

Classification par voisinages successifs sur des descriptions morphologiques complexes

David Grosser*, Noël Conruyt*
Henri Ralambondrainy*

*LIM-IREMIA, Université de la Réunion
Parc Technologique Universitaire, Bâtiment 2
2, rue Joseph Wetzell - 97490 Sainte-Clotilde
grosser, conruyt, ralambon@univ-reunion.fr

Résumé. Afin de classer des descriptions morphologiques issues de bases de connaissances en biologie, nous proposons une méthode de fouille de données incrémentale, interactive et semi-dirigée. Cette méthode est fondée sur la construction itérative du voisinage de la description partielle de l'objet à classer. Nous proposons différents indices de similarité adaptés à la nature complexe des données considérées (multi-valuées, incomplètes et structurées), pour sélectionner les descriptions les plus proches. Les connaissances du domaine sont utilisées aux différentes étapes du processus de classification, notamment pour le choix de variables discriminantes. A partir de la base de connaissances sur les coraux des Mascareignes, une application montre l'intérêt de cette approche.

1 Introduction

Outre l'étude des liens de parenté entre espèces (phylogénie), la description de la morphologie de *specimens* biologiques à des fins d'identification et de classification est une part essentielle du travail des systématiciens (phénétique). L'automatisation de ce processus par des outils informatiques dans le but de construire des systèmes classificatoires pose d'intéressants problèmes de représentation et de traitement de connaissances. Ceci est particulièrement vrai en biologie marine et pour certains taxons comme les coraux, où la nature polymorphe des individus (les colonies coralliennes) rend difficile leur description par des représentations classiques de type attribut-valeur. Les relations de dépendance entre caractères, induites par une forte variabilité morphologique engendrent notamment des difficultés d'observation et de description.

Pour pallier à ces contraintes et permettre la construction de bases de connaissances sur la biodiversité des milieux marins, une méthodologie d'acquisition des connaissances descriptives adaptée aux besoins particuliers des biologistes, s'appuyant sur *les logiques descriptives en sciences de la vie* a été proposée par Conruyt (1994) et Renard et al. (1996). Cette méthode permet de définir une connaissance d'ordre ontologique sur la nature des taxons et de décrire de manière structurée les spécimens utilisés pour la conception de la Taxonomie des milieux considérés. Elle offre un cadre signifiant pour représenter des objets biologiques complexes,