

TABLE DES MATIERES

INTRODUCTION	1
I LE CHEMINEMENT CONCEPTUEL	11
1.1 Les systèmes experts à l'INRA.....	11
1.1.1 Bien définir la cible des systèmes experts.....	12
1.1.1.1 Les experts	12
1.1.1.2 Les techniciens ou conseillers agricoles.....	12
1.1.1.3 Les agriculteurs.....	13
1.1.1.4 Conclusion	13
1.1.2 Bien définir les objectifs et les moyens.....	13
1.1.2.1 Savoir observer.....	14
1.1.2.2 Savoir raisonner	14
1.1.2.3 Conclusion	15
1.2 Le projet INSTIL.....	15
1.2.1 Neddie.....	16
1.2.2 Main	16
1.2.3 Maggy.....	17
1.2.4 Conclusion.....	18
1.3 Des systèmes experts à l'apprentissage.....	19
1.3.1 Les systèmes experts	19
1.3.2 Acquérir les connaissances de l'expert	21
1.3.3 Adaptation à l'utilisateur.....	24
1.3.4 Comparaison des deux approches.....	26
1.3.5 Conclusion.....	29
1.4 KATE.....	29
1.4.1 Une bonne représentation des connaissances.....	29
1.4.2 Des mécanismes d'exploitation adaptés.....	31
1.5 L'aide à la classification au MNHN.....	36
1.5.1 Comparaison avec l'opération SEPV de l'INRA.....	36
1.5.2 Utilisation des langages de frames et de l'hypertexte.....	40
1.5.3 L'application SPONTAX.....	42
1.6 Conclusion.....	45
II QU'EST-CE QUE LA ROBUSTESSE ?	49
2.1 Aspects théoriques	50
2.1.1 La robustesse statistique	50
2.1.2 Le formalisme mathématique de description.....	51
2.1.3 Combiner du numérique et du symbolique	51
2.2 Aspects pratiques.....	52
2.2.1 Les facteurs qualitatifs.....	52
2.2.1.1 Fiabilité	52
2.2.1.2 Compréhension.....	52
2.2.1.3 Précision.....	53
2.2.1.4 Exhaustivité	53
2.2.1.5 Cohérence	54
2.2.1.6 Redondance	54
2.2.1.7 Mise à jour.....	55
2.2.1.8 Ergonomie.....	55
2.2.1.9 Tolérance aux bruits.....	57
2.2.1.10 Adaptation aux besoins exprimés	58

2.2.2 Les critères globaux d'appréciation.....	59
2.2.2.1 Applicabilité à des domaines réels.....	59
2.2.2.2 Un langage de représentation puissant.....	60
2.2.2.3 Facilité de mise en oeuvre par les utilisateurs.....	60
2.2.2.4 Incrémentalité.....	60
2.3 Discussion	61
2.3.1 L'informaticien.....	61
2.3.2 Le statisticien.....	61
2.3.3 Le biologiste.....	63
2.3.4 Le béotien.....	64
2.3.5 L'expert.....	64
2.4 Notre méthode d'acquisition des connaissances.....	66
2.4.1 Différents types de connaissances à acquérir.....	66
2.4.1.1 Connaissances de base.....	66
2.4.1.2 Connaissances de faits observés.....	66
2.4.1.3 Connaissances produites	67
2.4.2 Deux types de traitements des exemples pour la classification et la détermination.....	68
2.4.3 La phase de validation des connaissances apprises.....	70

III TERMINOLOGIE ET CONCEPTS MIS EN ŒUVRE

3.1 Extension et compréhension.....	73
3.1.1 L'extension.....	73
3.1.1.1 Point de vue du philosophe et du mathématicien ..	73
3.1.1.2 Point du vue des biologistes et des naturalistes....	74
3.1.2 La compréhension.....	75
3.2 Classe et concepts.....	78
3.2.1 La classe.....	78
3.2.1.1 Point de vue des mathématiciens	79
3.2.1.2 Point de vue des systématiciens	79
3.2.2 Les concepts.....	83
3.2.2.1 Du point de vue naturaliste.....	84
3.2.2.2 Du point de vue mathématique.....	87
3.3 Classement et classification.....	89
3.3.1 Classer et le classement.....	89
3.3.2 Classifier et la classification.....	89
3.4 Détermination et identification.....	93
3.4.1 Détermination par comparaison directe.....	94
3.4.2 Détermination par comparaison avec des descriptions	94
3.5 Apprentissage et raisonnement	95
3.6 Individus, instances et objets.....	97
3.7 Synthèse des concepts utilisés dans cette thèse.....	99

IV LE ROLE CENTRAL DES DESCRIPTIONS EN SCIENCES

NATURELLES	101
4.1 Objectifs de la description.....	101
4.2 Qualités de la description.....	102
4.3 Qualités des descriptions.....	103
4.4 La représentation des données descriptives.....	103
4.4.1 Structuration naturelle.....	104
4.4.2 La logique de composition.....	105
4.4.3 La logique des points de vue	106
4.4.4 La logique de spécialisation.....	106
4.4.5 La logique de particularisation.....	107
4.4.6 La logique d' itération.....	107
4.4.7 Les conditions contextuelles.....	109

4.5	Représentation structurée selon un modèle descriptif.....	110
4.6	Illustration sur les données des éponges marines.....	112
4.6.1	Acquérir l'observable.....	112
4.6.2	Décomposition de l'entité globale en objets.....	113
4.6.3	Recherche des attributs de chaque objet.....	114
4.6.4	Recherche des valeurs de chaque attribut	114
4.6.5	Les propriétés d'un objet observable.....	115
4.6.5.1	Statuts.....	115
4.6.5.2	Attributs.....	117
4.6.5.3	Autres propriétés.....	122
4.6.6	Niveau de précision d'un objet.....	123
4.6.6.1	Spécialisation.....	123
4.6.6.2	Particularisation.....	124
4.6.7	Possibilité de multi-instanciation de chaque objet.....	125
4.6.8	Définition des dépendances inter / intra objets.....	126
4.6.9	Niveau de généralité d'un objet	127
4.6.9.1	Représentant unique.....	128
4.6.9.2	Représentant multiple	129
4.6.10	L'imprécision d'observation.....	129
4.6.10.1	Au niveau de l'observable.....	130
4.6.10.2	Au niveau de l'observé.....	131

V FORMALISATION DE LA NOTION D'OBJET EN BIOLOGIE...135

5.1	Les assertions composites.....	135
5.1.1	Rappel sur les assertions (symboliques)	135
5.1.2	Proposition : les objets assertions.....	137
5.1.3	Définition des assertions composites	137
5.1.4	Exemples.....	139
5.2	Les hordes composites.....	141
5.2.1	Rappel sur les hordes (symboliques).....	141
5.2.2	proposition : les objets hordes.....	141
5.2.3	Définition des hordes composites.....	142
5.2.4	Exemples.....	143
5.3	Les objets de synthèse.....	144
5.3.1	Rappel sur les objets de synthèse (symboliques).....	144
5.3.2	Proposition : les objets de synthèse.....	145
5.3.3	Exemple.....	145
5.4	Les objets munis de méthodes et de propriétés.....	146
5.4.1	Cas des variables :	146
5.4.2	Cas des objets :	147
5.5	Les objets classifiés	148
5.6	Les exemples d'apprentissage.....	149
5.7	Conclusion	151

VI FORMALISATION INFORMATIQUE DES DESCRIPTIONS :

HYPERQUEST	153
6.1	Pourquoi HyperQuest ?.....	153
6.2	Particularités d'HyperQuest.....	154
6.3	Le générateur de modèle.....	155
6.3.1	Conception d'un modèle descriptif	156
6.3.2	Règles d'élaboration du modèle descriptif.....	158
6.3.3	Dépendances.....	160
6.3.4	Spécialisations.....	162
6.3.5	Itérations.....	163
6.3.6	Edition d'un objet.....	167
6.3.7	Edition d'un attribut	169

6.3.8 Edition des valeurs classifiées.....	171
6.4 Le générateur automatique de questionnaire	172
6.4.1 Principe de construction d'un questionnaire.....	172
6.4.2 Entités conceptuelles et hypertextes	174
6.5 L'éditeur de cas : le questionnaire.....	176
Caractéristiques	176
6.5.1 Représentation de l'individu.....	176
6.5.2 Représentation des objets de l'individu.....	177
6.5.2.1 Un objet = une carte.....	177
6.5.2.2 Représentation des statuts possibles d'un objet..	178
6.5.3 Représentation des attributs d'un objet.....	179
6.5.3.1 Attributs qualitatifs.....	180
6.5.3.2 Attributs numériques.....	181
6.5.3.3 Attributs commentaires	181
6.5.3.4 Attributs classifiés.....	182
6.5.4 Représentation des spécialisations d'un objet.....	183
6.5.5 Représentation des multi-instanciations d'un objet.....	186
6.5.6 Représentation des règles entre objets.....	188
6.5.7 Personnalisation des objets.....	189
Fonctionnalités.....	189
6.5.8 Le menu "Cas"	191
6.5.8.1 Le choix "Nouveau"	192
6.5.8.2 Le choix "Dernier"	194
6.5.8.3 Le choix "Ancien ..."	194
6.5.8.4 Le choix "Enregistrer".....	195
6.5.8.5 Le choix "Consulter l'expert".....	196
6.5.8.6 Le choix "Voir la liste".....	197
6.5.8.7 Le choix "Convertir en Casuel".....	197
6.5.9 Le menu "Personnaliser"	198
6.5.9.1 Le choix "Préférences..."	198
6.5.9.2 Le choix "Scanner une image".....	198
6.5.9.3 Le choix "Tout importer".....	198
6.5.9.4 Le choix "Importer une image".....	199
6.5.9.5 Le choix "Déplacer l'image"	199
6.5.9.6 Le choix "Associer un bouton".....	199
Les boutons rectangulaires.....	200
Les boutons polygonaux.....	201
6.5.9.7 Le choix "Changer le titre".....	201
6.5.9.8 Le choix "Déplacer les boutons".....	201
6.5.9.9 Le choix "Eliminer un bouton"	201
6.5.9.10 Le choix "Déplacer les champs".....	201
6.5.9.11 Le choix "Sauver la carte"	202
6.5.10 Les boutons.....	202
6.5.10.1 Le bouton "Voir une image".....	202
6.5.10.2 Le bouton "Naviguer".....	204
6.5.10.3 Le bouton "Coursier".....	204
6.6 Conclusion.....	205

VII LE TRAITEMENT DES DESCRIPTIONS BIOLOGIQUES : KATE ET CASEWORK.....

7.1 Formalisation de l'approche inductive.....	207
7.1.1 Rappel des Notations.....	208
7.1.2 Principe de la classification par arbre de décision.....	209
7.1.3 Algorithme.....	210
7.1.4 Description des fonctions principales de l'algorithme.....	212
7.1.4.1 OrdonnerCritères (E, Yd)	212

7.1.4.2 Meilleure_division (E, s)	214
7.1.4.3 Critèred'Arrêt (E).....	215
7.1.4.4 ConstruireFeuille (E).....	217
7.1.4.5 CalculerSeuil (A,E).....	217
7.2 De l'induction au raisonnement par cas.....	219
7.2.1 Formalisation de la procédure de détermination.....	220
7.2.2 Limites de l'approche inductive.....	221
7.2.2.1 Apprentissage automatique = perte d'information.....	221
7.2.2.2 Gestion de l'inconnu en phase de consultation....	221
7.2.2.3 Rigidité de la consultation interactive.....	222
7.3 Le raisonnement par cas.....	223
7.3.1 Généralités.....	223
7.3.2 Notre procédure de raisonnement par cas : CaseWork.....	225
7.3.3 Formalisation.....	227
7.3.4 Comparaison des deux approches	228
CONCLUSION	231
8.1 Résultats.....	231
8.2 Limites actuelles.....	234
8.3 Perspectives.....	236
REFERENCES.....	241
LISTE DES FIGURES	251
ANNEXES	253
I L'induction en mathématiques*.....	255
II L'expert, son "raisonnement"*.....	261
III Répartition des descriptions de <i>Hyalonema</i> par Sous-Genre.....	263
IV Architecture d'HyperQuest.....	265
IV-1 Les piles d'HyperQuest.....	265
IV-1.1 Le dossier Source	266
IV-1.2 Le dossier de l'application du domaine.....	268
V Syntaxe BNF* pour le LCRC.....	271
V-1 Notes sur les grammaires BNF	271
V-2 Les définitions LCRC du modèle descriptif	272
V-2.1 Syntaxe des objets.....	272
V-2.2 Syntaxe des attributs.....	272
V-2.3 Syntaxe des valeurs.....	273
V-3 Les descriptions LCRC du questionnaire.....	273
V-3.1 Syntaxe des cas.....	273
VI Induction and Reasoning from Cases.....	275

