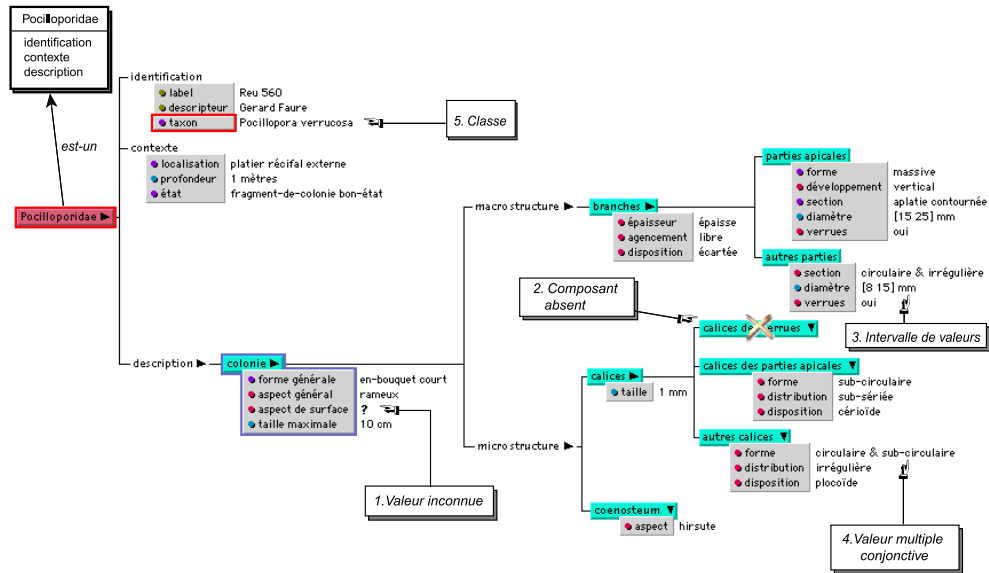


Figure 3. Exemple de cas du modèle CoDesc dans IKBS. La structure du cas est identique à celle du modèle descriptif dénotant le concept *Pocilloporidae* (fig. 1).

3. Aperçu de la plate-forme IKBS

IKBS⁵ est la plate-forme logicielle développée en Java, pour la mise en oeuvre du modèle CoDesc, dans le but de répondre à la nécessité, pour les biologistes, de disposer d'outils informatiques permettant l'acquisition, la gestion et le traitement des connaissances. IKBS peut interpréter des connaissances exprimées dans le langage définitionnel de CoDesc, propose des outils de visualisation des modèles et des cas, en garantissant la cohérence de la base et la synchronisation des différentes vues. Plusieurs méthodes d'analyse permettent d'exploiter les connaissances, telles que la classification, le raisonnement à partir de cas et l'identification de nouveaux spécimens par la méthode des arbres d'identification [GRO 02].

Pour un utilisateur de la plate-forme (biologiste ou informaticien), trois niveaux des connaissances sont considérés dans IKBS. D'une part, le **méta-niveau** ou le niveau sémantique, correspond à l'implantation du modèle de représentation CoDesc à l'aide de classes du langage Java : concepts, composants, valeurs, relations, etc. D'autre part, le niveau **Codesc** ou niveau connaissances : tout élément de la base de connaissances est un objet du langage, instance des classes définies au niveau méta. Enfin, le niveau **visuel** ou niveau intelligible : les connaissances sont manipulées par l'intermédiaire du langage de définition ou de leur représentation graphique. Les modifications apportées sont propagées par le système au niveau supérieur.

5. Iterative Knowledge Base System ou Système Itératif de gestion de Bases de Connaissances.

La modification d'un modèle descriptif entraîne automatiquement la modification correspondante aux niveaux des cas, avec éventuellement un réajustement des valeurs. *TKBS* permet de gérer l'évolution des connaissances, la migration d'un cas (changement de classe), la propagation des règles définies par l'utilisateur, activation des règles d'intégrité du modèle, l'ajout et la suppression d'un composant et la modification du domaine de valeurs d'un attribut (ajout d'un nouveau taxon par exemple).

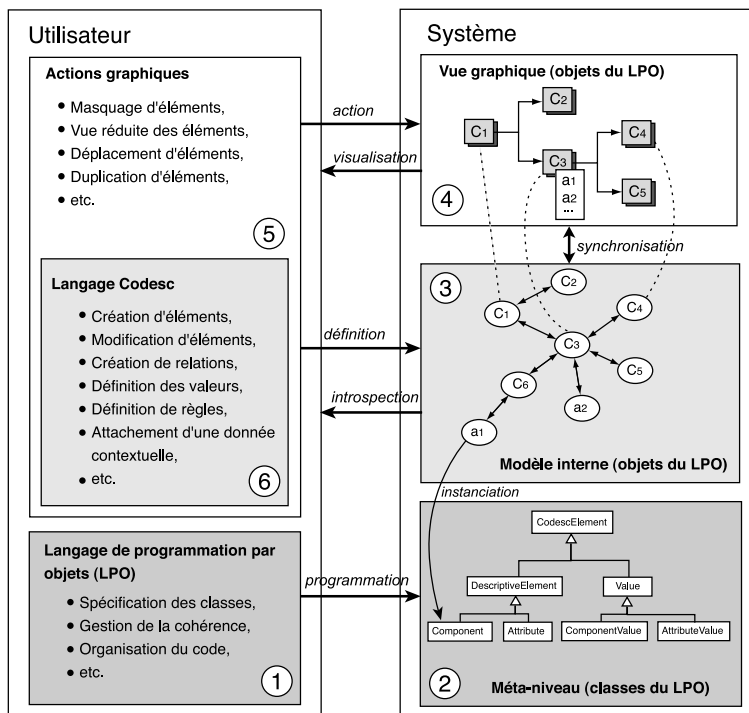


Figure 4. Les 3 niveaux de représentation du modèle *CoDesc* implanté dans *TKBS*, d'un point de vue utilisateur et système.

La figure 4 illustre ces 3 niveaux, du point de vue du programmeur et de l'utilisateur. Le modèle *CoDesc* est implanté par le concepteur (1) par un ensemble de hiérarchies de classes (2) Java. L'instanciation du modèle dans un domaine particulier (3) est une base de connaissances, qui peut être visualisée graphiquement (4) et manipulée par l'utilisateur à l'aide d'actions graphiques (5) ou bien par l'intermédiaire du langage *CoDesc* (6). Certaines actions au niveau graphique n'entraînent pas de modifications du modèle interne : afficher/masquer un sous arbre et ses attributs, agrandir la vue, modifier la présentation du modèle descriptif, etc. Différentes vues du modèle sont possibles simultanément et sont synchronisées par le système lors de la modification de l'une d'elle.

4. Conclusion

Dans le cadre de la recherche de modèles de représentation des connaissances et d'outils informatiques répondant aux besoins des biologistes, nous avons présenté un modèle conceptuel original de représentation des connaissances, ainsi qu'une plate-forme logicielle permettant de accompagner le travail des experts. Celle-ci offre un ensemble d'outils de haut niveau qui facilitent les différentes étapes de construction et de mise au point d'une base de connaissances, véritable ontologie du domaine. La méthodologie d'acquisition et de construction d'une taxonomie est fondée sur la méthode scientifique expérimentale des systématiciens, de nature inductive.

Depuis plusieurs années, la plate-forme *IKBS* a permis à un groupe d'experts français de construire une "base de connaissances sur les coraux des Mascareignes"⁶, à partir de près de 3000 spécimens récoltés dans l'Océan Indien et stockés dans la collection Gérard Faure [FAU 82]. Cette base de connaissances est une réponse à l'urgence qu'il y a de préserver le savoir-faire des experts liés à l'observation, la descriptions et la taxonomie de ces organismes marins. D'autres projets de conception de bases de connaissances sont en cours à l'aide de la plate-forme *IKBS*, pour l'identification de certaines plantes endémiques de l'île Maurice (projet Mauriflor@), sur d'autres familles d'organismes marins (les Hydraires), ainsi que pour l'aide au diagnostic médical en médecine nucléaire [CLE 01].

5. Bibliographie

- [BEA 93] BEACH J., PRAMANIK S., BEAMAN J., Hierarchic taxonomic databases , *Advances in computer methods for systematic biology : Artificial intelligence, databases, computer vision*, , 1993, p. 241-256, R. Fortuner (éd.).
- [BER 97] BERENDSOHN W. G., A taxonomic information model for botanical databases : the IOPI Model , *Taxon*, vol. 46, 1997, p. 283-309.
- [CLE 01] CLERMONT H. D., CONRUYT N., GROSSER D., ROBERT N., Etude préliminaire à l'utilisation de la plate-forme IKBS pour la création de bases de connaissances en Médecine Nucléaire : approche en cardiologie , *39ème Colloque de Médecine Nucléaire*, 2001.
- [CON 94] CONRUYT N., Amélioration de la robustesse des systèmes d'aide à la description, à la classification et à la détermination des objets biologiques , Thèse de doctorat, Université Paris IX-Dauphine, Mai 1994.
- [CON 97a] CONRUYT N., FAURE G., ANCEL G., RENARD J. L., GUILLAUME M., NAIM O., GRAVIER-BONNET N., A Knowledge Base for corals of the Mascarene Archipelago : Genus Pocillopora , *Proc 8th Int Coral Reef Symp*, 1997.
- [CON 97b] CONRUYT N., GROSSER D., Classification et identification d'espèces par discrimination à partir de connaissances structurées , *Actes des 5ème journées de classification et d'analyse de données (JOCLAD'97)*, 1997, p. 63-66.

6. Nous tenons à remercier les experts Gérard Faure (ISIM de Montpellier), Michel Pichon (EPHE de Perpignan) et Mireille Guillaume (MNHN de Paris) pour leur implication active dans ce projet, financé par la Région Réunion.

- [DAL 95] DALLWITZ M., PAIN T., ZURCHER E., User's guide to the DELTA system. A general system for coding taxonomic descriptions , csiro. div. entomol. édition, 1995.
- [FAU 82] FAURE G., Recherche sur le peuplements de Scléactiniaires des récifs coralliens de l'Archipel des Mascareignes , Thèse de doctorat d'Etat, Univ. Aix-Marseille II, 1982.
- [FAU 99] FAURE G., CONRUYT N., PICHON M., GUILLAUME M., GROSSER D., GEYNET Y., Development of a Knowledge Base for the Corals of the Mascarene Archipelago , *Int. Conf. on Scientific Aspects of Coral Reef Assessment*, ISRS/NCRI, 1999.
- [GRO 00] GROSSER D., DIATTA J., CONRUYT N., Improving Dissimilarity Functions with Domain Knowledge, applications with IKBS system , *PAKDD'2000, Int. conf. on Practical Applications of Knowledge Discovery in Database*, 2000.
- [GRO 02] GROSSER D., Construction itérative de bases de connaissances descriptives et classificatoires avec la plate-forme à objets IKBS : application à la systématique des coraux des Mascareignes , Thèse de doctorat, Université de la Réunion, Janvier 2002.
- [GRU 93] GRUBER T., A translation approach to portable ontology specifications , *Knowledge Acquisition*, vol. 5, n° 2, 1993, p. 199-220.
- [KIM 87] KIM W., BANERJEE J., CHOU H., GARZA J., WOELK D., Composite Object Support in an Object-Oriented Database System , *Proceedings of the 2nd OOPSLA*, ACM, Notices vol.22, n°12, 1987, p. 118-125.
- [LEB 96] LEBBE J., Quelques réflexions sur l'informatique appliquée à la systématique en France , *Biosystema 14, Informatique et Systématique*, , 1996, p. 5-10.
- [MAR 93] MARINO O., Raisonement classificatoire dans une représentation à objets multi-points de vue , Thèse de doctorat, Université Joseph Fourier, Octobre 1993.
- [MAS 89] MASINI G., NAPOLI A., COLENT D., LÉONARD D., TOMBRE K., *Les langages à objets*, InterEditions, 1989.
- [NAP 92] NAPOLI A., Représentation à objets et raisonnement par classification en intelligence artificielle , Thèse de doctorat d'état, Université Nancy 1, 1992.
- [PAG 00] PAGE M., GENSEL J., CAPPONI C., BRULEY C., GENOUD P., ZIÉBELIN D., Représentation de connaissances au moyen de classes et d'associations : le système AROM , HERMÈS, Ed., *Actes de LMO'00, langages et modèles à objets*, 2000, p. 91-106.
- [PAN 98] PANKHURST R., A historical review of identification with computers , *Information Technology Plant Pathology and Biodiversity*, , 1998, p. 289-303, Bridge P. eds.
- [PUL 00] PULLAN M. R., ATSON M. F., KENNEDY J. B., RAGUENAUD C., HYAM R., The Prometheus Taxonomic Model : a practical approach to representing multiple classifications , *Taxon*, vol. 49, 2000, p. 55-75.
- [REC 90] RECHENMANN F., FONTANILLE P., UVIETTA P., SHIRKA : système de gestion de bases de connaissances centrées-objets , Rapport de recherche, 1990, INRIA et Artémis.
- [ROU 01] ROUSSE G., Osis, an object model for systematics, TDWG, poster, <http://lis.snv.jussieu.fr/rousse/recherche/>, 2001.
- [SHE 95] SHERPA, TROPES 1.0 reference manual , Rapport de recherche, 1995, INRIA.
- [VAL 99] VALTCHEV P., Construction automatique de taxonomies pour l'aide à la représentation de connaissances par objets , Thèse de doctorat, Univ. Grenoble 1, 1999.
- [VER 00] VERON J., *Corals of the world*, Australian Institute of Marine Science, 2000.
- [VIG 00] VIGNES R., Informatique et Systématique : vers une ontologie de la Systématique , H.D.R., Université Paris-VI, Janvier 2000.