

# Approche hybride pour le diagnostic industriel basée sur le RàPC et le Datamining : Utilisation de la plateforme JCOLIBRI 2.1

Noureddine Mekroud<sup>1</sup>, Abdelouahab Moussaoui<sup>2</sup>

Département de l'informatique, Faculté des Sciences de l'Ingénieur, Université Ferhat Abbas, Sétif  
{Mekroud\_n, Moussaoui\_abdel}@yahoo.fr

**Résumé.** Le Raisonnement à Partir de Cas (RàPC) est un réflexe puissant et très naturel, qui vise la réutilisation des expériences passées dans la résolution des nouveaux problèmes ; ceci est confirmé par des expériences en psychologie et en sciences cognitives. Le RàPC, comme méthodologie d'ingénierie des connaissances, peut être renforcé dans les différentes étapes de son processus par la richesse des techniques du Datamining. Dans cet article, on propose une solution hybride, RàPC et Datamining, appliquée au domaine du diagnostic industriel. Le processus proposé commence par une fragmentation de la base des cas en deux espaces : Symptomes-Pannes & Symptomes-Solutions ; suivie d'un clustering des deux espaces, et d'un mappage entre leurs clusters ; on appliquera enfin un cycle RàPC pour chaque espace. La plateforme JCOLIBRI 2.1 sera utilisée pour l'implémentation de notre démarche. Les avantages de cette approche seront présentés, ainsi que des réflexes et perspectives diverses.

**Mots Clés:** ingénierie des connaissances, raisonnement à partir de cas, Datamining, diagnostic industriel, JCOLIBRI.

## 1 Introduction

L'application du Retour d'Expérience est fréquente dans la vie quotidienne de chacun, puisqu'il est bien évident et légitime que les problèmes similaires auront des solutions similaires, et qu'on se retrouve souvent face à un problème que l'on a déjà rencontré.

De nos jours, le savoir faire est le patrimoine principal des individus et établissements ; la valorisation des entreprises repose de plus en plus sur des facteurs immatériels. Selon une étude sur des centaines de sociétés industrielles américaines, l'estimation de la valeur de leur patrimoine de connaissances (Knowledge Capital) correspond à 217% de leur capital financier net [9]. La préservation et la capitalisation du savoir, du savoir faire et des meilleures pratiques existantes dans une entreprise, nécessite la mise en place d'un système permettant de fournir à une personne, pas nécessairement hautement qualifiée, l'information utile au moment où elle en a besoin, de façon exploitable pour superviser les décisions à prendre.

Les connaissances d'une entreprise, considérées comme un patrimoine fragile, circulent et s'enrichissent et sont exploitées plus ou moins avec fiabilité, mais peuvent également disparaître par un départ en retraite, une mutation ou un licenciement d'un expert. Cette richesse d'expertise, qui doit être capturée, capitalisée, protégée et distribuée, forme la partie principale d'une mémoire d'entreprise, qui rapproche le niveau de connaissances individuel au niveau d'expérience collective de l'organisme.

Le RàPC, comme une méthodologie de capitalisation des connaissances, propose des solutions aux problèmes actuels à résoudre, en utilisant les connaissances acquises des expériences passées, et en enrichissant en continu la base de connaissances. Cette méthodologie, qui forme un point de rencontre entre l'intelligence artificielle et les sciences cognitives ; repose sur des notions beaucoup plus théoriques que techniques, ce qui lui donne l'aspect d'une méthodologie et pas une technologie. Le Datamining, utilisé dans la découverte et la modélisation des informations utiles, cachées dans une masse de données grande et complexe, offre des solutions techniques incontournables dans la découverte et la capitalisation des connaissances de l'entreprise.

L'objectif de ce travail est d'étudier l'utilisation des techniques du Datamining, qui émergent actuellement, dans le cycle du processus RàPC, en éclairant l'appui qu'elles peuvent fournir pour améliorer la fiabilité de cette méthodologie. Visant le domaine du diagnostic industriel, notre approche est basée sur la mise en cause de la liaison directe et stable entre les pannes de leurs solutions possibles. On propose dans cet article la fragmentation verticale de la base des connaissances du domaine étudié en deux partitions : l'espace des symptômes et leurs pannes correspondantes, et l'espace des symptômes et leurs solutions possibles ; suivi d'un clustering du contenu de chaque espace, et d'un mappage entre les clusters des pannes et de leurs solutions possibles ; et on finira par l'application du cycle RàPC sur les deux partitions conjointement. L'approche proposée offre non seulement la possibilité de réutilisation des expériences passées en limitant la recherche aux segments les plus pertinents, mais aussi de traiter le cas où les mêmes symptômes donneront plusieurs pannes possibles, ce qui indique l'insuffisance de ces symptômes pour diagnostiquer les pannes. Aussi, si les mêmes symptômes auront plusieurs solutions possibles, on pourra choisir entre ces solutions suivant la stratégie de maintenance, corrective ou préventive. D'autres avantages et arguments pertinents seront fournis en ce qui suit.

Dans la section 2, on propose une présentation générale de la gestion des connaissances dans les entreprises. La section 3 étudie la méthodologie du raisonnement à partir de cas. Dans la section 4, on présente un état de l'art des travaux réalisés concernant l'intégration des techniques du Datamining dans les différentes étapes du cycle RàPC. La section 5 expliquera notre démarche, combinant quelques algorithmes de Datamining avec le processus RàPC. Dans la section 6, la plateforme JCOLIBI 2.1 sera utilisée dans l'implémentation et la validation de la solution proposée. À la fin, une conclusion et des perspectives seront proposées.

## **2 Les connaissances dans l'entreprise**

L'un des aspects de développement de l'efficacité d'une entreprise est la capitalisation de ses expériences développées au cours du temps, vu le risque de

centralisation du savoir faire dans les experts humains, qui peut causer la non disponibilité de cette richesse suite à une sur-occupation de l'expert, mutation, départ à la retraite, démission, licenciement ... [7].

Les connaissances sont une représentation réduite du monde réel [8]. Dans une entreprise, deux catégories de connaissances sont à distinguer, à savoir : les connaissances *Tacites* (implicites, non formalisables) : qui sont difficiles à décrire, comme les compétences, les habilités, la connaissance historique de l'organisation ...etc, c'est le savoir faire de l'entreprise ; et les connaissances *Explicites* (formalisables) : qui sont plus facilement codifiables (manuels, plans, modèles, documents d'analyse, données, ...), c'est le savoir de l'entreprise. Par exemple, avant l'invention du système de notation (solfège) au XII<sup>ème</sup> siècle, on ne pouvait pas représenter les connaissances musicales ; elles étaient apprises uniquement par l'expérience directe, l'écoute (tacites) ; mais grâce au solfège, elles sont devenues codifiables, donc explicites [5].

## 2.1 La gestion des connaissances

Dans la littérature de la gestion moderne des entreprises, de nouvelles notions sont évoquées, comme : les travailleurs de la connaissance, la société et l'économie de la connaissance, la mémoire d'entreprise, les entreprises apprenantes, la gestion des richesses immatérielles, la gestion de l'innovation...etc., où la richesse de l'entreprise est désormais basée sur ses activités intellectuelles, donnant naissance à une économie basée sur la connaissance [8]. La gestion des connaissances, ou Knowledge Management en anglais (baptisée KM), désigne l'ensemble de concepts et outils permettant la production des connaissances et le développement des compétences individuelles, collectives et organisationnelles. Le KM vise à rassembler le savoir et le savoir faire sur des supports accessibles, et faciliter leur transmission en temps réel à l'intérieur de l'établissement, ou les différer à nos successeurs [16]. C'est une application pratique des sciences cognitives, de l'intelligence artificielle et des sciences de l'organisation [15].

Le cycle de capitalisation des connaissances est composé de quatre phases : la première consiste à « repérer » les connaissances cruciales de l'entreprise ; Ensuite, il faut les « préserver », c'est-à-dire les modéliser, les formaliser et les conserver ; Puis, les « valoriser » et les mettre en exploitation en les rendant accessibles ; Après, les « actualiser » et les enrichir progressivement en les gérant continuellement [15].

Il existe une diversité de méthodologies de capitalisation des connaissances d'une entreprise, on peut distinguer celles spécifiques à la construction des mémoires d'entreprises, à savoir : REX, MEREX, CYGMA, atelier FX et Componential Framework ... etc ; et d'autres empruntées du domaine de l'ingénierie des connaissances, comme : KADS, CommonKADS, MKSM, MASK, et KOD [16].

## 3 Le Raisonnement à Partir de Cas

Le RàPC est un processus qui vise la réutilisation des expériences passées. Cette méthodologie, provenant du domaine de l'Intelligence Artificielle, a été utilisée dans

les systèmes experts et les sciences cognitives. Dans cette approche, l'utilisateur essaie de résoudre un nouveau problème en reconnaissant les similarités avec des problèmes préalablement résolus, appelés : cas. Un cas est communément un problème spécifique qui a été identifié, résolu, stocké et indexé dans une mémoire avec sa solution, et éventuellement le processus d'obtention de celle-ci [5]. Les systèmes de RàPC sont appliqués dans de nombreux domaines comme : la médecine, le commerce, le diagnostic industriel, le contrôle et l'analyse financière [16].

### 3.1 Le cycle du RàPC

Dans l'approche du RàPC, la résolution des problèmes est basée sur la réutilisation par analogie des expériences passées. Le cycle du RàPC se décompose habituellement en cinq phases principales [16] :

1. L'élaboration d'un nouveau problème (cas cible) : représente l'acquisition et la modélisation des informations connues sur le nouveau problème, pour lui donner une description initiale, d'une manière similaire aux cas existants dans la base des cas.
2. La remémoration des cas (cas sources) : rechercher les cas les plus similaires, cela signifie la recherche des correspondances entre les descripteurs des cas de la base et ceux du cas actuel à résoudre.
3. L'adaptation des cas (cas source) : réutiliser totalement ou partiellement la solution du cas trouvé le plus similaire, pour résoudre le nouveau problème.
4. La révision de la solution proposée (solution cible) : signifie l'évaluation de la solution proposée, par son application dans le monde réel.
5. La mémorisation d'un nouveau cas (cas cible) : représente l'ajout éventuel du cas cible dans la base des cas, c'est l'apprentissage d'un nouveau cas, qui pourra ainsi être utilisé pour la résolution des problèmes futurs.

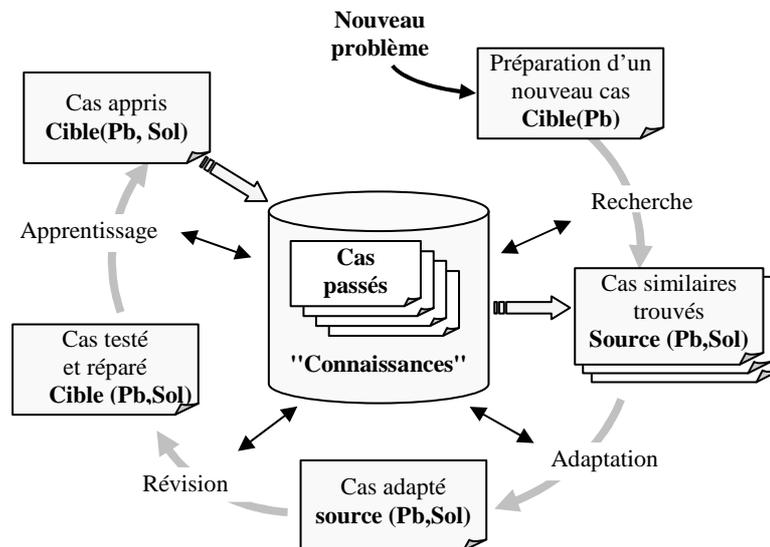


Fig 1 : Les étapes du cycle de RàPC [15]

### 3.2 Avantages et inconvénients du RàPC

Puisque cette approche n'a besoin que d'un ensemble de problèmes résolus pour commencer, le RàPC diffère des autres approches de l'IA, qui visent la représentation globale des connaissances utilisées dans un domaine, sous forme de système expert.

Ainsi, l'approche RàPC a l'avantage d'être une démarche plus simple à mettre en oeuvre que celles basées sur un modèle de domaine, puisque elle permet d'éviter les difficultés de modélisation du savoir-faire des experts (complexité des ontologies et des représentations logique...). Aussi, c'est un bon choix pour les domaines n'exigeant pas de solutions optimales, ou dont les principes sont mal formalisés ou peu éprouvés [16]. Le RàPC vise l'utilisation des connaissances spécifiques et pragmatiques, qui concernent les problèmes précédemment expérimentés, en les capitalisant d'une façon progressive avec le temps ; l'apprentissage sera ainsi incrémental et basé sur les expériences vécues [5].

Mais, par contre, le RàPC ne trouve pas nécessairement la solution concrète à un problème ; et parfois, juste proposer un ensemble de solutions possibles [6]. Aussi, vu la nécessité d'une intervention et d'une mobilisation en continu des experts, lors de la capitalisation progressive des connaissances, l'individualisme constitue un frein redoutable ; les experts hésitent de partager leurs connaissances, acquises après des années de travail, par méfiance d'une restructuration ou compression dans l'entreprise, ou ils estiment leurs savoir-faire comme leur plus grande assurance [7].

Un aspect, qu'on juge très intéressant, est que le RàPC est une méthodologie et pas une technologie de résolution des problèmes, et la majorité de ses définitions se basent sur la présentation du « Quoi » et pas le « Comment ». Ainsi, présenter le RàPC comme une méthodologie est important pour son développement. Si le RàPC est considéré comme une technologie, il pourrait sembler que les recherches dans ce domaine sont en grande partie achevées. Mais, considéré comme une méthodologie, les chercheurs auront toujours le défi de perfectionner le cycle de ce processus [17].

## 4 Quelques contributions d'intégration RàPC & Datamining

Le Datamining est l'extraction des informations utiles, et la découverte des connaissances et des modèles cachés dans des bases de données volumineuses et hétérogènes [10]. Dans la littérature, plusieurs contributions visant la combinaison des techniques du Datamining avec le processus du RàPC ont été proposées :

Clerkin et al. (2002) ont présenté une approche de génération automatique de la base des cas ; en segmentant, par l'algorithme des K-means, les bases de données contenant l'historique des opérations effectuées dans le domaine étudié.

Pan et Al (2007) ont étudié deux problèmes majeurs dans les algorithmes de recherche des cas similaires ; le premier est causé par les cas bruyants (un cas proche du problème à résoudre qui donne des solutions incorrectes pour ce dernier) et le deuxième problème est causé par l'inégalité de la distribution des cas (les problèmes similaires qui ont des solutions différentes). Les auteurs ont utilisé une analyse discriminante pour éliminer les cas bruyants, suivie d'une évaluation empirique et extensive des cas résultants, pour minimiser la marge d'erreur dans la recherche.

Kim et Han (2001) ont proposé une nouvelle méthode d'indexation et de classification de la base des cas, basée sur l'utilisation des réseaux neurones artificiels compétitifs, pour générer les valeurs centroïdes, qui seront ajoutées à la base sous forme de cas artificiels représentatifs, utilisés ensuite pour indexer la base des cas, ce qui facilitera la recherche des cas similaires au cas actuel à résoudre.

Zhuang et Al (2007) ont proposé la segmentation des bases de cas volumineuses en utilisant les cartes de Kohonen pour l'auto-organisation (SOM : Self-Organising Maps). Leur but était de trouver des groupements de cas avec un maximum de similitude intra-groupes, et de désimilitude inter-groupes, pour bien préparer la phase de remémoration, en limitant la recherche aux segments les plus pertinents.

À cause des incertitudes dans la représentation des connaissances et la description des attributs et des mesures de similarité, Chang et Al (2008) ont proposé d'incorporer la logique floue dans le processus du RàPC, pour avoir un modèle de prévision des ventes avec plus de souplesse et de précision ; cela en sélectionnant les ventes antérieures les plus similaires et les plus utiles au cas actuel.

N. Arshadi et Al (2000) ont proposé une nouvelle approche pour l'adaptation compositionnelle. Ils incitent la possibilité de combiner les solutions trouvées, pour permettre l'obtention d'une solution finale plus pertinente. Les solutions similaires peuvent être combinées par le choix de la valeur moyenne de leurs attributs. Une telle approche enrichisse la base par des cas généralisés plus ou moins précis.

Yang et Wu (2000) ont proposé une méthode pour simplifier la maintenance des bases de cas volumineuses, en les segmentant par l'algorithme du GDBSCAN, basé sur le regroupement par régions à densité plus élevée par rapport à leur entourage. Cette segmentation donnera, comme résultat, plusieurs bases de cas plus petites en taille ; ainsi, la maintenance de ces bases sera transparente et sans complexité.

## **5 Notre approche : Hybridation RàPC-Datamining**

### **5.1 Présentation générale**

Le RàPC est basé sur deux hypothèses concernant la nature du monde réel [1]: la *régularité* des situations, ou problèmes, similaires qui doit impliquer des solutions similaires ; et la *réurrence* qui assume que c'est fortement probable que les situations futures seront des variantes des situations actuelles. Ainsi, les approches standard du RàPC proposent un mappage entre le problème cible (actuel) et le problème source (déjà résolu), basé sur la similarité ; et la solution du cas similaire trouvé sera, suivant le raisonnement par analogie, la solution proposée au problème cible actuel. Mais, les contraintes internes et l'environnement extérieur changent au fil du temps. Ainsi, la liaison entre les descripteurs des cas et leurs solutions peut ne pas être stable, ce qui donne la possibilité d'avoir des expériences erronées [12]. Avec des systèmes RàPC utilisés à longs termes, on pourra être face à des situations dans lesquelles la réutilisation des expériences deviendra une erreur. Démarrant de ces hypothèses de base, notre approche propose un cycle de RàPC fondé sur des similarités appliquées dans deux espaces différents : espace des pannes et espace des solutions (Fig 2).

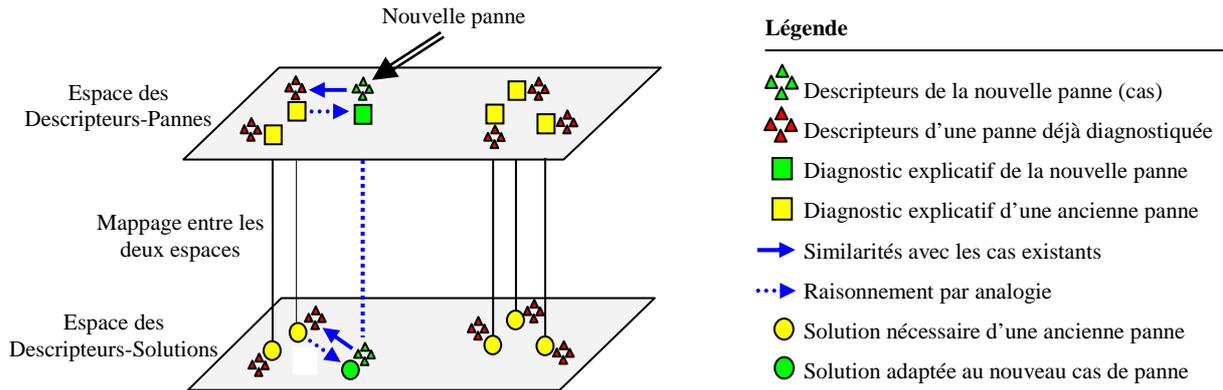


Fig 2 : Tracé de résolution des problèmes proposé par notre approche

Notre démarche commence par la fragmentation verticale de la base des cas en deux fragments : Descripteurs-Pannes & Descripteurs-Solutions, suivie d'un clustering de chacun des deux espaces, et d'un mappage (via des mesures de similarité globales) entre les clusters de ces deux fragments. Ensuite, et lors d'une nouvelle panne, on prépare les descripteurs de ce nouveau cas à résoudre ; et on applique dans l'espace des Descripteurs-Pannes, dans le but de trouver un diagnostic adéquat à cette panne, un cycle RàPC sur le cluster le plus similaire au cas actuel, en utilisant des mesures de similarité locales adéquates à chaque descripteur de panne, ainsi que des similarités basées sur les ontologies du domaine. Après, on mappe vers le cluster pertinent dans l'espace des solutions, où on recherche, suivant un autre cycle RàPC, la solution adéquate au cas de panne actuel à résoudre.

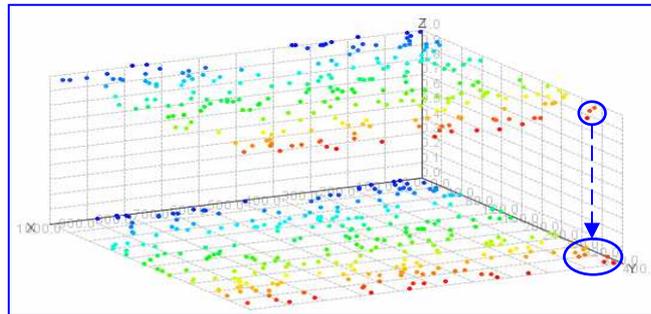


Fig 3 : Exemple de mappage entre deux espaces segmentés

Le schéma ci-dessous montre les attributs : descripteurs, pannes et solutions de notre domaine d'étude, partitionnés ensuite en deux espaces.

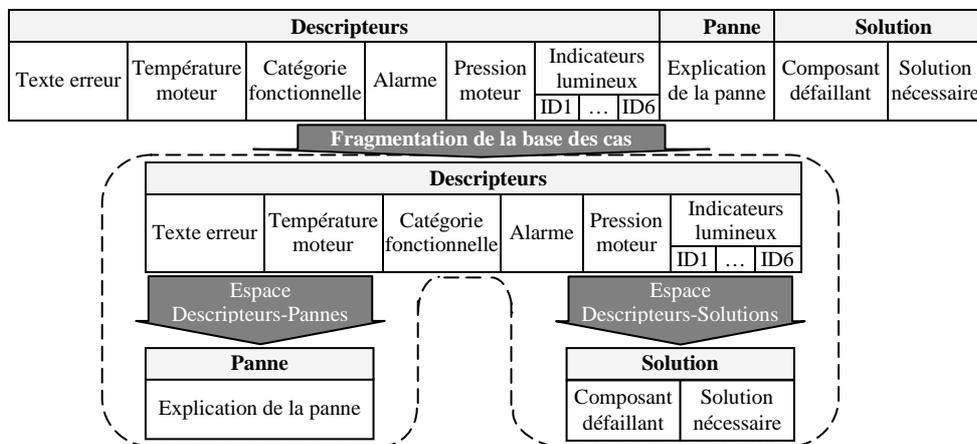


Fig 4 : Les attributs proposés de la base des cas, réparties en deux espaces

## 5.2 Les étapes du processus proposé

La figure suivante illustre la succession des étapes de notre approche :

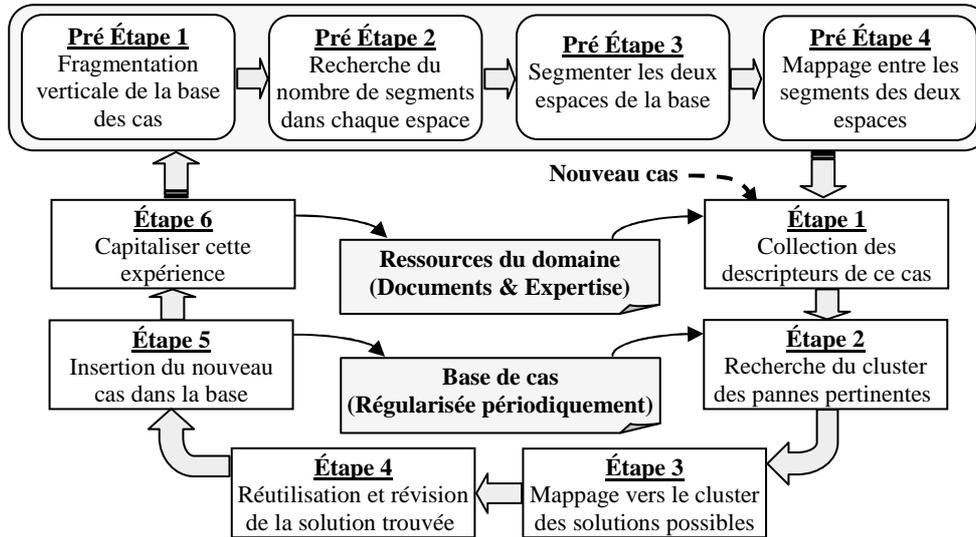


Fig 5 : Processus d'intégration des techniques du Datamining dans le cycle RàPC.

**Pré-Étape 1 :** Fragmenter la base des cas verticalement, en Descripteurs-Pannes et Descripteurs-Solutions, et isoler les causes des pannes de leurs solutions possibles.

**Pré-Étape 2 :** Rechercher le nombre de segments optimal dans chaque espace, par la segmentation hiérarchique ascendante (SHA), pour combler le point faible principal des K-Means qui est la fixation préliminaire de ce nombre.

**Pré-Étape 3 :** Segmenter les deux partitions de la base par un algorithme de K-Means. Des significations et utilités seront fournies lors d'une valeur différente du nombre de clusters dans les deux espaces (voir 5.3).

**Pré-Étape 4 :** Faire un tracé (mappage) entre les clusters Symptômes-Pannes et Symptômes-Solutions, suivant des mesures de similarités globales, pour lier les pannes à leurs solutions les plus pertinentes.

**Étape 1 :** Collection des descripteurs du nouveau cas de panne.

**Étape 2 :** Recherche du cluster des pannes pertinentes, suivant les Descripteurs-Pannes ayant la plus grande similarité avec les caractéristiques du nouveau cas. La recherche est basée sur l'algorithme des K Plus Proches Voisins (KPPV).

**Étape 3 :** Faire un mappage vers le (les) cluster(s) Descripteurs-Solutions correspondant(s) au cluster des pannes trouvé dans l'étape précédente.

**Étape 4 :** Réutiliser la solution, ayant les descripteurs les plus adéquats au cas actuel, dans le contexte du nouveau cas. Réviser et commenter les résultats trouvés.

**Étape 5 :** Maintenance de la base par l'insertion du nouveau cas résolu.

**Étape 6 :** Capitaliser cette expérience par la conservation des commentaires et des évaluations fournis, suivie d'une nouvelle segmentation des deux espaces de la base.

### 5.3 Avantages et inconvénients de cette approche

1. On a opté pour la mise en cause de la liaison directe entre les pannes et leurs solutions possibles, par l'isolation des descripteurs et leurs pannes correspondantes des descripteurs et leurs solutions adéquates, pour bénéficier des avantages suivants:
  - ☞ Détection de l'insuffisance des descripteurs : puisque si les mêmes descripteurs donneront plusieurs pannes possibles et/ou plusieurs solutions possibles, alors ces descripteurs ne sont pas suffisants et discriminants.
  - ☞ En se basant sur les algorithmes des règles d'associations utilisées en Datamining, on distingue que les relations de dépendance entre les champs de la base des cas sont plus fortes pour les relations entre les descripteurs et leurs pannes appropriées, d'un côté, et les descripteurs et leurs solutions possibles, de l'autre côté. Ces relations d'association aideront à mieux comprendre le comportement fonctionnel de l'équipement étudié.
2. Le mappage entre les pannes et leurs solutions possibles est très utile, puisque si on considère le mappage entre les segments des espaces Descripteurs-Pannes et Descripteurs-Solutions comme une relation par cardinalités (nombre d'occurrences possibles) dans les deux sens, cela est interprétable comme suite :
  - ☞ Les cardinalités dans le sens Descripteurs-Pannes vers Descripteurs-Solutions décrivent le nombre de solutions possibles pour les mêmes symptômes de panne, et ça donnera les interventions possibles sur cette panne dans les différentes stratégies de maintenance (préventive, corrective ...).
  - ☞ Les cardinalités dans le sens Descripteurs-Solutions vers Descripteurs-Pannes donneront le nombre de pannes qu'on peut résoudre par la même intervention, et ça aidera à planifier les interventions par priorité, selon le nombre de problèmes possibles à résoudre par chaque intervention.
  - ☞ Il est possible d'avoir une cardinalité nulle entre un segment des solutions et les segments de l'espace des pannes, c'est le cas d'un ensemble d'indicateurs (symptômes) qui prédisent une panne avant son arrivé, donc cela donnera la possibilité d'intervention par une maintenance préventive.
3. Les approches du K-Means et des KPPV ont une fiabilité reconnue, mais aussi une complexité faible (polynomiale), et cela facilitera la refragmentation de la base, qui est nécessaire après chaque modification de son contenu. La méthode SHA est très favorisée lors du choix du nombre de segments pour le K-Means.
4. L'utilisation de divers mesures de similarité, suivant le type et les valeurs de chaque champs, ainsi que la similarité basée sur les ontologies du domaine, renforcera l'efficacité de la recherche des cas similaires au problème à résoudre.

Mais, comme toute travail, notre approche présente quelques inconvénients :

1. Si la panne actuelle est mal classifiée, lors de la recherche du segment des Descripteurs-Pannes le plus pertinent, alors on aura un mappage vers un cluster de solutions non adéquates au problème à résoudre.
2. cette approche nécessite une bonne connaissance du domaine étudié, interprétée par une base de cas suffisamment grande, pour assurer la performance des résultats issus de l'application des algorithmes de segmentation et de mappage.

## 6 Implémentation

Pour étudier la faisabilité de notre démarche, nous avons développé un système de diagnostic industriel ; et comme équipement étudié, on a choisi une machine extrudeuse des tubes de plastique (*Extrudeuse Monovis BEX 1-90-30B* du fabricant allemand *BATTENFELD*). Notre implémentation est basée sur la conception d'un modèle du domaine, par la création de l'ontologie des composants de l'extrudeuse, et l'ontologie de ses catégories fonctionnelles, en utilisant l'éditeur d'ontologies Protégé et la technologie OWL (Web Ontology Language) qui permet la représentation des concepts et taxonomies du domaine, dans une hiérarchie exprimant les relations entre ces éléments [16]. Aussi, on a opté pour la personnalisation de la plateforme RàPC Open Source JCOLIBRI 2.1, suivant les caractéristiques de notre domaine d'étude. Le modèle du domaine est ensuite associé au modèle des connaissances du RàPC, formé d'une base de cas qui modélise l'historique des interventions. Cette association assurera une bonne représentation des descripteurs des cas, ce qui facilitera la recherche des cas similaires les plus pertinents, en utilisant des mesures de similarité diverses, basées sur les symptômes des pannes et sur les ontologies du domaine.

La figure suivante, projetée de notre implémentation, illustre la première phase du cycle de RàPC (élaboration d'un nouveau cas), en utilisant les descripteurs de panne et l'ontologie des composants de la machine à étudier, réalisée par l'éditeur Protégé, pour préparer une requête de recherche des cas similaires les plus proches.

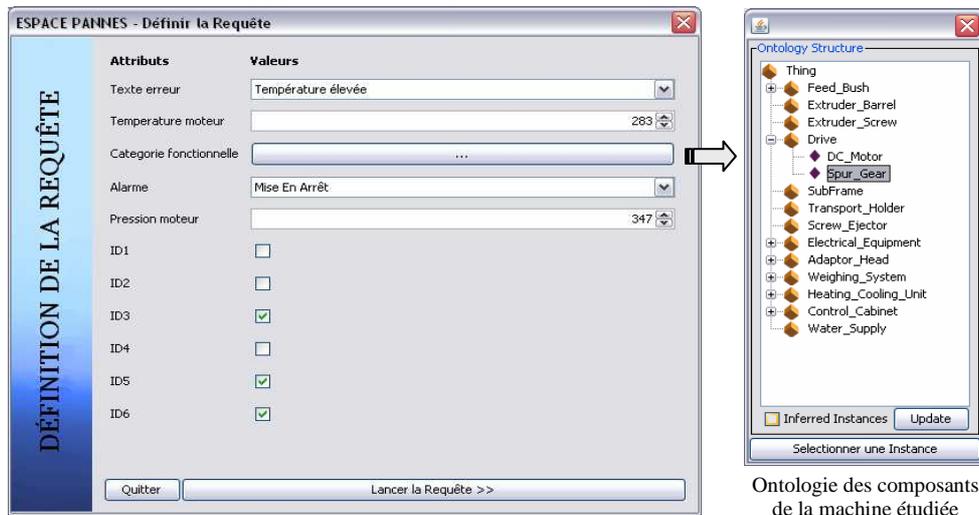


Fig 6 : Préparer une requête recherchant les cas similaires au problème à résoudre.

Après, le cycle RàPC recherche dans l'espace des Descripteurs-Pannes les pannes les plus similaires aux descripteurs de panne actuels. Et après avoir adapter et réviser le cas trouvé le plus pertinent, le nouveau cas résolu sera ajouté à la base des cas, bien sur si il n'existé par auparavant (ie : similarité moins de 100% par rapport au cas trouvé). À la fin, le processus continua par un deuxième cycle RàPC sur l'espace des Descripteurs-Solutions. Un nouveau clustering de la base des cas peut être nécessaire.

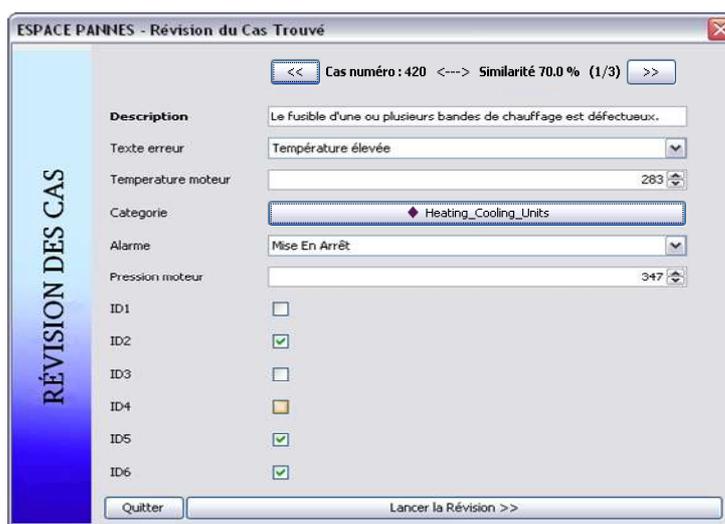


Fig 7 : Révision et adaptation du cas similaire trouvé

## 7 Conclusion

Le RàPC a apporté un courant d'air frais, et un certain degré d'optimisme au secteur de l'intelligence artificielle, d'une manière générale, et aux systèmes à base de connaissances en particulier. Poussé par l'intérêt de capitalisation du savoir-faire des entreprises, le renforcement du RàPC par les techniques du Datamining améliorera la fiabilité de cette méthodologie. Après avoir présenter un état de l'art sur la gestion des connaissances dans l'entreprise, et quelques travaux réalisés concernant l'intégration des solutions dans le processus du RàPC, on a proposé dans cet article une démarche pour l'intégration de plusieurs techniques, empruntées du Datamining, dans le processus du RàPC, surtout dans l'étape de recherche des cas similaires, par l'utilisation de diverses mesures de similarité locales, adaptées au type et à la modalité de chaque champs. Aussi, exploiter les ontologies du domaine d'étude dans la recherche, en utilisant des mesures de similarité basées sur les concepts, influa sans doute sur la pertinence des résultats de la recherche. On a renforcé notre approche par la séparation les pannes de leurs solutions possibles, suivie d'une étude sur ses avantages, et les connaissances qu'on peut extraire par les différentes formes de mappage entre les clusters de l'espace des pannes et ceux de l'espace des solutions.

Notre démarche a montrée des prérogatives claires, elle aidera à mieux comprendre le comportement fonctionnement des équipements, et facilite la planification des interventions de maintenance, préventives et correctives ; ainsi qu'une exploitation optimale de la base des connaissances, avec plus d'interactivité et de visibilité.

Les perspectives du couplage de la puissance et l'efficacité des solutions du Datamining avec la maturité du RàPC nous laisse estimer, dans l'avenir, des systèmes de retour d'expérience bien fondés, soit du coté méthodologique, ou de la qualité des outils d'implémentation utilisés. Dans notre domaine de diagnostic industriel, et parmi plusieurs perspectives possibles, les règles d'associations pourront aider à extraire les dépendances entre les symptômes des pannes ; où un symptôme pourra être le résultat d'un autre. Cela nous aidera à analyser l'enchaînement des défaillances, pour une meilleure conception des ontologies de l'équipement étudié.

## Références

- [1] Abasolo, J.M.: Towards a Component-based Platform for Developing Case-based Reasoning Systems. Projet de thèse doctorale en intelligence artificielle, Université polytechnique de Catalunya, Espagne (2004).
- [2] Arshadi, N., Badie, K.: A Compositional Approach to Solution Adaptation in Case-Based Reasoning and its Application to Tutoring Library. In: Proceedings of 8 th German Workshop on Case-Based Reasoning, Lammerbuckel (2000).
- [3] Chang, P.C., Liu, C.H., Lai, R.K.: A fuzzy case-based reasoning model for sales forecasting in print circuit board industries. In: Expert Systems with Applications 34, 2049–2058 (2008).
- [4] Clerkin, P., Hayes, C., Cunningham, P.: Automated case generation for recommender systems using knowledge discovery techniques. Technical Report, Computer Science Department, Trinity College Dublin, Ireland (2002).
- [5] Cortes Robles, G.: Management de l'innovation technologique et des connaissances : Synergie entre la théorie TRIZ et le Raisonnement à Partir de Cas, Application en génie des procédés et systèmes industriels. Thèse doctorale, Institut National Polytechnique de Toulouse (2006).
- [6] Devèze, B., Fouquin, M.: Case-Based Reasoning. Rapport des études en spécialité SCIA, école de l'ingénieur EPTA, France (2004).
- [7] Dieng-Kuntz, R., Corby, O., Gandon, F., Giboin, A., Golebiowska, J., matta, N., Ribière, M.: Méthodes et outils pour la gestion des connaissances : une approche pluridisciplinaire du Knowledge management. Dunod, 2<sup>ème</sup> édition, Paris (2001).
- [8] Duizabo, S., Guillaume, N.: Les problématiques de gestion des connaissances dans les entreprises. Centre de recherches DMSP (Dauphine Marketing Stratégie Prospectiv), Université Paris Dauphine. France (1997).
- [9] Ermine, J.L.: Enjeux, démarches et processus de la gestion des connaissances. Support de cours, université de trier, France (2005).
- [10] Gilleron, R., Tommasi, M.: Découverte de connaissance à partir de données. Technical report, Grappa, Université Lille 3, France (2000).
- [11] Kim, K.S., Han, I.: The cluster-indexing method for case-based reasoning using self-organizing maps and learning vector quantization for bond rating cases. In: Expert systems with Application 21, 147-156 (2001).
- [12] Leake, D.B., Wilson, D.C.: When Experience is Wrong : Examining CBR for Changing Tasks and Environments?. Computer Science Department, Indiana University, U.S.A (1999).
- [13] Liu, D.R., Ke, C.K.: Knowledge support for problem-solving in a production process: A hybrid of knowledge discovery and case-based reasoning. In: Expert Systems with Applications (2006).
- [14] Pan, R., Yang, Q., Pan, S.J.: Mining competent case bases for case-based reasoning. In: Artificial Intelligence 171, 1039–1068 (2007).
- [15] Rakoto, H.: Intégration du Retour d'Expérience dans les processus industriels, Application à Alstom Transport. Thèse doctorale. Institut National Polytechnique de Toulouse (2004).
- [16] Rasovska, I.: Contribution à une méthodologie de capitalisation des connaissances basée sur le raisonnement à partir de cas, Application au diagnostic dans une plateforme d'e-maintenance. Thèse doctorale, L'UFR des Sciences et Techniques de l'Université de Franche-Comté (2006).
- [17] Watson, I.: Case-based reasoning is a methodology not a technology. In: AI-CBR, University of Salford, United KingDoom (1999).
- [18] Yang, Q., Wu, J.: Keep it simple: A case-base maintenance policy based on clustering and information theory. In: Proceedings of the Canadian AI Conference, p. 102–114 (2000).
- [19] Zhuang, Z.Y., Churilov, L., Burstein, F., Sikaris, K.: Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. Revue Européenne de la recherche opérationnelle (2000).