INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*CBR*Tools : une bibliothèque objet pour l'indexation des cas par situation comportementale*

Michel Jaczynski - Brigitte Trousse

N° 3215

juillet 1997

THÈME 3

ISSN 0249-6399

CBR*Tools : une bibliothèque objet pour l'indexation des cas par situation comportementale

Michel Jaczynski[1] - Brigitte Trousse[2]

Thème 3 — Interaction homme-machine,
images, données connaissances

Action AID

Rapport de recherche n° 3215— juillet 1997 — 38 pages

Résumé : dans le raisonnement à partir de cas (RàPC), la situation d'un cas décrit dans quel cadre sa connaissance est pertinente et nous faisons dans nos travaux une distinction entre d'une part une situation instantanée, ensemble d'indices définissant un état à un instant particulier, et d'autre part une situation comportementale, ensemble d'indices décrivant principalement une évolution dans le temps de cet état. Dans cet article, nous nous intéressons uniquement à l'indexation de cas par situation comportementale dont nous présentons une modélisation objet. Un des points forts de cette modélisation réside dans la représentation flexible des situations comportementales, qui permet notamment de prendre en compte : des comportements échantillonnés et/ou à événements, des comportements futurs estimés, et des descriptions de situations temporelles simples ou complexes. Nous proposons également une stratégie de remémoration ouverte pouvant être spécialisée pour une application particulière. Enfin, nous présentons la bibliothèque logicielle générique CBR*Tools que nous avons développée pour l'indexation par situation comportementale en RàPC.

Mots-clé : raisonnement par cas, représentation des cas, comportement, situation comportementale, temps, indexation, stratégie de remémoration, modélisation objet, Java.

(Abstract : pto)

---

[1] Email : Michel.Jaczynski@sophia.inria.fr

[2] Email : Brigitte.Trousse@sophia.inria.fr

Unité de recherche INRIA Sophia Antipolis
2004, route de Lucioles - BP 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)
Téléphone : 04 93 65 77 77 – Télécopie : 04 93 65 77 65

CBR*Tools: an Object Oriented Library for Indexing Cases with Behavioral Situations

Abstract: In case-based reasoning, the situation of a case defines when its knowledge is relevant, and we make in our work a clear distinction between case indexing techniques based on an instantaneous situation, a set of indices giving the state of the world at a particular instant, and a behavioral situation, a set of indices describing mainly the evolution of this state. The focus of this article is on behavior-based indexing techniques and we introduce our object-oriented model for this kind of indexing. In this model we propose mainly a flexible representation of behavioral situation that takes into account: sampled and/or event-based behaviors, estimated behaviors, and the description of complex or simple temporal situations. In addition, we propose an open retrieval strategy that can be specialized for specific applications. Finally, we describe our generic software library CBR*Tools that supports our model.

Keywords: case-based reasoning, case representation, behavior, behavioral situation, indexing, time, retrieval strategy, objet-oriented model, Java.

Table des matières

1. Introduction 5

2. Présentation de l'indexation par situation comportementale 7

2.1 Définition d'une situation comportementale 7

2.1.1 Définition du concept de « comportement » 7
2.1.2 Définition du concept de « situation comportementale » 9

2.2 Intérêt de l'indexation par situation comportementale 9

3. Etat de l'art 10

3.1 Critères de classification 10

3.2 Analyse critique des systèmes existants 11

3.2.1 Type d'application 11
3.2.2 Observation du processus 12
3.2.3 Type de situations comportementales : explicite ou non-explicite 13
3.2.4 Structure des situations comportementales : figée ou dynamique 14
3.2.5 Composante instantanée d'une situation 14
3.2.6 Méthodologie de remémoration 15

3.3 Synthèse 15

3.3.1 Limites des systèmes existants 15
3.3.2 Notre approche 17

4. Modélisation objet de l'indexation des cas par situation comportementale 18

4.1 Modélisation de l'observation du processus 18

4.1.1 Regroupement des observations dans les « enregistrements » 18
4.1.2 Modèle objet de représentation des enregistrements 19

4.2 Représentation des cas et du problème courant 20

4.2.1 Représentation d'une situation comportementale : une structure flexible 20
4.2.2 Différents types de cas : abstraits, concrets et potentiels 23
4.2.3 Représentation du problème courant 24

4.3 Stratégie et représentation de la remémoration des cas 24

4.3.1 Définition de la stratégie de remémoration 25
4.3.2 Modèle objet de représentation de la stratégie 27

5. Réalisation et exemple d'utilisation de CBR*Tools 28

5.1 Réalisation de CBR*Tools 29

5.2 Application à la conduite de la nutrition de plantes 29

5.2.1 Cadre et objectifs de l'application 29
5.2.2 Représentation des données d'observation et des cas 30
5.2.3 Remémoration basée sur des situations comportementales 31

6. Conclusion 32

Références bibliographiques 34

Annexe 37

Table des figures

Figure 1 : Exemple de chronique échantillonnée numérique 8
Figure 2 : historique concis de lieux d'habitation 9
Figure 3 : exemple de cas avec une situation non-explicite dans SINS. 13
Figure 4 : caractéristiques des systèmes existants et objectifs de CBR*Tools 16
Figure 5 : diagramme de classes des enregistrements et des chroniques 20
Figure 6 : diagramme des classes pour la représentation des situations comportementales 21
Figure 7 : exemples de contraintes de localisation 22
Figure 8 : exemple de situation complexe 23
Figure 9 : décomposition en étapes de la stratégie de remémoration 25
Figure 10 : diagramme de classes des index 28
Figure 11 : exemple de cas concret pour la nutrition de plantes 31
Figure 12 : éléments utilisés de la notation UML 37

1. Introduction

De manière générale, le raisonnement par cas (RàPC) est une approche 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 (ou une méthode permettant de la générer). Le raisonnement se décompose habituellement en quatre phases principales [AAM 94] :

1. phase de *remémoration* dont le but est de rechercher des cas ayant des similarités avec le problème courant ;

2. phase de *réutilisation* permettant de construire une solution au problème courant en se basant sur les cas identifiés dans la phase précédente ;

3. phase de *révision* de la solution qui permet de l'affiner grâce à son évaluation ;

4. phase d'*apprentissage* chargée de mettre à jour les éléments du raisonnement en prenant en compte l'expérience qui vient d'être réalisée, et qui pourra ainsi être utilisée dans les raisonnements futurs.

Ainsi, un *cas* représente une expérience passée dont l'enseignement peut être utile lorsqu'un nouveau problème se présente. Généralement, un cas est indexé pour permettre de le retrouver suivant certaines caractéristiques pertinentes et discriminantes. Ces caractéristiques, aussi appelées *indices*, déterminent dans quelle situation (ou contexte) le cas peut être de nouveau réutilisé. La problématique de la phase de remémoration est donc de permettre d'identifier un certain nombre de cas ayant des indices similaires au problème courant : il est en effet peu probable de retrouver un cas correspondant exactement au problème courant. Un système de raisonnement par cas doit permettre l'expression des indices pour les différents cas, et doit disposer de structures d'indexation offrant une recherche efficace tout en utilisant des connaissances du domaine et/ou des connaissances induites à partir de son expérience. Dans ce sens, l'objectif de la phase de remémoration dépasse les approches classiques de recherche des bases de données même si des techniques issues de ce domaine sont parfois utilisées.

Dans notre étude, nous faisons une distinction entre d'une part une indexation basée sur une *situation instantanée*, ensemble d'indices définissant un état à un instant particulier et d'autre part une indexation basée sur une *situation comportementale*, ensemble d'indices décrivant une évolution dans le temps de cet état. L'indexation par situation instantanée est l'approche la plus répandue dans les systèmes de raisonnement à

partir de cas : par exemple les systèmes classiques[3] CHEF [HAM 86], JULIA [HIN 89], CASEY [KOT 89], et MEDIATOR [KOL 89] utilisent une indexation par situation instantanée. Le système CHEF permet de dériver des recettes suivant notamment les ingrédients et le mode de cuisson désirés en se basant sur d'autres recettes stockées dans des cas. Les cas sont indexés suivant un ensemble de buts qu'ils permettent de satisfaire ou d'échecs qu'ils permettent d'éviter. Les indices utilisés décrivent uniquement les objectifs pertinents à l'instant où le cas a été appris : il s'agit d'une situation instantanée.

L'aspect comportemental permet souvent de mieux décrire une situation et représente pour de nombreux problèmes une composante essentielle de la description de la situation. De manière générale en médecine, l'évolution de l'état d'un patient est souvent plus importante qu'un état particulier [KER 95], et les fiches médicales doivent être consultées. Dans les problèmes d'analyse et de compréhension du comportement d'un être humain utilisant une machine, il est nécessaire de prendre en compte des séquences temporelles d'interactions [RIT 94]. Dans le domaine de la conduite de procédés, des délais de réactions importants et variables peuvent être observés, ce qui oblige à prendre en compte des événements passés de plusieurs minutes et même de plusieurs heures : par exemple, dans la conduite de hauts fourneaux, une erreur d'alimentation peut être observée six heures plus tard [JAK 90]. La prise en compte d'un ensemble de comportements dans l'indexation a été nécessaire à la réalisation de plusieurs applications basées sur le raisonnement à partir de cas : prévision de l'évolution de feux de forêt [ROU 94], prévision économique [NAK 94], prévision médicale et épidémiologique [SCH 96, BUL 97], commande de robots mobiles [RAM 93] et aide à la supervision de processus [FUC 95].

Notre objectif est de définir des méthodes et des outils *génériques* pour le raisonnement à partir de cas dédiés à une classe de problèmes nécessitant une indexation par situation comportementale. Nous souhaitons faciliter le développement de telles applications grâce à une bibliothèque logicielle, offrant d'une part un ensemble de composants de RàPC directement réutilisables et d'autre part en permettant l'extension des méthodes proposées. La plupart des approches existantes sont dédiées à une application particulière et bien que certaines soient formulées en partie de manière plus générale [ROU 94, NAK 94, BUL 97], aucune d'entre elles ne satisfait ces objectifs.

[3] Ces systèmes sont décrits et analysés dans [KOL 93].

Plus précisément, notre but est de fournir dans cette bibliothèque des composants que le concepteur d'une application peut utiliser et étendre pour :

- représenter les données d'observations suivant les caractéristiques du processus,
- représenter les situations comportementales de manière adaptée aux types d'observations et aux connaissances du domaine d'application,
- et construire la phase de remémoration des cas sur la base d'une stratégie ouverte.

Dans ce cadre, nous avons conçu et développé une bibliothèque d'objets réutilisables, nommée CBR*Tools, pour la représentation et la recherche des cas nécessitant une indexation par situation comportementale.

La structuration de l'article est alors la suivante. Dans la section 2, nous présentons tout d'abord l'indexation par situation comportementale. Puis dans la section 3, nous mènerons une analyse critique des approches existantes afin de positionner nos objectifs. Ensuite nous décrirons dans la section 4 la modélisation proposée pour l'indexation des cas par situation comportementale. Dans la section 5, nous donnerons quelques éléments de la réalisation de CBR*Tools et l'utilisation de cette bibliothèque sera alors illustrée par une application de conduite de la nutrition de plantes issue de l'INRA[4] [BRU 93].

2. Présentation de l'indexation par situation comportementale

Nous définissons tout d'abord le concept de *situation comportementale* dont les éléments principaux sont les *comportements,* puis nous montrerons l'intérêt et les conditions d'application de ce type d'indexation.

2.1 Définition d'une situation comportementale

2.1.1 Définition du concept de « comportement »

L'indexation par situation comportementale suppose l'observation d'un *processus* qui évolue dans le temps comme c'est souvent la cas dans des problèmes de diagnostic, de prévision, ou de commande. Un processus représente un procédé physique ou une activité humaine se déroulant dans un environnement. Le processus possède un ensemble de

[4] Institut National de Recherche en Agronomie, unité de Sophia Antipolis.

variables caractéristiques de son état. Ces variables prennent des valeurs à différents instants, utilisant une unité de temps propre à l'observation du processus.

L'évolution et les séquences de valeurs d'une variable peuvent être formalisées par le concept de *chronique* aussi appelée *série chronologique* [BAR71]. Une chronique est une application d'un ensemble fini et totalement ordonné, noté *T*, dans un ensemble *E*, correspondant au domaine de valeurs d'une variable d'observation. L'ensemble *T*, appelé *chronologie*, est un sous-ensemble de l'échelle des temps, noté Ω, qui est un ensemble totalement ordonné. Une *sous-chronique* est une chronique définie sur un sous-ensemble d'une chronologie. Une chronique représente donc l'observation d'un processus suivant une variable sur un ensemble fini d'instants pour l'échelle des temps considérée. Un *comportement* est alors une sous-chronique particulière de l'observation du processus.

Différents types de chroniques peuvent être définis suivant les propriétés de l'échelle des temps, du domaine et de la sémantique des éléments de l'ensemble *E*. Par exemple, si la durée séparant deux valeurs successives de la chronologie est toujours constante[5], nous parlerons alors d'une *chronique échantillonnée* (cf. Figure 1). Nous considérons dans notre présentation que les chroniques échantillonnées sont observées à une fréquence suffisante, permettant de rendre compte de l'évolution de la variable. Nous n'abordons pas ici des problèmes de traitements du signal bien que ces techniques puissent être utilisées en amont.

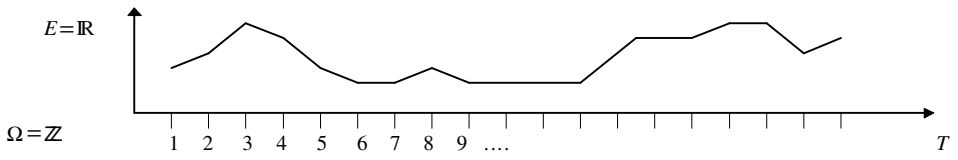

Figure 1 : Exemple de chronique échantillonnée numérique

Si la chronique n'est pas échantillonnée, nous parlerons de *chronique à événements* où un événement est une information datée et instantanée extraite du déroulement du processus [CAU 86]. Ces événements peuvent avoir différentes sémantiques : alarmes, début ou fin d'une action, déroulement d'une action instantanée pour l'échelle des temps considérée, changement d'état. Par exemple, si les événements possèdent la sémantique d'un changement d'état, nous obtenons un comportement qui représente un *historique concis* [WIL 86]. Dans ce type de chronique, la valeur de la variable est considérée comme constante entre deux mesures, dénotant un épisode de l'historique (cf. Figure 2).

[5] Ω est alors muni d'une structure de groupe abélien et d'une origine.

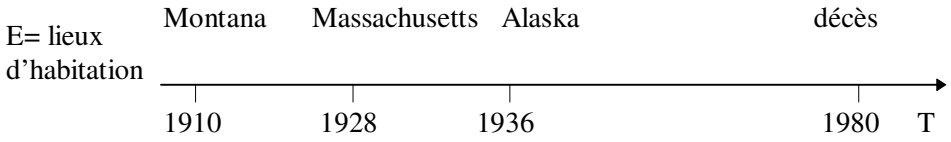

Figure 2 : historique concis de lieux d'habitation (adaptation de [WIL 86])

2.1.2 Définition du concept de « situation comportementale »

Nous utilisons ici le terme *situation* dans le sens qui lui est donné dans les travaux de raisonnement à partir de cas [KOL 93:145], à savoir qu'une situation contient l'ensemble des indices, caractéristiques pertinentes et discriminantes d'un cas ou du problème courant. Les valeurs de ces indices définissent dans quel cadre l'enseignement du cas peut être intéressant par rapport au problème courant et aux autres cas.

Nous définissons une *situation comportementale* comme étant un ensemble d'indices qui représente en particulier les comportements pertinents de certaines variables du processus et leurs relations temporelles. Ces indices doivent être pertinents pour l'activité de résolution de problème considérée et toute la problématique de la description d'une situation comportementale réside dans la sélection des comportements pertinents parmi l'ensemble des données d'observation. Une situation comportementale doit être vue comme une extension temporelle d'une situation instantanée, et donc n'exclut pas que certains indices représentent des valeurs instantanées.

2.2 Intérêt de l'indexation par situation comportementale

Dans le cadre des applications utilisant le RàPC et basées sur l'observation d'un processus, il s'agit de déterminer quelles sont les caractéristiques de la classe de problèmes nécessitant l'indexation par situation comportementale. L'observation d'un processus ne représente en effet qu'un prérequis. Dans certaines approches [DET 94, KOP 88, BRA 91], l'observation d'un réseau de communication fait partie inhérente de l'application mais l'aspect comportemental n'a pas été jugé pertinent pour l'indexation, bien que la dimension temporelle ait été prise en compte dans la description incrémentale du problème [DET 94] ou dans la représentation de plans et de comportements attendus dans la solution des cas [KOP 88, BRA 91].

L'indexation par situation comportementale doit être utilisée pour les problèmes qui présentent les deux caractéristiques suivantes :

- les évolutions dans le temps des variables du processus sont pertinentes pour l'activité de résolution de problèmes ; et/ou les relations temporelles entre les valeurs des différentes variables d'un processus sont pertinentes ;

- il n'est pas possible par manque de connaissance, d'abstraire ou de modéliser les différents comportements et leurs relations pour décrire une situation instantanée.

Ces caractéristiques se retrouvent dans de nombreux systèmes existants en RàPC qui abordent d'une part des problèmes de prévision et d'autre part des problèmes de conduite et de supervision de processus (cf. §3.2.1). De manière générale, un cas est indexé suivant une situation comportementale et décrit la solution associée à cette situation (par exemple l'évolution observée du processus ou les paramètres de commande utilisés). Le raisonnement permet alors de déduire la solution du problème courant à partir des cas remémorés dont la situation est similaire à celle du problème.

3. Etat de l'art

Nous présentons tout d'abord les critères de classification que nous utiliserons pour l'analyse comparative des systèmes existants. Enfin, nous présenterons notre problématique prenant en compte les limites des approches actuelles.

3.1 Critères de classification

Nous utilisons six critères de classification des systèmes existants qui sont pertinents pour la représentation et la remémoration des cas indexés par une situation comportementale :

1. *Type d'application* : les systèmes existants abordent des problèmes ayant des domaines d'application ou des objectifs communs.

2. *Observation du processus* : les caractéristiques de l'observation sont importantes dans la mesure où elles engendrent des contraintes fortes dans la représentation et la remémoration des cas.

3. *Type de situation comportementale* : la situation peut être explicite (chaque cas ayant une situation prédéterminée) ou non-explicite (la situation est alors construite lors de la remémoration).

4. *Structure de la situation comportementale* : la structure peut être dynamique (chaque situation est composée d'un ensemble variable d'indices) ou figée (toutes les situations ont la même structure).

5. *Composante instantanée d'une situation* : cette composante peut être présente ou non en conjonction avec une composante comportementale dans une situation.

6. *Méthodologies de remémoration* : différentes techniques de recherche sont employées et certaines plus générales, présentent un aspect méthodologique pour la remémoration.

3.2 Analyse critique des systèmes existants

A partir des six critères précédents, nous présentons une analyse critique des principaux systèmes existants utilisant une indexation par situation comportementale. Les différents éléments de la comparaison sont réunis à la fin de cette section dans la Figure 4.

3.2.1 Type d'application

Nous classons les approches existantes en considérant leur type d'application : *prévision* ou *commande* de processus dynamiques. Dans le domaine de la prévision de l'évolution de processus, le système REBECAS [ROU 94] aborde la prévision de l'évolution des incendies de forêt. L'approche de Nakhaeizadeh [NAK 94] a été appliquée dans le domaine économique à la prévision de taux d'intérêt et à la prévision du nombre d'immatriculations de voitures et de camions en France. Le système de Bull *et al* [BUL 97] détecte des risques épidémiologiques et permet également la prévision de l'expansion des épidémies ou de ses conséquences. Dans le domaine médical, le système ICONS [SCH 96] effectue des estimations de l'évolution de la fonction rénale d'un patient.

Dans les problèmes de commande et de supervision, le système SINS « Self-Improving Navigation System » [RAM 93] a été conçu dans le cadre de la commande d'un robot mobile. Ce système est très intéressant dans la mesure où il montre l'amélioration du raisonnement par rapport à une indexation basée sur une situation instantanée, implantée dans le système ACBARR [MOO 92], prédécesseur de SINS. Le

système PADIM « Poste d'Aide à la Décision Immédiate » [FUC 95] aide un opérateur en charge de la supervision d'un procédé complexe en lui proposant les descriptions visuelles les plus adaptées.

3.2.2 Observation du processus

Relativement à l'observation du processus, nous avons identifié trois caractéristiques : le nombre de variables, les types de chroniques, et la prise en compte d'observations futures estimées dans la définition du problème à résoudre. Les systèmes existants sont basés soit sur une *unique* variable soit sur une nombre *multiple* de variables. La majorité des systèmes utilisent un seul flux de données [NAK 94, SCH 96, BUL 97, FUC 95] mais SINS et REBECAS définissent plusieurs variables. Dans REBECAS, différentes chroniques sont prises en compte représentant l'évolution : du relief, de la végétation, du vent et de la propagation du feu. Dans SINS, l'observation du robot évoluant dans son environnement est basée sur des chroniques représentant notamment : la densité des obstacles, la progression du robot vers l'objectif, et les différents paramètres de commande. L'utilisation de plusieurs variables est intéressante car elle permet de garder la séparation entre différentes grandeurs et différents modes d'observation ; il est alors possible de spécifier des méthodes de mise en correspondance spécifiques et adaptées à chaque variable.

Différents types de chroniques sont utilisés mais les chroniques *échantillonnées* sont les plus répandues [NAK 94, BUL 97, SCH 96, RAM 93]. Dans le système de Bull *et al* [BUL 97], une situation comportementale est constituée d'une séquence de scénarios où chaque scénario donne l'état de la santé publique et de la charge des services de santé pour la durée d'une semaine. L'observation du système ICONS [SCH 96] est basée sur plusieurs dizaines de paramètres qui sont traités pour obtenir l'état journalier d'un patient. Seuls PADIM et REBECAS gèrent des chroniques *à événements*. PADIM intègre dans la description de la situation des séquences d'événements qui sont issus de l'évolution du procédé et des actions de l'opérateur. Dans REBECAS, le feu est observé grâce à des relevés effectués sur le front du feu et les informations continues provenant de cartes sont transformées en chroniques à événements, (par exemple, le relief est transformé grâce à la reconnaissance de portions de relief). Les chroniques à événements permettent une meilleure représentation car chaque événement possède une sémantique précise. Cependant si les observations ne sont pas par nature événementielles, des méthodes de transformations des données échantillonnées en événements peuvent être difficiles à définir, ce qui oblige les applications à considérer directement les chroniques échantillonnées. REBECAS utilise de plus un type de chroniques particulier, que nous appellerons *chroniques spatiales*, où la chronologie est basée sur la distance parcourue depuis le point de départ du feu.

Enfin, certaines données d'observations peuvent être spécifiques à l'expression du problème à résoudre : REBECAS prend en compte les observations estimées ou *comportements futurs estimés* permettant de mieux définir la situation du problème.

3.2.3 Type de situations comportementales : explicite ou non-explicite

Nous pouvons définir deux types de situations comportementales : l'une *explicite* et l'autre *non-explicite*. Dans les approches existantes, cette distinction provient de la solution choisie pour la granularité temporelle des cas où il s'agit de déterminer les frontières temporelles d'un cas : un cas représente-t-il une longue expérience ou une expérience particulière et assez précise dans le temps ? Ce problème, déjà analysé dans [KOL 93:168], est d'ailleurs commun à tous les systèmes de raisonnement à partir de cas opérant sur un processus évoluant dans le temps, qu'ils utilisent ou non l'indexation par situation comportementale.

SINS et REBECAS considèrent tous deux qu'un cas est une expérience étendue dans le temps. Pour REBECAS, un cas représente l'évolution du relief, de la végétation, du vent et de la propagation pour un feu tout entier depuis son début jusqu'à sa fin, selon une direction de propagation définie. Pour SINS, un cas est un historique des évolutions des attributs de l'environnement et des paramètres de commande pour une durée déterminée. Dans ces deux approches, un cas est alors indexé par une situation comportementale non-explicite dans le sens où le cas sera remémoré si une partie variable de sa description correspond au problème courant. En fait dans SINS et REBECAS, un cas n'a pas de séparation figée entre sa situation comportementale et la solution qu'il représente : la situation et la solution sont déterminées après mise en correspondance avec le problème courant (cf. Figure 3).

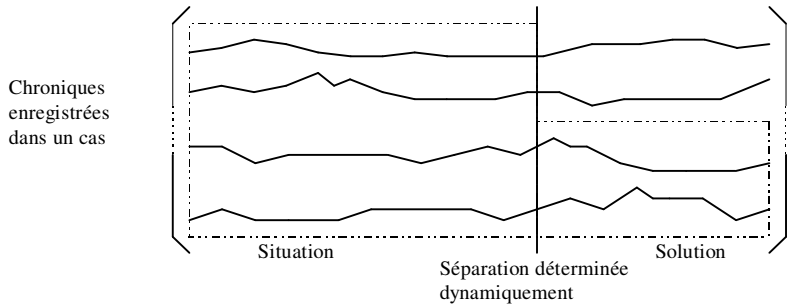

Figure 3 : exemple de cas avec une situation non-explicite dans SINS.

Dans les autres approches, un cas représente une expérience élémentaire qui définit une situation comportementale explicite. Par exemple, dans PADIM la chronique des

événements du système et de l'opérateur est explicitement associée au cas de supervision. Dans le système ICONS, l'expérience représentée par un cas contient une situation comportementale donnant l'évolution du patient dans les sept jours précédents.

L'utilisation de situations explicites ou non-explicites présente chacune des avantages et des inconvénients. La représentation de situations explicites permet d'associer à chaque situation des connaissances spécifiques mais mène à la création de nombreux cas. D'un autre côté, la gestion de situations non-explicites évite la création de cas élémentaires mais impose des méthodes de recherche plus coûteuses.

3.2.4 Structure des situations comportementales : figée ou dynamique

Nous avons identifié deux types de structures des situations comportementales dans les systèmes existants : une *structure figée* ou une *structure dynamique*. Dans les structures figées [NAK 94, SCH 96], toutes les situations comportementales sont décrites de la même façon et par le même nombre de valeurs. Cette méthode est utilisée par exemple dans ICONS où la situation comporte un ensemble fini de descriptions sur des horizons de temps constants et limités à sept jours. En effet, le système prend l'hypothèse, confirmée par les médecins, que les comportements plus anciens ne sont jamais pertinents.

Les structures dynamiques permettent de définir pour chaque cas, une situation composée de comportements pertinents spécifiques qui sont sélectionnés suivant différentes méthodes : *sélection d'événements* pertinents [FUC 95], sélection de sous-chroniques basée sur un intervalle de temps (*fenêtres temporelles*) [RAM 93, ROU 94, BUL 97].

Les structures dynamiques sont plus générales que les structures figées et surtout permettent de mieux définir une situation comportementale. En effet, l'échec d'un raisonnement peut être causé par la prise en compte de données non pertinentes ou par l'absence de représentation de données pertinentes ; une structure dynamique permet de mieux satisfaire ces contraintes. D'un autre côté, les structure figées peuvent être justifiées par des connaissances du domaine et permettent une mise en correspondance plus simple.

3.2.5 Composante instantanée d'une situation

La situation peut être strictement comportementale ou référencer également une *composante instantanée*. La majorité des travaux se limitent à l'utilisation d'une situation strictement comportementale et seul PADIM [FUC 95] réutilise des cas indexés par une situation composée d'une partie instantanée donnant notamment l'état du système, le

profil de l'utilisateur et le but de la supervision, et d'une autre partie comportementale composée d'une chronique d'événements.

3.2.6 Méthodologie de remémoration

Toutes les approches existantes proposent des techniques de mise en correspondance basées sur la *recherche de chroniques* similaires. Ces méthodes sont souvent dédiées à une application, cependant certaines caractéristiques de la remémoration nous semblent plus générales. Dans REBECAS, l'utilisation de *points de vue* permet de guider la recherche suivant une des variables d'observation du processus (le vent, la végétation ou le relief). ICONS met en œuvre une recherche basée sur des *tendances*, représentant le comportement du patient sur différents intervalles de temps de un à sept jours. Les points de vue ou les tendances permettent une mise en correspondance plus efficace en ne considérant dans un premier temps qu'une partie de la situation. PADIM commence la remémoration par un *filtrage sur la composante instantanée* de la situation puis prend en compte l'aspect comportemental, lequel est souvent plus complexe et d'un niveau d'abstraction moins élevé que la composante instantanée.

3.3 Synthèse

3.3.1 Limites des systèmes existants

Suite à l'analyse présentée précédemment, nous faisons le constat qu'aucune approche existante n'est conçue dans le but d'aborder l'indexation par situation comportementale de façon globale et paramétrée ; les représentations et les techniques utilisées sont donc plutôt dédiées à un type d'application.

En effet, pour l'observation du processus, deux grands types de chroniques sont utilisés : les chroniques échantillonnées et à événements. Cependant aucune approche existante ne permet de sélectionner suivant les propriétés du processus, une observation échantillonnée ou à événements. De plus, les systèmes existants ne permettent pas l'observation des systèmes dynamiques *hybrides* où il est nécessaire de représenter des chroniques à la fois échantillonnées et à événements.

La flexibilité de la représentation des situations comportementales est basée sur les trois caractéristiques suivantes : situations explicites et/ou non-explicites, structures figées ou dynamiques, et intégrant ou non une composante instantanée. Aucun système n'offre de cadre unifié pour prendre en compte l'ensemble de ces aspects.

	REBECAS [ROU 94]	Nakhaeizadeh [NAK 94]	Bull *et al* [BUL 97]	ICONS [SCH 96]	SINS [RAM 93]	PADIM [FUC 95]	**Objectifs de CBR*Tools**
Type d'application	prévision	prévision	prévision	prévision	commande	commande (supervision)	***bibliothèque générique***
Observation							
variables	multiple	unique	unique	unique	multiple	unique	**multiple**
type chroniques	événements et spatiales	échantillonnée	échantillonnée	échantillonnée	échantillonnée	événements	• **échantillonnée** • **événements** • ***hybride***
comportements futurs estimés	oui	non	non	non	non	non	**oui**
Type situation	non-explicite	explicite	explicite	explicite	non-explicite	explicite	• **explicite** • **non-explicite**
Structure situation							
type structure	dynamique	figée	dynamique	figée	dynamique	dynamique	• **figée** • **dynamique**
type sélection	fenêtres temporelles	-	fenêtre temporelle	-	fenêtre temporelle	événements	• **fenêtres temporelles** • **événements** • ***contraintes temporelles***
Composante instantanée	non	non	non	non	non	oui	**oui**
Méthodologie de remémoration	recherche de chroniques, points de vue	recherche de chroniques	recherche de chroniques	recherche de chroniques, tendances	recherche de chroniques	recherche de chroniques, filtrage	***stratégie ouverte***

Figure 4 : caractéristiques des systèmes existants et objectifs de CBR*Tools

De plus, les structures existantes ne prennent pas en compte des contraintes temporelles explicites entre comportements. Ces contraintes sont nécessaires pour indexer des cas suivant des scénarios ou les explications d'un expert. Les scénarios mettent l'accent sur les relations temporelles entre événements et sont connus par exemple dans des problèmes de supervision de systèmes industriels [GHA 89]. Des explications peuvent être données par un expert (réel ou virtuel) pour décrire l'utilité des cas ; les cas doivent alors être indexés suivant une situation composée des indices observables de ces explications [KOL 93:272] qui peuvent inclurent des contraintes temporelles explicites entre comportements.

Certains systèmes existants reposent sur des aspects méthodologiques intéressants pour la remémoration : points de vue [ROU 94], tendances [SCH 96], étape de remémoration basée sur la composante instantanée [FUC 95]. Cependant, aucune méthodologie de recherche n'intègre ces différents aspects.

3.3.2 Notre approche

Notre objectif est de concevoir et de réaliser une bibliothèque logicielle, appelée CBR*Tools, pour l'indexation par situation comportementale. Cette bibliothèque doit dépasser les limites existantes en offrant un ensemble de composants réutilisables permettant :

- *la représentation des données d'observation suivant les caractéristiques du processus*. Nous souhaitons supporter une observation suivant une ou plusieurs variables et avec des chroniques échantillonnées et/ou à événements. La situation du problème à résoudre peut également intégrer des observations estimées ;

- *la représentation des situations comportementales de manière adaptée aux types d'observation et aux connaissances du domaine*. Il s'agit donc de définir un cadre flexible de représentation qui doit permettre de choisir ou d'utiliser en conjonction différents types de situations (explicites ou non-explicites), et différents types de structures (figées ou dynamiques) avec l'intégration optionnelle d'une composante instantanée. La structure des situations doit également intégrer la sélection de fenêtres temporelles et des événements pertinents, et doit permettre la représentation de contraintes temporelles explicites entre comportements ;

- *la construction de la phase de remémoration des cas sur la base d'une stratégie ouverte*. Notre but est de proposer une structuration des étapes de remémoration qui peut être restreinte ou étendue suivant les besoins et les connaissances disponibles.

Nous prenons également en compte des contraintes pratiques (occupation mémoire et efficacité de remémoration) inhérentes aux problèmes d'indexation. Enfin, nous avons adopté une approche objet de conception car elle permet à la fois de définir des composants finaux directement réutilisables tout en offrant un modèle ouvert qui peut être adapté par spécialisation ou par composition. Nos principaux, objectifs en comparaison aux approches existantes, sont résumés dans la Figure 4.

4. Modélisation objet de l'indexation des cas par situation comportementale

Nous présentons la modélisation objet proposée dans notre bibliothèque logicielle CBR*Tools ainsi que la justification des choix effectués suivant nos trois principaux objectifs concernant l'observation du processus, la représentation des situations comportementales utilisée dans les cas et le problème courant, et la remémoration. CBR*Tools, implantée en Java, offre un ensemble de classes directement utilisables tout en définissant des classes abstraites et des protocoles de méthodes permettant son extension.

4.1 Modélisation de l'observation du processus

Notre modèle de l'observation du processus s'appuie sur le concept d'enregistrement qui regroupe les observations et permet de définir les situations comportementales en référençant les données d'un enregistrement.

4.1.1 Regroupement des observations dans les « enregistrements »

Un *enregistrement* regroupe un ensemble de chroniques correspondant aux variables d'observation du processus, pour une période de temps définie et pour un contexte donné. La notion d'enregistrement et son utilisation présentent quatre avantages principaux. Tout d'abord, l'observation du processus est segmenté vis-à-vis d'un contexte donnant les caractéristiques constantes d'un enregistrement, par exemple la localisation de l'expérience, les modes opératoires ou des objectifs. Il est en effet courant de pouvoir identifier des périodes d'observation, parfois appelées sessions, sans lien de causalité.

Ensuite, toutes les chroniques d'un enregistrement sont définies sur la même échelle des temps. Nous posons ainsi un ordre total, local à un enregistrement, entre les éléments des chronologies. Ainsi les relations temporelles inutiles entre les mesures de chroniques d'enregistrements différents ne sont pas représentées.

Le stockage et la manipulation des chroniques en mémoire peuvent de plus être optimisés. En effet, chaque chronique gère sa représentation tout en maintenant des fonctions d'accès uniformes. Les chroniques numériques peuvent alors être compressées par approximation (compression avec perte) par des approches de segmentations [GAR 91]. Les chroniques symboliques peuvent également être compressées sur l'alphabet défini par les valeurs des variables durant un enregistrement, en utilisant la compression de Huffman [HEL 91]. Ces deux méthodes de compression sont très efficaces étant donné que l'ensemble des éléments de chaque chronique est fini et connu pour un enregistrement. La décompression se fait lors du premier accès à la chronique[6].

Enfin, les situations comportementales sont construites en référençant certaines valeurs des chroniques (cf. §4.2), ce qui évite la duplication des données dans chaque situation qui peuvent se recouvrir. De même, le contexte représentant une des parties de la composante instantanée des situations n'est pas non plus dupliqué dans chaque situation.

4.1.2 Modèle objet de représentation des enregistrements

Nous donnons dans la Figure 5 la modélisation objet des enregistrements et des chroniques qui est implantée dans CBR*Tools. La hiérarchie de chroniques montre les différents types de chroniques possibles et notamment la chronique avec prévision. Ce type de chronique permet de prendre en compte les prévisions des observations (comportements futurs estimés) de manière transparente tout en facilitant leur mise à jour par composition d'une chronique d'observations et d'une chronique de prévisions.

Cette modélisation permet de satisfaire nos objectifs en permettant la représentation des données d'observation suivant les caractéristiques d'un processus particulier. En effet, elle offre la possibilité de définir une sous-classe d'un enregistrement qui contiendra le nombre de variables et les types de chroniques requis : une ou plusieurs variables, chroniques échantillonnées ou à événements. De plus, la hiérarchie des chroniques peut être étendue pour permettre par exemple la représentation des chroniques spatiales (cf. [ROU 94]) ou d'autres types de chroniques (domaine de valeurs spécifique, sémantique particulière des événements), grâce à la définition de nouvelles classes.

[6] Une méthode plus complexe, semblable au mécanisme de mémoire cache, peut être envisagée pour la gestion de la décompression et de la compression des chroniques.

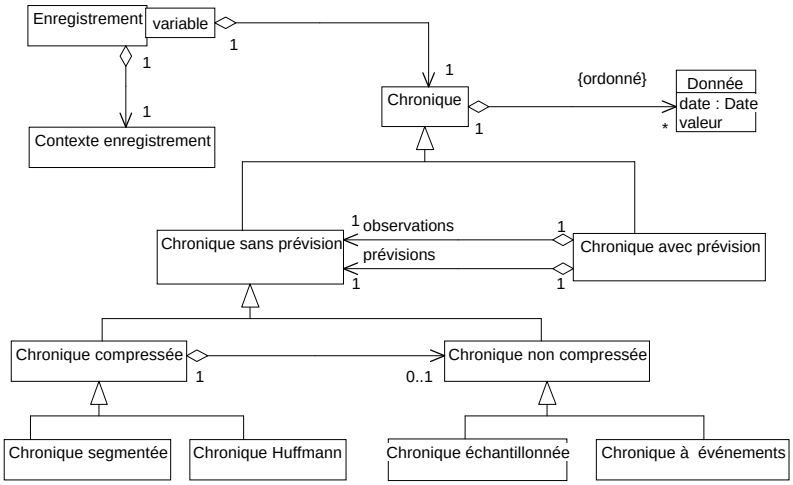

Figure 5 : diagramme de classes des enregistrements et des chroniques (notation UML, cf. annexe)

4.2 Représentation des cas et du problème courant

Nous proposons la représentation de cas abstraits, concrets et potentiels. Ces différents types de cas et le problème courant reposent sur la représentation de situations comportementales.

4.2.1 Représentation d'une situation comportementale : une structure flexible

Une situation comportementale doit décrire de manière relative par rapport à un instant particulier les comportements pertinents et leur relations. Ainsi, lorsqu'un nouveau problème est à résoudre, il est possible d'identifier dans les données d'observation du problème et par rapport à l'instant courant, les comportements pertinents et leurs relations. Cette étape d'identification sera effectuée lors de la remémoration, mais implique une représentation des situations adéquates.

La modélisation objet de la représentation que nous proposons est donnée dans la Figure 6. Une *situation mixte* est constituée d'une *composante instantanée* et d'une *composante comportementale*. Dans CBR*Tools, nous appellerons *situation comportementale*, une situation mixte ayant au moins une composante comportementale. La composante instantanée comprend le contexte de l'enregistrement et d'autres données à caractère instantané qui peuvent être définies pour une application particulière (par exemple le profil d'un utilisateur ou l'état global du système comme dans PADIM). La

composante comportementale référence un enregistrement à une date précise, appelée *date de référence*, et définit un ensemble de *comportements élémentaires* pertinents pour chaque variable ainsi que des *contraintes temporelles* entre ces comportements.

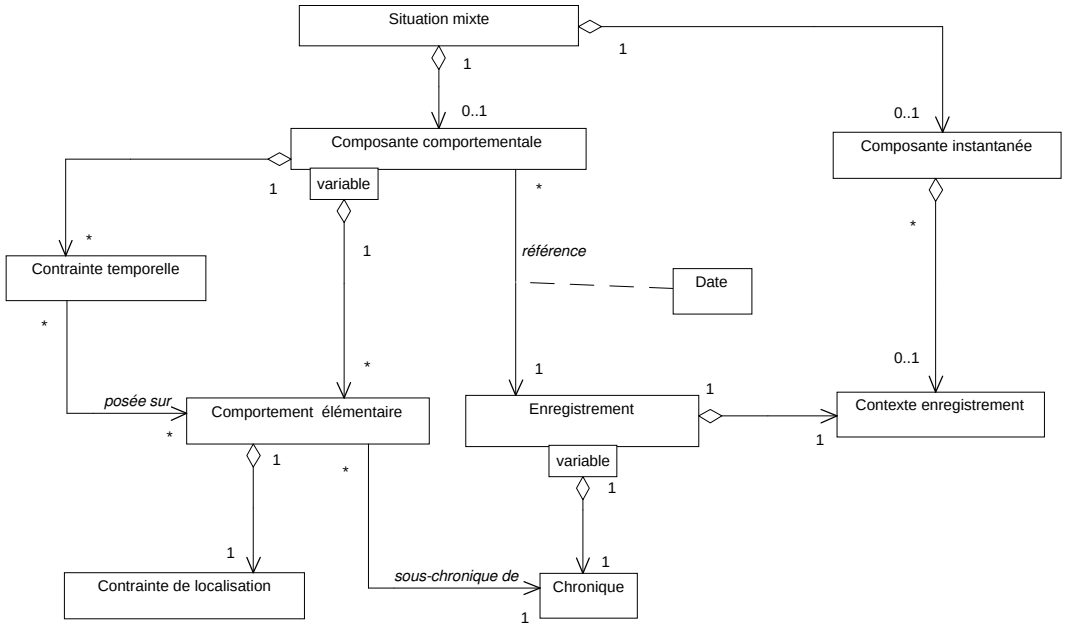

Figure 6 : diagramme des classes pour la représentation des situations comportementales

Un comportement élémentaire définit une sous-chronique dans l'enregistrement de la situation dont la sous-chronologie est spécifiée en donnant l'élément minimum et l'élément maximum. Chaque comportement élémentaire possède une contrainte de localisation. Cette contrainte est exprimée de façon relative par rapport à la date de référence et doit permettre d'identifier le comportement homologue dans la situation d'un problème.

Nous avons défini trois types de contraintes de localisation (cf. Figure 7) :

- *par position* : les bornes de la sous-chronologie sont identifiées par le nombre de mesures depuis la date de référence,

- *par temps écoulé* : les bornes de la sous-chronologie sont identifiées par le temps écoulé relativement à la date de référence,

– *par valeur* : les bornes de la sous-chronologie doivent être recherchées de telle manière que le comportement soit similaire au sens de la similarité définie pour la variable considérée. Des contraintes additionnelles peuvent être définies de manière à restreindre l'espace de recherche.

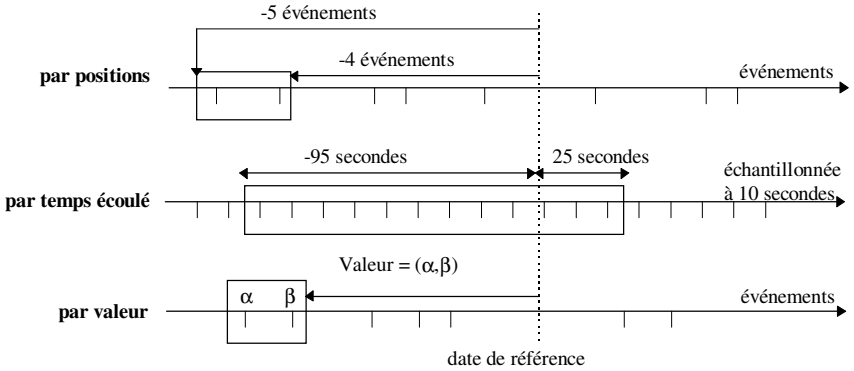

Figure 7 : exemples de contraintes de localisation

La définition de ces comportements pertinents élémentaires permet d'obtenir un ensemble d'éléments de même nature bien que les chroniques de l'enregistrement puissent être différentes (échantillonnées, à événements). Ces comportements peuvent alors être reliés par des contraintes temporelles entre intervalles (relations d'Allen [ALL 84]) définis par le début et la fin de chaque comportement élémentaire. Un ensemble de contraintes temporelles peut être généré en se basant sur les contraintes temporelles satisfaites dans la situation. Toutefois, il est possible de sélectionner un sous-ensemble de ces contraintes de manière à ne prendre en compte que celles jugées réellement pertinentes.

Le formalisme objet que nous avons choisi pour la représentation des situations comportementales est très flexible puisque notre modèle permet :

– la sélection d'un ensemble de comportements pertinents élémentaires pour chaque variable, chacun étant défini par une fenêtre temporelle,

– la sélection d'événements pertinents, lorsqu'un comportement pertinent élémentaire est réduit à un événement,

– la description de contraintes temporelles entre comportements élémentaires,

– la définition d'une composante instantanée *optionnelle* qui prend notamment en compte les éléments pertinents du contexte associé à un enregistrement.

Par exemple, nous pouvons représenter des situations :

– simple : « *La situation A est définie à l'instant t_0 par les cinq dernières valeurs de toutes les chroniques* ». Dans ce cas, les comportements élémentaires sont définis par position et il n'y a pas de contraintes temporelles.

– ou complexe : « *La situation B est définie à l'instant t_1 par les valeurs de la variable V_1 dans les 10 dernières secondes, par une forte augmentation des valeurs de la variable V_2, et par l'événement α suivi de β sur la variable V_3 , β se produisant pendant l'augmentation de V_2* ». Cette situation (cf. Figure 8) montre également comment il est possible de prendre en compte des chroniques échantillonnées (V_2) et à événements (V_1 et V_3) grâce aux comportements élémentaires.

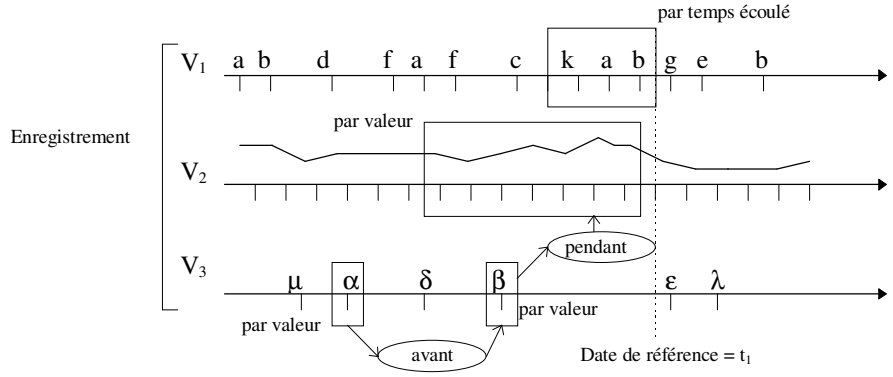

Figure 8 : exemple de situation complexe

4.2.2 Différents types de cas : abstraits, concrets et potentiels

Les cas abstraits peuvent être disponibles dans les connaissances du domaine, ou issues de la généralisation automatique[7] ou manuelle d'expériences concrètes [BER 96]. Dans notre modèle, un cas abstrait est un cas possédant une situation comportementale dont les chroniques ne représentent pas directement des observations réelles : les chroniques peuvent être construites à partir de connaissances expertes ou induites à partir d'expériences. Des enregistrements fictifs sont alors construits. Contrairement à un cas

[7] Toutefois, nous n'abordons pas les aspects de généralisation automatique dans la cadre de la version actuelle de CBR*Tools.

abstrait, un cas concret possède une situation comportementale basée sur des observations réelles (*vécues*). Les cas abstraits et concrets reposent sur l'utilisation d'une situation explicite.

Un *cas potentiel* est une notion que nous introduisons dans nos travaux pour prendre en compte les situations non-explicites. Un cas potentiel n'a pas d'existence propre, mais il est défini à travers un *patron de situations comportementales*. Un patron est composé d'une situation comportementale sans relation avec un enregistrement particulier mais qui définit un ensemble de contraintes de localisation et de contraintes temporelles. Ce patron peut être instancié pour une date de référence donnée et ainsi devenir un cas concret. Les patrons sont utilisés lors de la remémoration pour expliciter des connaissances jusqu'alors cachées dans l'ensemble des données des enregistrements.

Dans notre modèle, les cas abstraits, les cas concrets et les patrons sont des objets référençant une situation comportementale. Notre approche permet de représenter des situations de structure figée, grâce à l'utilisation d'un unique patron de situations et généralise cette méthode en offrant la possibilité de définir plusieurs patrons. Le modèle permet, de plus, la définition de situations ayant une structure dynamique, c'est-à-dire spécifique à chaque cas et dépendante des connaissances propres à chaque expérience.

4.2.3 Représentation du problème courant

Le problème courant référence une situation comportementale dans laquelle aucun comportement pertinent élémentaire n'a été défini. Ces comportements seront identifiés lors de la mise en correspondance avec la situation des cas et suivant les indices de chaque cas. La situation du problème peut également être liée à un enregistrement dont certaines chroniques gèrent les comportements futurs estimés.

4.3 Stratégie et représentation de la remémoration des cas

Nous présentons tout d'abord notre stratégie de remémoration basée sur la représentation des situations comportementales et des différents types de cas que nous avons définis. Puis, nous décrirons la modélisation ouverte adoptée pour représenter cette stratégie, permettant ainsi son extension suivant les besoins spécifiques inhérents à toute application de notre classe de problèmes.

4.3.1 Définition de la stratégie de remémoration

La stratégie de remémoration que nous proposons est basée sur les heuristiques suivantes :

- il est préférable de réutiliser des cas abstraits plutôt que des cas concrets puisqu'ils sont issus des connaissances du domaine ou de généralisation ;

- il est préférable de réutiliser des cas concrets plutôt que des cas potentiels, car les cas concrets peuvent provenir d'experts, être reliés à des cas d'échec ou de succès, ou porter des informations spécifiques quant à leur réutilisation ;

- il est inutile de chercher le cas le plus similaire, mais il suffit d'identifier des cas satisfaisant un critère minimum de similarité.

Suivant ces heuristiques, notre stratégie se décompose en trois alternatives : remémoration des cas abstraits, remémoration des cas concrets et remémoration des cas potentiels. Le résultat global de la remémoration correspond à celui de la première alternative dont le résultat est non vide. Chaque alternative se décompose à son tour en étapes de filtrage et de sélection (cf. Figure 9). Les étapes de filtrage ont pour but de restreindre l'espace de recherche selon des contraintes qui représentent notamment le critère minimum de similarité exigé. Les étapes de sélection ordonnent les meilleurs cas remémorés suivant leur similarité.

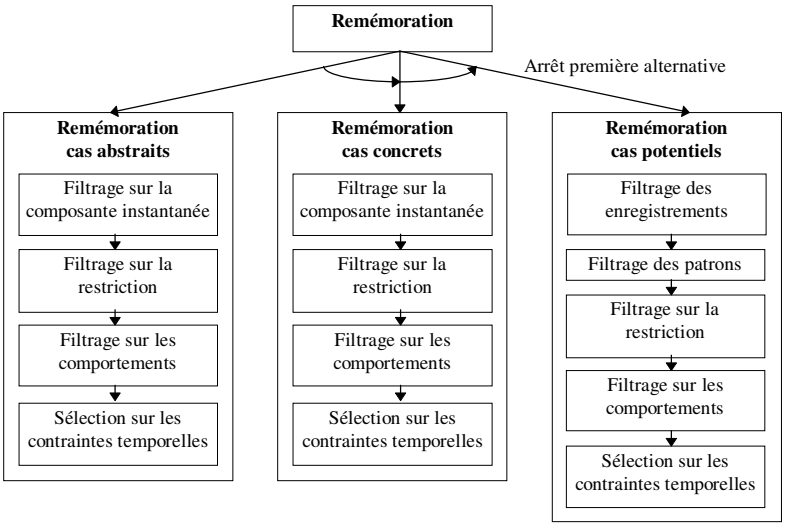

Figure 9 : décomposition en étapes de la stratégie de remémoration

Ainsi, nous proposons la décomposition suivante pour la remémoration des cas abstraits et des cas concrets :

1. *Filtrage sur la composante instantanée*. Nous commençons, tout comme dans PADIM [FUC 95], par un filtrage suivant la composante instantanée. La composante instantanée comprend dans notre modèle le contexte de l'enregistrement de la situation, et il est souvent utile d'éliminer de la remémoration les cas ayant un contexte complètement différent car les connaissances d'adaptation peuvent être inexistantes ou difficiles à mettre en œuvre.

2. *Filtrage sur une restriction de la situation comportementale*. Une restriction correspond à la généralisation de la notion de point de vue de REBECAS [ROU 94] (restriction sur les variables) et des tendances de ICONS [SCH 96] (restriction sur des horizons temporels). Ce filtrage permet de ne retenir que les cas ayant un minimum de similarité sur certaines variables et sur un horizon de temps restreint. Cette étape nécessite une mise en correspondance moins complexe et moins coûteuse que l'étape suivante ce qui rend plus efficace l'ensemble de la remémoration.

3. *Filtrage sur les comportements élémentaires pertinents*. L'ensemble des comportements élémentaires est pris en compte pour la mise en correspondance des cas et du problème courant. Avant chaque comparaison entre un cas et un problème, une étape d'identification des comportements élémentaires doit être effectuée. En effet, à partir des contraintes de localisation de chaque comportement d'un cas, les comportements homologues dans le problème sont identifiés (par position, par temps écoulé ou par valeur). Ensuite la mise en correspondance entre les comportements élémentaires peut être effectuée.

4. *Sélection basée sur les contraintes temporelles*. Grâce à l'étape précédente, les contraintes temporelles sont instanciées dans le contexte du problème courant. Les contraintes sont alors vérifiées et les meilleurs cas sont sélectionnés suivant une mesure de similarité.

Nous proposons également l'algorithme suivant pour la remémoration des cas potentiels :

1. *Filtrage des enregistrements*. Les enregistrements ayant un contexte suffisamment similaire avec l'enregistrement courant sont identifiés.

2. *Filtrage des patrons*. Cette étape permet d'obtenir les patrons qui peuvent être appliqués étant donnée la situation du problème.

3. *Filtrage et sélection des cas potentiels*. Sur tous les enregistrements identifiés, chaque patron est instancié pour un ensemble de dates de référence et donne des cas concrets temporaires. Les dates de références sont déterminées en parcourant les valeurs d'une chronique, appelée *chronique de déplacement*. Les étapes de filtrage sur une restriction et sur les comportements puis l'étape de sélection sur les contraintes temporelles sont exécutées de la même façon que précédemment. Les cas concrets temporaires remémorés deviendront définitifs suivant l'utilisation qui en sera faite dans la suite du raisonnement[8].

L'introduction du concept de *cas potentiel* d'une part et l'utilisation de *cas concrets* et *abstraits* d'autre part permettent de manipuler des situations non-explicites et explicites dans le même cadre. Nous réunissons dans notre modélisation ces deux types de situations en évitant les inconvénients des situations non-explicites des approches existantes [ROU 94, RAM 94]. En effet, nous offrons la possibilité de définir des patrons de situations (sorte de situations non-explicites) mais lorsqu'un cas basé sur une telle situation est utilisé, il est transformé en un cas concret, permettant d'associer à cette expérience d'autres informations importantes pour le raisonnement, notamment des liens avec des cas d'échec et de succès, et l'évaluation de sa réutilisation. De plus, la mise en évidence de situations explicites permet d'améliorer la remémoration en évitant une recherche linéaire à chaque raisonnement pour retrouver une situation non-explicite. Cette méthode engendre la création de nombreux cas dont la représentation est optimisée grâce à la factorisation des données dans les enregistrements.

4.3.2 Modèle objet de représentation de la stratégie

Nous proposons un modèle objet basé sur les *index* et leur structuration pour construire des stratégies complexes de remémoration. Un index est une structure de données permettant d'accéder via la description d'indices à un ensemble de cas. Nous donnons, dans la Figure 10, un extrait de la hiérarchie d'index que nous avons définie dans CBR*Tools. Cette hiérarchie est basée sur deux grands types d'index : les *index simples*, représentant les index existants en RàPC (par exemple KNN[9], K-d Trees[10]

[8] Tous les cas concrets ainsi identifiés ne seront pas forcément pertinents à long terme ; il appartient donc à la phase d'apprentissage de décider si ces cas doivent être mémorisés pour les prochains raisonnements.

[9] KNN (K-Nearest Neighbours) est une méthode de recherche permettant d'identifier les k-plus proches cas du problème courant étant donnée une similarité (complexité linéaire en fonction du nombre de cas O(n)).

[10] Les K-d Trees permettent de réduire la complexité de la recherche KNN, mais nécessitent la construction au préalable d'un arbre de discrimination (la complexité moyenne est de $O(k\ n\ \log_2 n.)$ pour la construction de l'arbre, et de $O(\log_2 n)$ pour la recherche).

[WES 94], Prototypes[11] [JAC 94]), et les *index composites*, permettant de construire un index complexe par composition. Parmi les index composites, les index d'alternatives permettent de gérer différentes branches d'indexation et les index séquentiels permettent de connecter des index pour former des séquences.

Ainsi la structure de base de notre stratégie de remémoration est construite grâce à un index « première alternative » composé de trois index de type séquentiel. Ensuite, il est possible de modéliser la remémoration particulière liée à une application en spécifiant les index à utiliser pour chaque étape avec beaucoup de flexibilité : ajout ou suppression d'étapes suivant les besoins. L'atout de cette modélisation est de pouvoir ajouter des nouvelles étapes complexes de manière transparente puisqu'un index composite hérite de la classe de base des index. D'ailleurs, nous proposons d'autres index composites comme des index faisant l'union ou l'intersection des résultats.

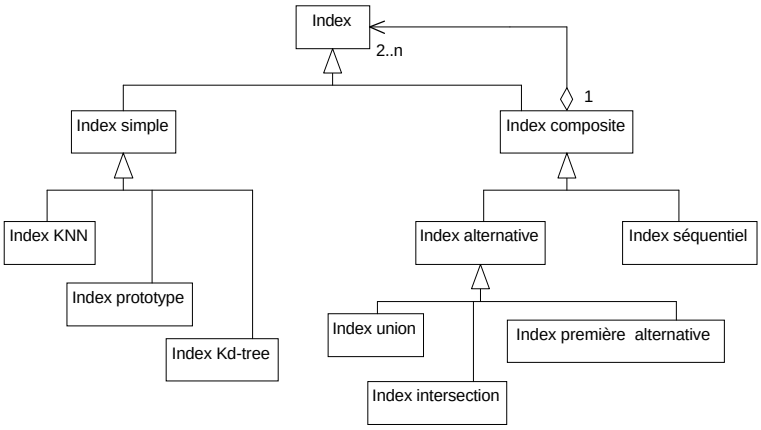

Figure 10 : diagramme de classes des index

5. Réalisation et exemple d'utilisation de CBR*Tools

Nous décrivons tout d'abord l'architecture générale de réalisation de la bibliothèque CBR*Tools. Puis, afin d'illustrer la modélisation introduite précédemment, nous présentons une application en cours d'implantation à partir de CBR*Tools, dans le domaine de la nutrition de plantes.

[11] Cette méthode de recherche est basée sur une hiérarchie de prototypes. Chaque prototype définit un ensemble de contraintes dures ou flexibles (utilisation d'ensembles flous et de préférences). L'algorithme de recherche effectue une classification du problème courant pour identifier les prototypes les plus spécifiques auxquels ont été rattachés les cas compatibles.

*5.1 Réalisation de CBR*Tools*

CBR*Tools est implanté avec le langage de programmation JAVA et repose actuellement sur une structure à deux niveaux : représentation et gestion des mécanismes de base du raisonnement à partir de cas, et une spécialisation dédiée à l'indexation par situation comportementale. Le premier niveau offre des hiérarchies de classes permettant notamment la représentation de la base de cas, des indices, des index (prototypes exacts et flous [15], Kd-tree [27], KNN) et des similarités. Ce niveau définit également des classes gérant le raisonnement à partir de cas et un ensemble de protocoles pour l'indexation et la remémoration. Le deuxième niveau propose les composants spécifiques que nous avons décrits dans cet article dont notamment : les enregistrements, les chroniques, les situations comportementales, les patrons de situations. Cette structuration permet de mettre en évidence dans le premier niveau les concepts dont la portée est plus générale que l'indexation par situation comportementale.

Ainsi, grâce à l'utilisation des classes d'objets fournis dans CBR*Tools, le développement de systèmes de RàPC qui nécessitent une indexation par situation comportementale est simplifié. De plus, la modélisation objet que nous proposons peut être étendue pour prendre en compte les aspects spécifiques inhérents à toute application de notre classe de problèmes : types de chroniques, contextes des enregistrements, similarités entre comportements, étapes de remémoration particulières.

5.2 Application à la conduite de la nutrition de plantes

5.2.1 Cadre et objectifs de l'application

L'objectif des techniques mises en place pour la nutrition automatique des plantes à l'INRA [BRU 93] est de gérer la nutrition minérale et hydrique en évitant à la fois les excès et les déficits. Plus précisément, il s'agit de calculer chaque jour la concentration de la solution d'irrigation de manière à garder la concentration de la solution de drainage autour d'une consigne. Les connaissances sur le domaine sont très peu élaborées et se réduisent à des graphes de relations qualitatives. Une approche basée sur le raisonnement à partir de cas alliant une indexation par situation comportementale est utilisée pour assister les chercheurs de ce domaine à mettre en évidence les données pertinentes pour la nutrition.

L'indexation par situation comportementale nous est apparue indispensable dans ce problème. En effet, les experts du domaine ont observé des délais de réaction variables dans le temps : une modification de la solution d'irrigation a un impact sur la

concentration du drainage après un délai de 2 à 5 jours suivant la période de l'année (été ou hiver). Ainsi, il s'agit de tenir compte dans la régulation : des actions passées de régulation, des évolutions passées de l'environnement et également des prévisions météorologiques à court terme pour anticiper l'évolution de l'environnement et améliorer la régulation [BRU 93].

Notre objectif est donc de réaliser grâce à CBR*Tools un système de raisonnement à partir de cas permettant :

- de représenter dans un cas les actions de régulation prises dans une situation où les données pertinentes sont fournies,
- de mettre à jour les connaissances de chaque cas en modifiant la sélection des données pertinentes,
- de proposer des actions de régulation basées sur des cas remémorés.

5.2.2 Représentation des données d'observation et des cas

Dans cette application, le processus à observer comprend un ensemble de plantes et l'environnement. Le processus est observé par différentes variables donnant pour chaque jour : la température, le rayonnement solaire, l'humidité, et les concentrations des solutions d'irrigation et de drainage. Ainsi un enregistrement doit stocker les différentes mesures de ces variables dans des chroniques échantillonnées numériques où l'unité de temps est le jour. Le contexte de l'enregistrement permet de lui associer la valeur de la consigne à maintenir pour la concentration du drainage. L'implantation de l'observation du processus est alors effectuée grâce à la spécialisation des classes des enregistrements et des contextes de CBR*Tools. Plusieurs expériences sont menées simultanément avec des consignes différentes et ainsi plusieurs enregistrements sont créés.

Un cas est alors composé d'une situation comportementale, de la concentration de la solution d'irrigation correspondante et de l'évaluation de la régulation. Cette évaluation est calculée à partir de l'évolution de la concentration de la solution de drainage. Deux patrons de cas potentiels sont définis pour prendre en compte de manière standard les différentes configurations de l'environnement : sélection des évolutions passées sur les 2 derniers jours pour l'été, et sur les 5 derniers jours pour l'hiver. L'utilisateur du système peut : soit créer directement un nouveau cas concret suivant ces observations et ses connaissances, soit créer un nouveau cas à partir des cas potentiels remémorés, soit modifier des anciens cas concrets suivant de nouvelles connaissances mises en évidence durant les expériences de régulation. Des cas abstraits sont également définis et représentent des cas limites dans lesquels le processus sort des seuils acceptables de régulation. Les cas concrets, abstraits et les patrons sont définis par spécialisation des

classes existantes dans CBR*Tools. La Figure 11 représente un exemple de cas concret où la situation comportementale est composée d'une part de comportements sélectionnés par position issus d'un patron de situations, et d'autre part d'un comportement par valeur défini par un expert.

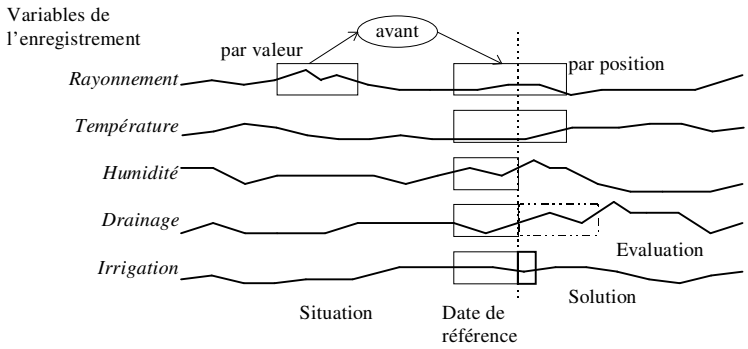

Figure 11 : exemple de cas concret pour la nutrition de plantes

5.2.3 Remémoration basée sur des situations comportementales

Nous avons construit la phase de remémoration en se basant sur notre stratégie de remémoration en trois alternatives et en composant les index disponibles dans CBR*Tools. La remémoration des cas abstraits permet d'anticiper des échecs du système et ainsi donne la possibilité de réagir à temps avec d'autres méthodes de régulation. Cette fonctionnalité nous semble très importante lorsqu'il s'agit d'opérer sur un processus réel. Si aucun cas abstrait n'est remémoré, la recherche des cas concrets est exécutée. La remémoration des cas abstraits et concrets utilise les quatre étapes que nous avons définies :

1. Filtrage sur la composante instantanée : la recherche est restreinte aux cas provenant d'enregistrements ayant la même consigne pour la concentration du drainage. En effet, nous n'avons pas encore de connaissances pour réutiliser les cas issus de contraintes de régulation différentes.

2. Filtrage sur une restriction : la restriction correspond à la sélection des évolutions passées sur les 2 derniers jours sachant que ces données sont pertinentes dans tous les cas. Ce filtrage est effectué en utilisant un index de type Kd-tree [WES 94] utilisant une similarité numérique basée sur les écarts de valeurs.

3. Filtrage sur les comportements élémentaires : chaque comportement est mis en correspondance avec son comportement homologue grâce à une mesure de similarité basée sur la moyenne des écarts de chaque mesure de la chronique.

Ensuite, une similarité globale est calculée en faisant la moyenne des similarités de chaque comportement. Un index de type KNN avec une contrainte de similarité minimale permet d'identifier un sous-ensemble de cas.

4. Sélection basée sur les contraintes temporelles : un index de type KNN permet d'identifier les meilleurs cas en utilisant une mesure de similarité calculant le ratio de contraintes satisfaites.

Finalement, si aucun cas abstrait ni concret n'est remémoré, la recherche des cas potentiels est effectuée. Le filtrage des enregistrements se fait par une requête de correspondance exacte sur la consigne de régulation pour les mêmes raisons que précédemment. Le filtrage des patrons de situations comportementales est basé sur les périodes de l'année, été ou hiver, auxquelles ces patrons sont associés. Ensuite, la sélection des cas potentiels est effectuée en une seule étape étant donnée la structure simple des patrons. Nous utilisons alors un index KNN adapté, qui parcourt les enregistrements, instancie les patrons et classe les cas suivant leur similarité.

6. Conclusion

Nous avons présenté dans cet article la modélisation objet de l'indexation des cas par situation comportementale que nous avons implantée dans la bibliothèque logicielle CBR*Tools. Notre démarche se distingue des approches existantes qui se limitent aux besoins d'une application particulière et qui ne proposent pas une synthèse de ce type d'indexation. Notre bibliothèque offre un ensemble de composants logiciels réutilisables facilitant le développement de systèmes de RàPC qui nécessitent ce type d'indexation :

- Nous proposons tout d'abord un modèle objet permettant la représentation des données d'observation suivant les caractéristiques du processus : observation sur une ou plusieurs variables, utilisation des chroniques échantillonnées et/ou à événements, et la description de comportements futurs estimés dans la situation du problème. Nous offrons une représentation optimisée grâce aux enregistrements qui évitent la duplication des données dans les situations comportementales et qui permettent l'application de méthodes de compression des comportements pour réduire les coûts de stockage.

- Nous offrons ensuite un cadre flexible de représentation des situations comportementales. CBR*Tools permet en effet de représenter à la fois des situations explicites dans les cas concrets et abstraits, et des situations non-explicites dans les cas potentiels à travers les patrons de situations comportementales. La structure des situations peut être figée si l'on utilise un seul

patron pour tous les cas, ou dynamique si la situation de chaque cas est décrite plus précisément par la sélection de fenêtres temporelles, d'événements et par la représentation de contraintes temporelles.

– Pour la remémoration, nous proposons une stratégie de recherche ouverte basée sur la remémoration des cas abstraits, concrets et potentiels. La représentation de cette stratégie repose sur les index simples et composites qui permettent l'ajout et la modification de certaines étapes.

Ainsi, l'originalité de notre modèle réside dans l'unification de certaines caractéristiques des systèmes existants (notamment les modes d'observation, les types de représentation des situations, les étapes de remémoration), et dans la proposition d'extensions : observation hybride et compression des comportements, représentation de contraintes temporelles dans les situations, stratégie de remémoration ouverte. De plus, nous proposons une bibliothèque objet générique en RàPC, CBR*Tools, qui implante ce modèle.

Outre l'atelier de conduite de la nutrition des plantes que nous avons décrit, deux autres applications utilisant CBR*Tools sont en cours de réalisation : assistance à la navigation hypertextuelle, et aide à l'argumentation sur Internet dans un contexte de décision collective [KAR 97]. Ces applications ont également pour but d'illustrer et de valider l'approche générique de la conception de CBR*Tools. De plus, nous sommes actuellement à l'étude de plusieurs perspectives et extensions de nos travaux. Tout d'abord, nous pensons enrichir la représentation des chroniques en prenant en compte des occurrences vagues des événements et différentes échelles de temps (heures, jours, mois etc.) qui sont utiles dans le domaine médical [KER 95]. Nous envisageons également de proposer des index permettant une remémoration guidée par les contraintes temporelles en intégrant des travaux de reconnaissance de scénarios [GHA 89, FON 96, DOU 94]. Enfin, nous étudions des mécanismes d'apprentissage automatique permettant de mettre à jour la description de la situation de chaque cas. En effet, nous pensons que l'échec d'une remémoration (impossibilité de trouver des cas similaires ou remémoration de cas inadéquats) est dû entre autres à la prise en compte de données non pertinentes dans les chroniques, ou au contraire au manque de données pertinentes. Ainsi, l'objectif de l'apprentissage est de déterminer pour chaque cas, l'ensemble des données pertinentes dans la description d'une situation comportementale. L'utilisation de notre représentation permet une telle approche et devrait mener à un apprentissage plus sémantique que les méthodes d'apprentissage par renforcement utilisées dans SINS [RAM 93].

Remerciements

Les auteurs tiennent à remercier Michel Rueher, Professeur à l'Université de Nice Sophia-Antipolis, pour ses remarques constructives lors de la rédaction de ce rapport.

Références bibliographiques

[AAM 94] A. Aamodt and E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System. *AI Communications*, 7(1) :36–59, March 1994.

[ALL 84] J. F. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23:123–154, 1984.

[BAR71] M. Barbut and C. Fourgeaud. *Eléments d'analyse mathématique des chroniques*. Hachette, 1971.

[BER 96] R. Bergmann, W. Wilke. On the role of abstraction in case-based reasoning. In I. Smith and B. Faltings, editors, *Advances in Case-Based Reasoning*, volume 1168 of *Lecture Notes in Artificial Intelligence*, pages 28–43. Springer, 1996.

[BRA 91] R. Brandau, A. Lemmon, and C. Lafond. Experience with extended episodes: Cases with complex temporal structure. In *Case-Based Reasoning Workshop*, pages 1–12, Washington, D.C., May 1991.

[BRU 93] R. Brun, B. Paris, and I. Hammelin. Fertigation management of rose plants grown in greenhouse on rockwool. *Adv. Hort. Sci.*, 7:109–111, 1993.

[BUL 97] M. Bull, G. Kundt, and L. Gierl. Case-based risk detection and forecasting in a geographic-medical system. In R. Bergmann and M. Wilke, editors, *German Workshop on Case-Based Reasoning*, pages 59–64, March 1997.

[CAU 86] S. Cauvin *et al*. Surveillance et interprétation d'alarme en milieu industriel. In *Actes des 6e journées nationales PRC-GDR Intelligence artificielle*, pages 9–30, 1986.

[DET 94] R. Deter. CBR for maintenance of telecommunication network. In M. Keane, J.P. Haton, and M Manago, editors, *Second European Workshop on Case-Based Reasoning*, pages 23–31, Chantilly, November 1994.

[DOU 94] C. Dousson and M. Ghallab. Suivi et reconnaissance de chroniques. *Revue d'intelligence artificielle*, 8(1):29–61, 1994.

[FON 96] D. Fontaine. Une approche par graphes pour la reconnaissance de scénarios temporels. *Revue d'intelligence artificielle*, 10(4):439–468, 1996.

[FUC 95] B. Fuch, A. Mille, and B. Chiron. Operator decision aiding by adaptation of supervision strategies. In M.Veloso, K.D.

Althoff, and M. M. Richter, editors, *Case-Based Reasoning Research and Development*, volume 1010 of *Lecture Notes in Artificial Intelligence*, pages 23–32. Springer-Verlag, 1995.

[GAR 91] Philippe Garnesson and Gérard Giraudon. Polygonal approximation : Overview and perspective. Technical Report 1621, INRIA, June 1991.

[GHA 89] M. Ghallab and A. Mounir-Alaoui. Managing efficiently temporal relations through indexed spanning tree. In *Proc. 11th IJCAI*, pages 1297–1303, 1989.

[HAM 86] K. Hammond. CHEF: A model of case-based planning. *In Proceedings of AAAI-86*, pages 267–271, 1986.

[HEL 91] G. Held, and T.R. Marshall. *Data compression*. J. Wiley and Sons, 1991.

[HIN 89] T.R. Hinrichs. Strategies for adaptation and recovery in a design problem solver. In *Proceedings of the DARPA Workshop on Case-Based Reasoning*, Morgan Kaufmann, 1989.

[JAC 94] M. Jaczynski et B. Trousse. Fuzzy logic for the retrieval step of a case-based reasoner. In M. Keane, J.P. Haton, M. Manago, editors, *EWCBR-94 : Second European Workshop on Case-Based Reasoning*, pages 313-321, Chantilly, France, 1994.

[JAK 90] F. Jakob and P. Suslenschi. Situation assessment for process control. *IEEE Expert*, 5(2):49–59, 1990.

[KAR 97] N.I. Karacapilidis, B. Trousse and D. Papadias. Using Case-Based Reasoning for Argumentation with Multiple Viewpoints. In *Proceedings of the 2nd International Conference on Case-Based Reasoning (ICCBR'97)*, , volume 1266 of *Lecture Notes in AI*. Springer-Verlag, 1997.

[KER 95] E. T. Keravnou. Modelling medical concepts as time-objects. In M. Stefanelli and Jeremy Wyatt, editors, *Artificial Intelligence in Medicine*, volume 934 of *Lecture Notes in Artificial Intelligence*, pages 67–90. Springer-Verlag, 1995.

[KOL 89] J. Kolodner and R. L. Simpson. The MEDIATOR: Analysis of an early case-based problem solver. *Congnitive Science*, 13(4) :507–549, 1989.

[KOL 93] J. Kolodner. *Case-Based Reasoning*. Morgan Kaufmann, 1993.

[KOP 88] L. Kopeikina, R. Brandau, and A. Lemmon. Case based reasoning for continous control. In J. Kolodner, editor, *Proceedings of a Workshop on Case-Based Reasoning*, pages 250–259, Clearwater Beach, Florida, May 1988.

[KOT 89] P. Koton. *Using experience in learning and problem solving*. PhD thesis, Departement of Computer Science, MIT, 1989.

[MOO 92] K. Moorman and A. Ram. A case-based approach to reactive control for autonomous robot. In *Proceedings of the AAAI Fall Symposium on AI for Real-World Autonomous Mobile Robots*, Cambridge, October 1992.

[MUL 97] P.A. Muller. *Modélisation objet avec UML*. Eyrolles, 1997.

[NAK 94] G. Nakhaeizadeh. Learning prediction from time series: a theoretical and empirical comparison of CBR with some other approaches. In *Topics in Case-Based Reasoning*, volume 837 of *Lecture Notes in Artificial Intelligence*, pages 65–76. Springer, 1994.

[RAM 93] A. Ram and J.C. Santamaria. Continous case-based reasoning. In *AAAI Case-Based Reasoning Worksop*, pages 86–93, 1993.

[RIT 94] F. E. Ritter and J. H. Larkin. Developing process models as summaries of HCI action sequences. *Human -Computer Interaction*, 9:345–383, 1994.

[ROU 94] S. Rougegrez. *Prédiction de processus à partir de comportement observés : le système REBECAS*. Thèse de Doctorat, Institut Blaise Pascal, 1994.

[SCH 96] R. Schmidt, B. Heindl, B. Pollwein, and L. Gierl. Abstraction of data and time for multiparametric time course prognoses. In I. Smith and B. Faltings, editors, *Advances in Case-Based Reasoning*, volume 1168 of *Lecture Notes in Artificial Intelligence*, pages 377–391. Springer Verlag, 1996.

[WES 94] S. Wess, K.D. Althoff, et G. Derwand. Using K-d Trees to improve the retrieval step in case-based reasoning. In S. Wess, K.D. Althoff et M. M. Richter, editors, *Lecture Notes in Artificial Intelligence, Topics in Case-Based Reasoning*, pages 167-181, Springer-Verlag, 1994.

[WIL 86] B. C. Williams. Doing time: putting qualitative reasoning on firmer ground. In *Proceedings of the AAAI Natianal Conference on Artificial Intelligence*, pages 105–112, 1986.

Annexe

Dans la Figure 12, nous donnons les éléments de la notation UML (*Unified Modeling Language*) [MUL 97] nécessaires pour la compréhension des diagrammes présentés dans cet article. La notation UML est une notation conçue pour la modélisation objet d'applications et fait suite notamment aux notations des méthodes OMT et Booch.

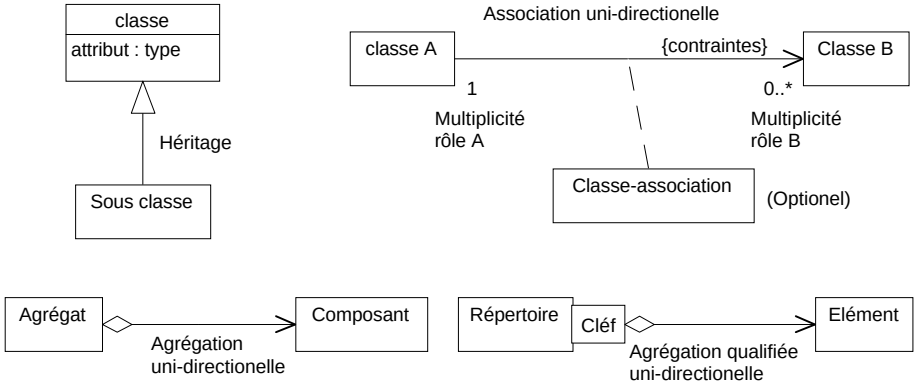

Figure 12 : éléments utilisés de la notation UML

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique
615 rue du jardin botanique, BP 101, 54600 VILLERS-LES-NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Editeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, LE CHESNAY Cedex (France)
ISSN 0249-6399