

N° d'ordre :

# THÈSE

présentée à

**L'UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS**

pour obtenir le titre de

**DOCTEUR EN SCIENCES**

spécialité

**INFORMATIQUE**

par

**Michel JACZYNSKI**

Sujet de la thèse :

**Modèle et plate-forme à objets pour l'indexation des cas par situations comportementales : application à l'assistance à la navigation sur le Web**

Soutenue le 3 décembre 1998 devant le jury composé de :

<b>M. Pierre Bernhard</b>	Président
<b>M. Jean-François Perrot</b>	Rapporteur
<b>M. Joël Quinqueton</b>	Rapporteur
<b>M. Alain Mille</b>	Examineur
<b>M. Michel Plu</b>	Invité
<b>M. Michel Rueher</b>	Examineur
<b>Mme Brigitte Trousse</b>	Directeur



## **Remerciements**

---

C'est avec une certaine émotion et beaucoup de sincérité que je voudrais remercier toutes les personnes ayant soutenu et apprécié mon travail.

En premier lieu, je remercie M. Pierre Bernhard de me faire l'honneur de présider le jury. Je le remercie également pour m'avoir accueilli dans l'unité de recherche de l'INRIA Sophia Antipolis. Je remercie M. Jean-François Perrot pour avoir accepté d'être rapporteur de ma thèse et pour m'avoir donné la possibilité de présenter mes travaux dans son équipe. Je remercie M. Joël Quinqueton, également rapporteur de cette thèse, pour ses commentaires qui ont permis d'améliorer mon rapport.

Je tiens tout particulièrement à remercier Mme Brigitte Trousse qui a dirigé cette thèse, pour sa disponibilité, son soutien et ses conseils qui ont été nombreux et importants. Mme Brigitte Trousse a beaucoup œuvré pour la mise en valeur de mon travail, et j'ai apprécié l'ambiance d'équipe qui s'est formée. J'ai beaucoup appris à ses côtés et je lui adresse toute ma reconnaissance.

Je remercie également M. Alain Mille pour sa participation au jury et pour son activité au sein du groupe RàPC français avec lequel j'ai passé des moments agréables dans différents lieux d'Europe. Je remercie M. Michel Plu du CNET pour son intérêt dans les applications de Broadway et pour sa participation au jury. Je remercie enfin M. Michel Rueher pour tous ses conseils au cours de ma thèse, sans oublier que j'ai eu le privilège d'être son élève.

Je remercie Rushed Kanawati, Marc Golinelli, Mourad Elhaddad et Youness Taheri pour leurs commentaires et leur participation à divers projets de l'équipe AID. Je remercie M. Brun de l'INRA Sophia Antipolis pour le temps qu'il m'a consacré. Je remercie également Mme Bonnardel et ses étudiants de l'université de Provence pour leur étude de Broadway.

La réalisation de ce travail s'appuie également sur un environnement qui est essentiel. A ce titre, je voudrais remercier l'INRIA et son personnel, notamment l'équipe SEMIR et l'équipe de la documentation. Je remercie également l'INRIA et la Région PACA pour leur participation financière dans cette étude.

Mes plus profonds remerciements vont à mes parents. Tout au long de mon cursus, ils m'ont toujours soutenu, encouragé et aidé. Ils ont su me donner toutes les chances pour réussir. Qu'ils trouvent, dans la réalisation de ce travail, l'aboutissement de leurs efforts ainsi que l'expression de ma plus affectueuse gratitude.

Je remercie ma famille et en particulier mon frère pour m'avoir fait partager leur joie de vivre et m'avoir ainsi soutenu dans mes efforts. Plus personnellement, je remercie ma fiancée, Anne-Marguerite de Boussineau, pour son aide, son écoute et surtout son amour qui m'a été essentiel durant ces années. Je remercie tous mes plus proches amis, en ayant une pensée pour Rémy à qui je souhaite santé et bonheur.

Je voudrais également remercier tous mes collègues et amis du projet Secoia. Je garderai un bon souvenir des discussions animées au cours des repas ensoleillés près de la piscine... Plus particulièrement, je remercie Stéphane Demphlous à qui je souhaite une très grande réussite personnelle. J'ai également une pensée amicale pour Wided Lejouad, dont je garderai l'exemple ainsi que le souvenir de nos discussions.

Je remercie enfin toutes les personnes intéressées par mon travail, en espérant qu'elles puissent trouver dans mon rapport des explications utiles pour leurs propres travaux.

*à ma mère et à mon père*





# ***Table des matières***

- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
-



<b>Remerciements</b>	<b>3</b>
<b>Table des matières</b>	<b>7</b>
<b>Introduction</b>	<b>15</b>
<b>1 Problématique</b>	<b>17</b>
<b>2 Contributions</b>	<b>18</b>
2.1 Présentation des contributions	19
2.2 Evaluation expérimentale	20
<b>3 Plans de lecture</b>	<b>20</b>
<b>Chapitre I : le RàPC pour une classe de problèmes</b>	<b>23</b>
<b>1 Motivations</b>	<b>25</b>
1.1 Régulation de la nutrition des plantes	25
1.1.1 Problème de la nutrition	25
1.1.2 Vers la réutilisation d'expériences de régulation	26
1.2 Aide à la navigation sur le World-Wide-Web	27
1.2.1 Présentation du Web	28
1.2.2 Assistance à la navigation sur le Web	30
<b>2 Présentation du raisonnement à partir de cas</b>	<b>31</b>
2.1 Principes du raisonnement	32
2.1.1 Représentation des cas et des indices	32
2.1.2 Cycle du raisonnement	32
2.1.3 Importance de l'indexation pour le raisonnement et terminologie	34
2.2 Utilisation du raisonnement à partir de cas	35
2.2.1 Avantages, inconvénients et conditions d'application	35
2.2.2 Un exemple : le système CLAVIER	37
<b>3 Application du RàPC à une classe de problèmes</b>	<b>38</b>
3.1 Classe des problèmes nécessitant la prise en compte d'historiques	38
3.2 Intérêt et spécificité du RàPC	39
3.2.1 Justification et intérêt du raisonnement à partir de cas	39
3.2.2 Gestion d'indices sous forme d'historiques	39
3.3 Démarche générale de l'étude	40
3.3.1 Nécessité d'un modèle d'indexation générique	40
3.3.2 Besoin d'un outil logiciel ouvert pour la gestion du modèle et du RàPC	41
<b>Chapitre II : modèle d'indexation par situations comportementales</b>	<b>43</b>
<b>1 Approches existantes d'indexation par historiques</b>	<b>45</b>
1.1 Approches basées sur le raisonnement à partir de cas	45
1.1.1 Présentation des systèmes existants	45
1.1.1.1 Prévision de l'évolution des feux de forêts : le système REBECAS	46
1.1.1.2 Commande de robots mobiles : le système SINS	47
1.1.1.3 Autres systèmes et approches de RàPC	49
1.1.2 Limites des systèmes existants	52
1.1.2.1 Absence d'approche générale pour l'indexation	52
1.1.2.2 Gestion des cas insuffisante pour la découverte de connaissances	52
1.1.2.3 Absence d'indices basés sur des observations hybrides	55
1.2 Autres approches	55
1.2.1 Reconnaissance d'historiques	56
1.2.1.1 Approches de traitement du signal	56
1.2.1.2 Approches syntaxiques	57
1.2.1.3 Approches basées sur des scénarios temporels	59

1.2.2 Bases de données temporelles et gestion d'historiques	62
1.2.2.1 Systèmes de gestion de bases de chroniques	62
1.2.2.2 Le système TEMPOS	63
1.2.3 Limites de la reconnaissance d'historiques et des bases de données temporelles	64
<b>2 Proposition du « modèle d'indexation par situations comportementales »</b>	<b>66</b>
2.1 Modèle de représentation des indices et des cas	66
2.1.1 Hypothèses	66
2.1.2 Présentation générale	67
2.1.3 Description détaillée	68
2.1.3.1 Granularité des cas	68
2.1.3.2 Représentation des données d'observation	69
2.1.3.3 Représentation des situations comportementales	72
2.1.3.4 Représentation des cas	76
2.2 Guide d'utilisation de notre modèle de représentation	76
2.2.1 Principes de recherche	77
2.2.1.1 Types de filtrage pour la recherche	77
2.2.1.2 Exemple d'une stratégie séquentielle de recherche	78
2.2.2 Principes de réutilisation	79
2.2.3 Principes d'apprentissage	79
2.3 Un exemple : recherche des cas pour la nutrition des plantes	80
2.3.1 Identification des variables d'observation	80
2.3.2 Détermination de la sémantique d'un enregistrement et de son contexte	81
2.3.3 Définition de la représentation des cas	81
2.3.3.1 Cas cible	82
2.3.3.2 Cas potentiels	82
2.3.3.3 Cas concrets	82
2.3.4 Conception des phases de raisonnement	83
<b>3 Conclusion</b>	<b>84</b>
<b>Chapitre III : plate-forme à objets pour la réalisation de systèmes de RàPC</b>	<b>87</b>
<b>1 Outils logiciels existants pour le RàPC</b>	<b>89</b>
1.1 Présentation des outils	89
1.1.1 REMIND	90
1.1.2 REPRO	93
1.1.3 KATE	94
1.1.4 FABEL	94
1.1.5 Autres outils	96
1.2 Limites des outils existants	99
1.2.1 Difficulté de modifier les composants proposés	99
1.2.2 Difficulté d'intégrer de nouveaux composants	100
<b>2 Proposition d'une plate-forme à objets pour le RàPC</b>	<b>100</b>
2.1 Principes d'une plate-forme à objets	100
2.1.1 Définition et propriétés d'une plate-forme	100
2.1.2 Conception d'une plate-forme	101
2.1.3 Rôles des patrons de conception dans une plate-forme à objets	103
2.1.4 Utilisation d'une plate-forme	104
2.2 Choix d'une plate-forme comme outil réutilisable et ouvert	105
2.3 Analyse des points d'ouverture de notre plate-forme à objets	106
2.3.1 Points d'ouverture du RàPC	107
2.3.1.1 Gestion du raisonnement	107
2.3.1.2 Représentation des cas	108
2.3.1.3 Organisation de la mémoire	109
2.3.2 Points d'ouverture pour l'indexation par situations comportementales	110
2.3.2.1 Gestion du raisonnement	110
2.3.2.2 Représentation des cas	110
2.3.2.3 Organisation de la mémoire	112
<b>3 Conclusion</b>	<b>112</b>

<b>Chapitre IV : patrons de conception et modèles à objets ouverts pour le RàPC</b>	<b>115</b>
<b>1 Modèles à objets ouverts pour le RàPC</b>	<b>117</b>
1.1 Gestion du raisonnement	117
1.1.1 Patrons de conception et représentation du cycle du raisonnement	117
1.1.1.1 Contrôle du raisonnement	118
1.1.1.2 Cohérence de l'assemblage des phases	120
1.1.2 Schémas de base pour les phases du raisonnement	123
1.1.2.1 Schéma de configuration pour la recherche	124
1.1.2.2 Deux schémas pour la réutilisation	125
1.1.2.3 Schéma de décision de l'ajout d'un cas pour l'apprentissage	127
1.2 Représentation des cas	128
1.2.1 Cas et situation	128
1.2.2 Indices structurés et poids d'importance	129
1.3 Organisation de la mémoire	130
1.3.1 Indépendance des index et du stockage des cas	131
1.3.2 Bases de cas structurées : organisation logique et physique des cas	131
1.3.3 Index et stratégie de recherche	133
1.3.3.1 Index et filtrage des indices	133
1.3.3.2 Hiérarchie extensible d'index	133
1.3.3.3 Communication entre les index et cohérence de l'assemblage	136
1.3.4 Similarités structurées, modifiables et extensibles	138
1.3.4.1 Domaines de valeurs des similarités	138
1.3.4.2 Similarités structurées et modifiables	139
1.3.4.3 Hiérarchies extensibles de fonctions élémentaires de similarité et d'agrégation	140
<b>2 Modèles à objets ouverts pour l'indexation par situations comportementales</b>	<b>142</b>
2.1 Représentation des cas	143
2.1.1 Représentation des données d'observation	143
2.1.1.1 Chroniques	143
2.1.1.2 Enregistrements et curseurs	144
2.1.2 Représentation des cas et des situations comportementales	145
2.1.3 Représentation des patrons de cas potentiels	147
2.2 Organisation de la mémoire	148
<b>3 Réalisation et utilisation de CBR*Tools</b>	<b>149</b>
3.1 Réalisation de la plate-forme	149
3.1.1 Structuration en sous-systèmes	150
3.1.2 Implantation en Java	150
3.2 Utilisation de CBR*Tools	152
3.2.1 Environnement d'utilisation	152
3.2.2 Estimation de la complexité d'utilisation de CBR*Tools	153
3.2.2.1 Effort minimum	153
3.2.2.2 Dépendances systématiques entre les points d'ouverture	154
<b>4 Conclusion</b>	<b>156</b>
<b>Chapitre V : application pour l'assistance à la navigation sur le Web</b>	<b>159</b>
<b>1 Présentation de Broadway</b>	<b>161</b>
1.1 Contraintes de l'application	161
1.2 Interface utilisateur et architecture de Broadway	162
1.2.1 Interface utilisateur	162
1.2.1.1 Barre d'outils	162
1.2.1.2 Contrôleur	163
1.2.2 Architecture logicielle	164
1.2.2.1 Architecture globale	164
1.2.2.2 Architecture détaillée de Broadway	166
1.2.2.3 Broadway*Tools : des composants ouverts et réutilisables pour Broadway	167
1.3 Serveur de recommandations	169
<b>2 Evaluation de notre modèle d'indexation et de CBR*Tools</b>	<b>170</b>
2.1 Application du modèle d'indexation par situations comportementales	170

2.1.1	Identification des variables d'observation	170
2.1.1.1	Adresse d'une page Web	170
2.1.1.2	Contenu d'une page Web	172
2.1.1.3	Evaluation explicite des pages	174
2.1.1.4	Evaluation implicite des pages : le ratio d'affichage	174
2.1.2	Détermination de la sémantique d'un enregistrement et de son contexte	174
2.1.3	Définition de la représentation des cas	176
2.1.3.1	Cas cible	176
2.1.3.2	Cas potentiels et cas concrets	176
2.1.4	Conception des phases de raisonnement	178
2.1.4.1	Phase de recherche	178
2.1.4.2	Phase de réutilisation	181
2.1.4.3	Phase de révision	182
2.1.4.4	Phase d'apprentissage et maintenance de la base de cas	183
2.2	Utilisation de la plate-forme CBR*Tools	184
2.2.1	Gestion du raisonnement	184
2.2.2	Représentation des cas	186
2.2.3	Organisation de la mémoire	188
2.3	Bilan	190
2.3.1	Modèle d'indexation par situations comportementales	190
2.3.1.1	Généricité du modèle de représentation	190
2.3.1.2	Mise en évidence de cas : vers la découverte des connaissances	191
2.3.2	Plate-forme à objets CBR*Tools	191
2.3.2.1	Evaluation du degré d'expertise nécessaire de CBR*Tools pour Broadway	192
2.3.2.2	Evaluation de l'aide apportée par CBR*Tools pour la réalisation de Broadway	192
<b>3</b>	<b>Proposition d'un nouveau type d'assistants à la navigation</b>	<b>193</b>
3.1	Evaluation expérimentale de Broadway	193
3.1.1	Objectifs et procédure d'expérimentation	194
3.1.2	Résultats de l'expérimentation	194
3.1.3	Configuration de Broadway utilisée	195
3.2	Comparaison avec les systèmes existants d'aide à la navigation	197
3.2.1	Situation de recommandation	197
3.2.2	Apprentissage	200
3.2.3	Architecture : indépendance des navigateurs, assistance multi-sites et ouverture	200
<b>4</b>	<b>Conclusion</b>	<b>201</b>
<b>Conclusion</b>		<b>203</b>
<b>1</b>	<b>Bilan</b>	<b>205</b>
1.1	Modèle d'indexation par situations comportementales	205
1.2	Plate-forme à objets pour le RàPC	206
1.3	Broadway : un nouveau type d'assistants à la navigation sur le Web	206
<b>2</b>	<b>Perspectives</b>	<b>207</b>
2.1	Vers l'extension et la généralisation de notre modèle d'indexation	207
2.2	Vers des méthodes d'utilisation d'objets-métier pour le RàPC	208
2.3	Vers des assistants basés sur les comportements	209
<b>Références bibliographiques</b>		<b>211</b>
<b>1</b>	<b>Liste des références</b>	<b>213</b>
<b>2</b>	<b>Index des références</b>	<b>224</b>
<b>Table des illustrations</b>		<b>227</b>
<b>Annexe A : éléments de notation UML</b>		<b>233</b>
<b>1</b>	<b>Diagramme de cas d'utilisation</b>	<b>235</b>
<b>2</b>	<b>Diagramme de classes</b>	<b>237</b>
<b>3</b>	<b>Diagramme d'objets</b>	<b>238</b>

---

<b>4 Diagramme de séquence</b>	<b>239</b>
<b>Annexe B : patrons de conception et CBR*tools</b>	<b>241</b>
<b>Annexe C : exemple d'utilisation de CBR*Tools</b>	<b>255</b>
<b>1 Présentation de l'application</b>	<b>257</b>
<b>2 Construction du système de raisonnement</b>	<b>258</b>
2.1 Représentation des cas	258
2.2 Organisation de la mémoire	260
2.2.1 Base de cas	261
2.2.2 Mesure de similarité	261
2.2.3 Base d'index	262
2.3 Gestion du raisonnement	264
2.3.1 Recherche	265
2.3.2 Réutilisation	265
2.3.3 Apprentissage	265
2.3.4 Fabrique des objets	266
<b>3 Exemples d'exécution</b>	<b>266</b>
3.1 Premier raisonnement	266
3.2 Deuxième raisonnement	267
<b>4 Code source complet de l'exemple</b>	<b>268</b>
4.1 Classe CarCase	268
4.2 Classe CarCaseSituation	268
4.3 Classe CarFileSimpleCaseBase	270
4.4 Classe CarIndexBase	271
4.5 Classe CarReasonerFactory	273
4.6 Classe CarRetain	274
4.7 Classe CarRetrieve	274
4.8 Classe CarReuse	274
4.9 Classe CarsApp	275
4.10 Classe CarSimilarity	277
<b>Annexe D : paramètres et algorithmes de Broadway</b>	<b>279</b>
<b>1 Paramètres de la modélisation</b>	<b>281</b>
<b>2 Algorithmes</b>	<b>284</b>
2.1 Instanciation du patron de cas potentiels	284
2.2 Filtrage sur les comportements élémentaires	285





## ***Introduction***

- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
-



# 1 Problématique

Notre travail a pour but l'étude du *raisonnement à partir de cas* (paradigme de résolution de problèmes par réutilisation d'expériences passées) pour une classe de problèmes complexes, ainsi que la conception d'outils logiciels pour faciliter la réalisation de systèmes intégrant ce type de raisonnement. Plus précisément, notre étude a été motivée par deux applications concrètes : tout d'abord la régulation de la nutrition des plantes, puis l'assistance à la navigation sur le World-Wide-Web.

La première application s'inscrit dans l'objectif général de réguler la nutrition minérale et hydrique des plantes, de manière à maîtriser les dérives minérales dans le substrat de culture et à assurer une meilleure adaptation de l'offre aux besoins des plantes. Le problème de régulation considéré est issu des travaux menés à l'INRA Sophia-Antipolis (Brun *et al.*, 1993 ; Brun & Settenbrino, 1993) : il s'agit de maintenir autour d'une consigne la concentration en sels minéraux de la solution de drainage qui est récupérée après l'arrosage des plantes. La commande, c'est-à-dire le paramètre contrôlé, est la concentration de la solution d'irrigation. Cette régulation est influencée par des perturbations dont certaines sont effectivement mesurées (température, rayonnement solaire, humidité). Bien que des approches de régulation aient été mises en place, ce problème reste ouvert notamment à cause de l'inertie du processus, du nombre important de variables pouvant entrer en jeu et de la nécessaire prise en compte des prévisions météorologiques. Une approche basée sur la réutilisation d'expériences passées doit permettre de reproduire les expériences de commande réussies, tout en évitant les échecs passés. De plus, cette approche doit permettre l'identification de comportements plus généraux, pour ainsi venir enrichir les connaissances du domaine.

Dans la deuxième application, nous désirons assister un groupe d'utilisateurs durant leurs navigations sur le World-Wide-Web (Web). Le Web est un vaste hypermédia qui est en expansion croissante, et offre une énorme quantité de données. Deux principales méthodes permettent de retrouver des documents pertinents dans cet espace virtuel : l'interrogation et la navigation. L'interrogation suppose l'utilisation d'un moteur de recherche et l'expression d'une requête. Toutefois, l'expression d'une requête n'est pas toujours facile si l'objectif de recherche est mal défini, et les réponses sont généralement accompagnées de beaucoup de bruit. La navigation est alors nécessaire et l'utilisateur peut parcourir à son gré les pages Web en suivant les liens proposés. Cependant, la complexité et l'étendue du Web peuvent entraîner la désorientation de l'utilisateur lors d'une navigation. La navigation est donc importante dans une tâche de recherche d'information sur le Web et nécessite une assistance. Notre approche vise alors une recherche d'information coopérative (de type indirect), où les expériences passées d'un groupe de personnes sont mises à profit. Plus précisément, nous désirons suivre un ensemble d'utilisateurs durant leurs navigations pour déduire leur intérêt potentiel à partir des comportements observés, et réutiliser des expériences passées similaires au sein du groupe pour proposer des recommandations.

Ces deux types d'applications, bien qu'issues de domaines différents, partagent des caractéristiques communes, ce qui nous a amenés à considérer la classe de problèmes définie par les critères suivants :

1. La prise en compte *d'historiques* est choisie ou nécessaire pour formuler le problème.
2. Un *processus dynamique* est observé.
3. Un ensemble *important* de données brutes est collecté.
4. Les données brutes peuvent être obtenues par *échantillonnage* et/ou lors *d'événements*.
5. Les connaissances du domaine sont *réduites* et une meilleure compréhension est recherchée.

Notre travail consiste alors à faciliter la conception et la réalisation de systèmes de raisonnement à partir de cas pour cette classe de problèmes. De manière générale, le raisonnement à partir de cas (RàPC) est une technique de résolution de problèmes basée sur la réutilisation par analogie d'expériences passées appelées *cas*. Un cas représente notamment un problème et la solution qui lui a été appliquée. Un cas est indexé pour permettre de le retrouver suivant certaines caractéristiques pertinentes et discriminantes. Ces caractéristiques, aussi appelées *indices*, déterminent dans quelle situation le cas peut être de nouveau réutilisé. Le raisonnement se décompose habituellement en quatre phases principales (Aamodt & Plaza, 1994) : la *recherche* des cas pertinents dans la mémoire, la *réutilisation* des solutions passées pour constituer une solution au problème courant, la *révision* de la solution pour l'affiner et enfin *l'apprentissage* de la nouvelle expérience pour améliorer le système et enrichir sa mémoire.

L'application du RàPC à des problèmes ayant ces caractéristiques présente une spécificité qui touche principalement la gestion des indices des cas sous la forme d'historiques. Peu de travaux en RàPC ont abordé cette spécificité : prévision de l'évolution de feux de forêt (Rougegrez, 1994a), prévision économique (Nakhaeizadeh, 1994), prévision médicale et épidémiologique (Schmidt *et al.*, 1996 ; Bull *et al.*, 1997), commande de robots mobiles (Ram & Santamaria, 1997), aide à la supervision de processus (Fuchs *et al.*, 1995), et assistance à la navigation sur le Web (Corvaisier *et al.*, 1998). De plus, la plupart de ces approches sont dédiées à une application particulière, et aucun de ces travaux ne vise une classe de problèmes aussi générale, ni ne fournit des outils réutilisables. Enfin, les approches existantes ne permettent pas une gestion des cas soutenant l'effort de découverte des connaissances, nécessaire dans ces problèmes où les connaissances initiales sont réduites.

Nous nous fixons alors deux objectifs principaux :

- la définition d'un *modèle générique d'indexation des cas* adapté à la classe de problèmes visée,
- et la réalisation d'un *outil logiciel ouvert* pour faciliter son utilisation.

En définitive, nous abordons trois des problèmes importants soulevés par Janet Kolodner, un pionnier dans le domaine du raisonnement à partir de cas, qui citait parmi les objectifs de recherche du domaine (Kolodner, 1993, pages 571–576) : « l'extraction de cas depuis des situations continues » et la « définition d'un vocabulaire général d'indexation » qui doivent être considérées dans le modèle d'indexation, ainsi que la conception « d'outils pour la réalisation de systèmes RàPC ».

## 2 Contributions

Les deux premières contributions de notre étude sont :

- La proposition d'un modèle d'indexation générique, appelé *modèle d'indexation par situations comportementales*, pour faciliter l'application du RàPC au sein de la classe de problèmes visée.
- La proposition d'un nouveau type d'outils logiciels pour le RàPC à travers la conception et la réalisation d'une plate-forme à objets (*object-oriented application framework*), nommée CBR\*Tools. Cette plate-forme repose sur des points d'ouverture et des modèles à objets exprimés systématiquement en termes de patrons de conception (*design patterns*). Cet outil facilite l'utilisation de notre modèle d'indexation et son intégration dans des systèmes de RàPC.

Pour évaluer ces deux contributions, nous avons principalement réalisé un *système d'aide à la navigation sur le Web*, nommé Broadway. De par la nature de notre modèle d'indexation, l'approche concrétisée dans Broadway est originale. Nous présentons ainsi une troisième contribution, avec la

---

proposition d'un nouveau type d'assistants à la navigation sur le Web, basé sur le calcul de recommandations par réutilisation de comportements observés à partir d'un ensemble extensible de variables.

## 2.1 Présentation des contributions

Premièrement, nous proposons pour la classe de problèmes visée, une approche générale de l'indexation alliant un modèle de représentation générique avec un guide d'utilisation présentant différents principes de gestion des indices dans un cycle de raisonnement à partir de cas. Ces principes donnent des exemples de types de traitements identifiés et analysés lors de l'étude des applications réalisées. Le modèle de représentation générique permet de séparer les données brutes (structurées en variables et en enregistrements) des cas qui référencent ces données à travers des situations comportementales. Une situation comportementale est principalement formée de comportements élémentaires (séquences issues d'historiques) reliés par des contraintes temporelles. Ce modèle permet d'une part de représenter des indices sur des historiques échantillonnés et/ou à événements, et d'autre part de manipuler les cas pour mener à la découverte des connaissances. En effet, l'introduction du concept de patron de cas potentiels permet d'extraire, suivant des hypothèses préétablies, des enseignements structurés à partir des données d'observation. Ces extractions peuvent être stockées sous forme de cas concrets dont les représentations peuvent être modifiées suivant les raisonnements. De plus, ces cas référencent toujours les données brutes d'observation ce qui permet de faciliter leur interprétation.

Nous proposons deuxièmement un nouveau type d'outils logiciels dans le domaine du RàPC tirant profit des concepts de plate-forme à objets (*object-oriented application framework*), de patron de conception (*design pattern*) et de cas d'utilisation (*use case*). Nous proposons tout d'abord une description de haut niveau, sous la forme de cas d'utilisation liant les tâches de conception d'un système de RàPC vers les points d'ouverture nécessaires à l'application du RàPC et de notre modèle d'indexation. Ces points d'ouverture sont intégrés dans une architecture abstraite au sein d'une plate-forme à objets dont la modélisation s'appuie systématiquement sur l'utilisation de patrons de conception. Puis, la plate-forme réalisée, appelée CBR\*Tools, fournit également un ensemble de classes concrètes pouvant être directement utilisées (index, similarité, base de cas). Cette plate-forme facilite la spécialisation concrète de notre modèle d'indexation et son intégration avec des méthodes spécifiques de recherche, d'adaptation, de révision et d'apprentissage nécessaires à la réalisation d'une application. Notre approche permet également la capitalisation d'objets-métier réutilisables sous la forme, par exemple, d'index, de fonction de similarité et de schémas de phase de raisonnement.

Enfin, nous proposons un nouveau type d'assistants à la navigation sur le Web dont le prototype Broadway<sup>1</sup> est un premier exemple. Cette assistance est basée sur la réutilisation des navigations passées d'un groupe d'utilisateurs dont les comportements sont observés suivant un ensemble extensible de variables. Dans Broadway, quatre variables sont utilisées : adresse, contenu, évaluation explicite, et évaluation implicite. Cette observation par variables, combinée avec notre modèle d'indexation, permet la gestion de comportements détaillés dans un cadre flexible et générique.

Nos trois contributions ont fait l'objet de sept publications scientifiques nationales et internationales : modèle d'indexation par situations comportementales (Jaczynski, 1997 ; Jaczynski & Trousse, 1997a), plate-forme à objets pour le RàPC (Jaczynski & Trousse, 1999 ; Jaczynski &

---

<sup>1</sup> [Browsing advisor reusing pathways](#).

Trousse, 1998a) et le système Broadway (Jaczynski & Trousse, 1998b ; Jaczynski & Trousse, 1998c ; Jaczynski & Trousse, 1997b).

## 2.2 Evaluation expérimentale

L'évaluation de notre modèle d'indexation a été réalisée dans le cadre de la nutrition des plantes et dans celui de l'assistance à la navigation sur le Web. Plus précisément, la première application nous a permis de vérifier que les caractéristiques du processus de nutrition peuvent être prises en compte par le modèle de représentation. Un prototype de recherche des cas a été réalisé en se basant sur des données brutes enregistrées manuellement par R. Brun, chercheur de l'INRA Sophia Antipolis. Toutefois, la spécification et la validation d'un système de régulation avec une approche de RàPC demandent une mise en opération effective qui n'a pas été effectuée. En effet, la réalisation d'un tel système requiert des compétences métier et une lourde disponibilité des experts du domaine. C'est pourquoi nous avons retenu le domaine de la navigation sur le Web : les données réelles sont plus facilement disponibles et nous avons été plus autonomes pour concevoir et tester l'application de notre modèle d'indexation. Nous avons alors réalisé le système d'assistance Broadway qui montre l'utilisation opérationnelle de notre modèle d'indexation ainsi que son intégration dans un cycle complet de raisonnement à partir de cas.

L'approche de plate-forme à objets pour le RàPC et les modèles à objets proposés ont mené à la réalisation concrète de CBR\*Tools comprenant plus de 200 classes programmées dans le langage Java et conçues sous l'atelier de génie logiciel Rational Rose (Rose, 1998). En plus des deux applications précédentes réalisées avec CBR\*Tools, nous avons également implanté un système classique de RàPC permettant l'affectation d'un facteur de risque à une voiture pour un assureur. Ce système n'utilise pas notre modèle d'indexation par situations comportementales car il avait uniquement pour but de valider l'architecture et les composants de base.

Notre assistant de navigation Broadway fut évalué lors d'une expérimentation qui a été conçue et supervisée par des étudiants en DESS d'ergonomie (Hébraud *et al.*, 1998). Cette expérimentation a mis en jeu dix utilisateurs réels (étudiants en psychologie n'ayant pas de connexion avec nos travaux) et a nécessité l'utilisation continue de deux systèmes Broadway pendant dix jours. Les résultats de cette expérimentation montrent que l'approche est bénéfique dans le cadre restreint choisi.

## 3 Plans de lecture

Le document est organisé en cinq chapitres et permet de développer notre approche visant des modèles et des outils pour l'application du RàPC à une classe de problèmes. Un aperçu rapide de la thèse peut être obtenu en complétant la lecture de l'introduction et de la conclusion de la thèse, par les résumés placés au début de chaque chapitre.

Nous conseillons à tout lecteur de commencer par le chapitre I qui présente plus précisément nos motivations et notre démarche pour l'application du raisonnement à partir de cas à la classe des problèmes nécessitant la prise en compte d'historiques. Le raisonnement à partir de cas et la notion d'indexation sont également introduits.

Dans le cadre d'une lecture complète, nous proposons ensuite, dans le chapitre II, notre modèle d'indexation des cas par situations comportementales, après avoir mené un état de l'art des approches existantes d'indexation par historiques. Nous décrivons également l'utilisation du modèle d'indexation dans le cadre de la nutrition des plantes. Dans le chapitre III, nous menons un état de l'art des outils

logiciels pour le RàPC et nous présentons le concept de plate-forme à objets qui nous a permis de dépasser leurs limites en terme d'ouverture. Nous analysons alors les points d'ouverture nécessaires pour le raisonnement à partir de cas et plus spécifiquement pour l'indexation par situations comportementales. Dans le chapitre IV, nous présentons l'utilisation des patrons de conception pour la modélisation ouverte du cycle du RàPC et de l'indexation par situations comportementales. Nous introduisons alors la plate-forme CBR\*Tools que nous avons réalisée. Enfin, dans le chapitre V, nous détaillons la conception de Broadway, notre assistant à la navigation sur le Web qui met en œuvre le modèle d'indexation par situations comportementales ainsi que la plate-forme CBR\*Tools.

Un lecteur intéressé par le raisonnement à partir de cas et par l'indexation par situations comportementales pourra directement lire la proposition du modèle (cf. chapitre II, §2 et §3), les points d'ouverture associés de notre plate-forme (cf. chapitre III, §2.3), le cycle de raisonnement de Broadway (cf. chapitre V, §1.3 et §2.1) et l'évaluation du modèle dans le cadre de Broadway (cf. chapitre V, §2.3.1).

Un lecteur intéressé par notre plate-forme à objets pour la réalisation de systèmes de RàPC pourra directement lire la proposition de cette approche (cf. chapitre III, §2) et les modèles à objets associés (cf. chapitre IV). La description du niveau spécifique pour l'indexation par situations comportementales, présentée séparément (cf. chapitre III, §2.3.2 ; chapitre IV, §2), repose sur les concepts du modèle d'indexation qui peuvent être rapidement cernés (cf. chapitre II, §2.1.2 et §2.2). Nous décrivons également un exemple d'utilisation de la plate-forme pour Broadway ainsi que son évaluation (cf. chapitre V, §2.2 et §2.3.2).

Un lecteur intéressé par notre application Broadway pourra se reporter à la présentation du système (cf. chapitre V, §1) et à l'évaluation de notre approche (cf. chapitre V, §3). Cependant, la représentation des cas et des indices ainsi que la réalisation (cf. chapitre V, §2) s'appuient sur les notions introduites précédemment (cf. chapitre II, §2 ; chapitre III, §2.3 ; chapitre IV, §1 et §2).

Nous fournissons également plusieurs annexes pour compléter notre étude. L'annexe A décrit les éléments de la notation UML (*Unified Modelling Language*) utilisée dans les chapitres III, IV et V. L'annexe B présente les patrons de conception (*design patterns*) utilisés dans CBR\*Tools. L'annexe C décrit un exemple complet de l'utilisation de CBR\*Tools dans le cadre du calcul du facteur de risque d'une voiture pour un assureur. Et enfin, l'annexe D liste principalement les paramètres de Broadway permettant de modifier le cycle du raisonnement suivi.

