

**VERS UNE CARTOGRAPHIE DES CONNAISSANCES GUIDÉE PAR FOUILLE DE
DONNÉES : 1^{ÈRE} ÉTAPE – MODELISATION BOOLEENNE.**

Brahami Menaouer,

Maître assistant à l'École Normale Supérieures de l'Enseignement Technologique - ENSET d'Oran

brahami.menaouer@enset-oran.dz , mbrahami@gmail.com, + 2137 79 30 84 42

Atmani Baghdad,

Maître de conférences à l'université d'Oran – département d'Informatique

atmani.baghdad@univ-oran.dz , atmani.baghdad@gmail.com + 213 6 64 45 82 51

Adresse professionnelle

ENSET d'Oran ★ BP 1523 ★ El M'naouer 31000 Oran, Algérie

Université d'Oran ★ BP 1524 ★ El M'naouer 31000 Oran, Algérie

Résumé : Nous présentons dans cet article la plateforme d'une nouvelle approche booléenne de cartographie des connaissances guidée par fouille de données pour le transfert des connaissances en vue de définir la stratégie de gestion à engager. La démarche que nous avons adoptée utilise une nouvelle méthode d'apprentissage artificiel cellulaire nommée CASI (Cellular Automata for Symbolic Induction). La première étape consiste en la réalisation d'une cartographie des domaines de connaissances biologiques réalisée en collaboration avec le Laboratoire de « *génétique et d'amélioration des plantes* » de l'université d'Oran. La deuxième étape concerne la construction d'un arbre de décision à partir de la cartographie des connaissances.

Le but, après une modélisation booléenne de la cartographie des domaines de connaissances, est double : d'une part proposer une plateforme de cartographie guidée par fouille de données, et, d'autre part réduire la complexité de stockage, ainsi que le temps de calcul. Seule la modélisation booléenne de la cartographie est décrite dans ce papier.

Mots clés : Cartographie des connaissances, ingénierie des connaissances, arbre de décision, machine cellulaire, biologie, génétique des plantes.

Summary : We present in this article the platform of a new boolean approach of mapping of knowledges guided by body search of data for the transfer of knowledge in order to define the strategy of management to be engaged. The step that we adopted uses a new method of cellular artificial training named CASI (*Cellular Automata for Symbolic Induction*). The first stage consists of the realization of a mapping of the domains of biologicals knowledges realized in collaboration with the laboratory of « *genetics and improvement of the plants* » of the university of Oran. The second stage relates to the construction of a decision tree from the mapping of knowledges.

The purpose, after a boolean modeling of the mapping of the domains of knowledges, is double: on the one hand to propose a platform of cartography guided by body search of data, and, on the other hand to reduce the complexity of storage, as well as the computing time. Only the boolean modeling of the mapping is described in this paper.

Key words : knowlege mapping, knowledge engineering, Decision tree, cellular machine, biology, genetic of the plants

<http://isdsm.univ-tln.fr>

VERS UNE CARTOGRAPHIE DES CONNAISSANCES GUIDÉE PAR FOUILLE DE DONNÉES : 1^{ÈRE} ÉTAPE – MODELISATION BOOLEENNE

Dans une perspective d'innovation, la gestion de connaissances (Prax, 2007) est une discipline en pleine expansion dans le monde des entreprises dont les objectifs visent la formalisation et le transfert des savoirs spécifiques à l'organisation, la capitalisation et l'exploitation de ces savoirs en vue d'améliorer la performance organisationnelle.

Plusieurs étapes ont été identifiées dans un processus de gestion de connaissances : il s'agit de l'explicitation de connaissances tacites repérées comme cruciales pour l'entreprise, du partage du capital des connaissances rendues explicite sous forme de mémoire, de l'appropriation et de l'exploitation d'une partie de ces connaissances par les acteurs de l'entreprise (Nonaka & Takeuchi, 1997) (Figure. 1).

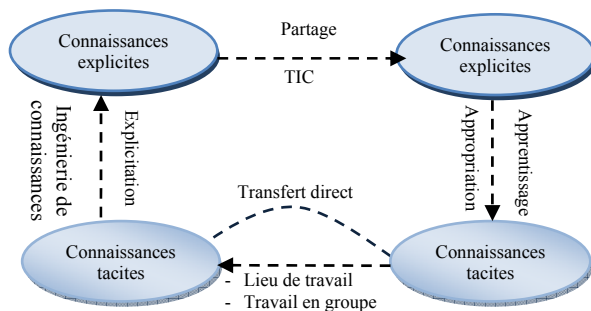


Figure.1 – Quatre modes de création et de transfert des connaissances

L'objectif prioritaire de la gestion des connaissances est de préserver le patrimoine stratégique des connaissances, développé au fil de son existence, mais qui reste peu explicite et demeure tacite chez les acteurs de métiers (Lavinia, 2006). A cet effet, il y a un certain nombre de méthodologies qui sont aujourd'hui disponibles pour formaliser ce patrimoine stratégique de connaissances. Nous nous intéressons dans ce papier à l'une de ces méthodes : la cartographie des connaissances.

Dans ce papier, nous nous intéressons à une cartographie des domaines de savoir-faire biologique réalisée par l'équipe BIOSIF « *Simulation, Intégration et Fouille de données*

Biologiques » en collaboration avec le Laboratoire de « *Génétique et d'Amélioration des Plantes* » de l'université d'Oran. Nous avons utilisé la méthode M3C « *Méthode de Cartographie des Connaissances Critiques* » (Aubertin *et al.*, 2004) pour définir les meilleurs savoir-faire biologiques concernant la génétique des plantes. Le résultat de la cartographie ainsi obtenu, est affiné par un processus de modélisation booléenne. Ce raffinement se fait par la machine cellulaire CASI (Atmani & Beldjilali, 2007) qui va assister la méthode M3C à l'optimisation de la carte et, assurer par la suite une modélisation booléenne.

1 – LA CARTOGRAPHIE DES CONNAISSANCES

Les connaissances cartographiées peuvent être à acquérir, à surveiller, à créer, à partager et enfin à capitaliser et à transférer. La méthode de cartographie des connaissances permet de représenter et analyser les connaissances d'une entreprise en les regroupant par domaine et en les visualisant sous forme de carte.

Dans toute opération de gestion des connaissances, la cartographie des connaissances a pour objectif de mettre en valeur les connaissances critiques métiers de l'entreprise (Ermine, 2006). La cartographie des connaissances est d'abord une démarche managériale dont la finalité est d'identifier les patrimoines de savoir-faire stratégique aux actions métiers dans l'organisation. L'identification de cette dernière dans une organisation consiste à pérenniser, développer les connaissances liées au métier de l'entreprise ainsi qu'à sa stratégie de travail. Son but principal consiste à montrer rapidement aux collaborateurs d'une organisation, d'un réseau ou d'une filière où se trouve l'expertise recherchée. De même elle indique l'importance d'un certain savoir et lequel risque de se perdre et qui doit donc être préservé (Karina, 2008). La cartographie des

connaissances devient alors un outil d'aide à la décision (Boughzala, 2007).

Plusieurs approches d'évolution de la cartographie ont été proposées pour organiser les ressources cognitives d'une entreprise. Aubertin *et al.*, (2004) ont proposé trois approches la première se focalise sur la classification fonctionnelle qui s'appuie sur un organigramme, la deuxième sur la classification par processus et la troisième sur la classification par domaines qui organise la connaissance autour de sujets, objets ou finalités. Clément .G *et al.* (2007) ont mené un projet pour cartographier des connaissances et des compétences techniques critique au sein de la direction de l'innovation et de la recherche de la SNCF. Ermine J.L (2005) a réalisé un projet piloté chez Chronopost International (*observatoire des métiers*) et qui repose sur deux objectifs ; d'une part, identifier les savoir-faire métiers impactés par la stratégie et, d'autre part, réfléchir à l'évolution des compétences critiques dans le futur. Pour cela, Ermine J.L à élaboré le projet en plusieurs phases : la première phase est la réalisation d'une cartographie stratégique aux actions métiers formaliser par l'approche du modèle graphique « *une carte des domaines de connaissances* ». La deuxième phase, consiste à une analyse des savoir-faire métiers critiques par l'utilisation des critères de criticité en tenant compte des spécificités de Chronopost International. Chabot J.L (2006) a proposé une cartographie complète des différents domaines d'expertise pour l'entreprise HYDRO-Québec. Cependant, l'objectif prioritaire est d'identifier, d'une part, les domaines de connaissances et, d'autre part, faire une étude de criticité pour faire ressortir les domaines de connaissances critique avec l'aide de la société française Kadrant. De même, Barroso A.C.O. et Ricciardi R. I. (2003) ont réalisé un projet piloté dans le centre de radiopharmacie de Sao Paulo (*IPEN*). Etant donné que le domaine nucléaire souffre de problèmes liés à cette accumulation considérable des connaissances : tels que le risque de non-préservation, la difficulté de transfert, etc. ils ont élaboré le projet en plusieurs étapes en utilisant l'approche par processus. Le processus a été décrit d'une manière classique sous la forme de diagrammes de flux liant des activités relatives aux processus (Aubertin *et al.*,2004).

Enfin, nous terminons cette brève présentation du rôle de la cartographie par les deux projets gouvernés par l'INRS « *Institut National de Recherche et de Sécurité* ». Le premier projet est piloté par Matta *et al.*, (2001) et qui a pour objectif de montrer l'apport de la méthode MASK pour une capitalisation de connaissances. La démarche a conduit à la réalisation d'un livre de connaissances. Dans cette optique, l'INRS a lancé le deuxième projet basé sur la méthode de cartographie des connaissances par domaine pour mettre en valeur les domaines de connaissances critiques et qui organise les connaissances autour de sujets, d'objets ou de finalités (Aubertin *et al.*, 2004).

Dans ce contexte, et comme nous l'avons déjà souligné, nous nous sommes intéressés à une cartographie des domaines de savoir-faire biologiques réalisée en collaboration avec le laboratoire de « *Génétique et d'Amélioration des Plantes* » de l'université d'Oran. Les fondements de notre projet se basent essentiellement sur deux concepts : la cartographie biologique des connaissances génétiques des plantes et, la criticité, telle que développé par le club de gestion des connaissances (Ermine & Boughzala, 2005).

Parmi les questions clés de cette démarche nous citons : Quelles sont les connaissances qu'on peut partager dans ce cadre? Comment identifier les savoirs et savoir-faire? Quelles sont les connaissances à y capitaliser et par quels moyens?. Pour répondre à ces questions, nous avons décidé de réaliser la cartographie des connaissances cruciales, en appliquant la démarche proposée par Aubertin *et al.*, (2004). Pour ces auteurs, il faut tout d'abord déterminer les domaines de savoir ou les processus sensibles. Ensuite, pour chaque domaine du savoir il faut délimiter les connaissances liées, et finalement, il faut évaluer le niveau de criticité de chaque connaissance pour déterminer les connaissances cruciales de l'organisation. La carte est le plus souvent représentée par un modèle graphique.

La figure 2 illustre la cartographie des domaines de connaissances biologiques

cruciales obtenues par l'outil « *MindManager Pro 8* »¹.

¹ Michael & Battina : *MindManager Pro 8*, <http://www.mindmanager.fr>, 2009.

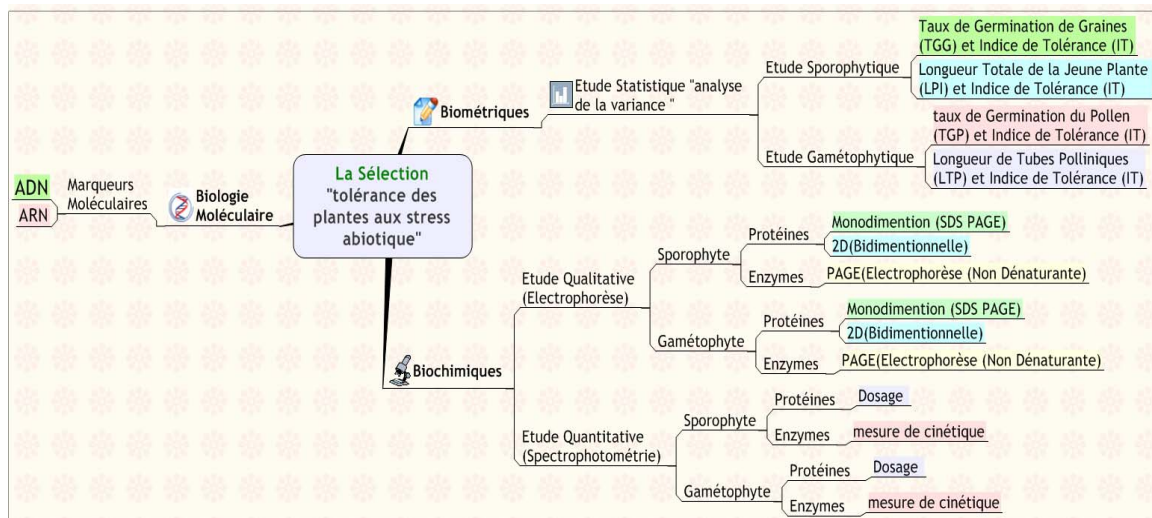


Figure. 2 – Cartographie des domaines génétique des plantes : vue générale

2 – MODELISATION BOOLEENNE DES CONNAISSANCES

Fascinées par la capacité impressionnante des outils informatiques, les entreprises se sont dotées de vastes systèmes de traitement et de communication de l'information espérant, sans doute, dominer les problèmes liés à la création de la connaissance. Cet écoulement, continu et croissant d'informations, peut maintenant être stocké et préparé à l'étude grâce aux nouvelles techniques d'Entrepôt de Données (ou Data Warehouse). Parmi les obstacles rencontrés pour réussir la création de connaissance par Fouille de données (ou Data Mining) nous citons: la quantité croissante des informations qui est générée et, la mise à la disposition des services concernés la bonne information au bon moment. En conséquence de l'arrivée de ces deux champs d'application, Fouille et Entrepôt de Données, une idée nouvelle s'impose : Pourquoi ne pas associer toutes ces techniques afin de créer des méthodes puissantes de recherche de connaissances, intégrant toutes les étapes allant du recueil des données jusqu'à l'évaluation de la connaissance acquise. C'est ainsi qu'est née l'idée d'une création des connaissances par un processus de fouille automatique à partir des données « apprentissage à partir de l'expérience passée ».

2.1 – Introduction à l'apprentissage automatique

L'apprentissage automatique est certainement, en intelligence artificielle, le champ d'application le plus fertile de ces dernières

années. On sait de manière générale qu'une des prérogatives de l'intelligence artificielle est d'apprendre à partir de l'expérience passée de sorte que son comportement devient adaptable. L'apprentissage automatique est ainsi le champ d'étude où l'on essaie de reproduire la capacité de l'homme à apprendre. Les pionniers interprètent l'*apprentissage automatique* comme un ensemble de changements dans un système qui permet à ce dernier d'accomplir mieux la même tâche, ou, une tâche similaire dans la même population dans l'avenir. On distingue ainsi trois niveaux de description d'un système d'apprentissage (Rakotomalala, 1997):

- Un système qui ne reçoit aucune entrée et qui accomplit le mieux une tâche,
- Un système qui reçoit des connaissances en entrée, mais n'accomplit aucune induction,
- Et enfin, un système qui reçoit des entrées et en extrait des connaissances qui ne sont connues ni implicitement ni explicitement, c'est l'apprentissage *inductif*.

C'est la dernière description qui nous intéresse dans cette problématique, plus particulièrement l'apprentissage empirique qui vise à produire une nouvelle connaissance à partir de l'expérience passée : cas pratiques, exemples, observations, etc... C'est ce que nous avons baptisé « *vers une cartographie automatique à partir d'une série d'observations* ».

Ce papier présente une nouvelle approche de cartographie des connaissances biologiques guidée par apprentissage automatique en vue de définir la stratégie de gestion des connaissances à engager.

La démarche que nous avons expérimentée s'appuie sur une nouvelle méthode cellulaire d'extraction de règles de cartographie à partir des données nommée *CASI* « *Cellular Automata for Symbolic Induction* » (Atmani & Beldjilali, 2007).

La première étape consiste à construire une cartographie des domaines de connaissances biologiques. La deuxième étape concerne la transformation de cette cartographie sous forme de partitions en respectant le principe booléen de *CASI*.

Le but, après une modélisation booléenne de la cartographie des domaines de connaissances, est double : d'une part affiner la cartographie par une fouille de données orchestrées par *CASI*, et d'autre part réduire la complexité de stockage, ainsi que le temps de calcul. Seule la modélisation booléenne de la cartographie est décrite dans ce papier. Le processus général d'apprentissage automatique que la machine cellulaire *CASI* applique à une population d'apprentissage est organisé en cinq étapes (voir figure 3):

1. A partir de l'expérience passée, appelé échantillon d'apprentissage Ω_a , nous commençons le traitement symbolique pour la construction du graphe d'induction « *méthode SIPINA²* ».
2. Initialiser le graphe d'induction par automate cellulaire « *coopération COG et CIE* »;
3. Optimiser le graphe d'induction booléen « *coopération COG et CIE* »;
4. Générer les règles cellulaire de cartographie « *coopération COG et CIE* »;
5. Valider la connaissance extraite « *coopération entre les modules CV et CIE* »;

² SIPINA : méthode non arborescente pour la construction des graphes d'induction par apprentissage automatique symbolique à partir des données.

L'objectif est de rechercher un modèle ϕ de création de connaissance permettant, pour une observation ω issue de Ω_a pour laquelle nous ne connaissons pas sa classe $Y(\omega)$ dans la carte mais dont nous connaissons l'état

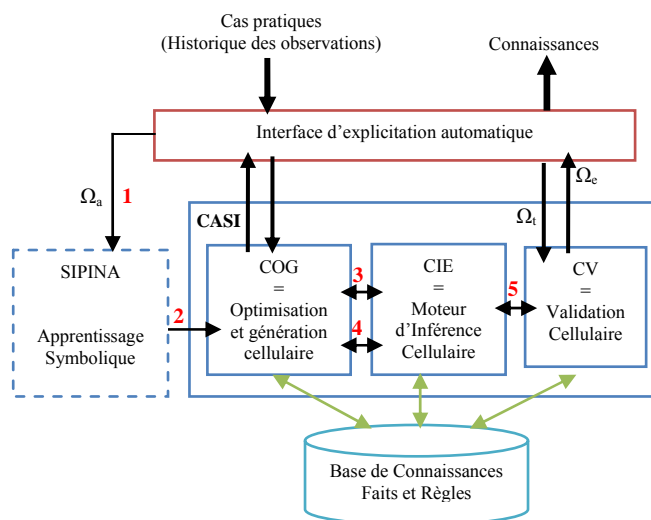


Figure. 3 – Diagramme général du système cellulaire (*CASI*)

2.2 – Représentation booléenne du graphe d'induction

Pour représenter notre cartographie des domaines de connaissances biologiques critiques, nous considérons, pour illustrer l'architecture et le principe de fonctionnement du module *CIE*, la partie du graphe illustrée par la figure 4, extraite de la cartographie des domaines génétique des plantes de la figure 2. Cette carte explique l'axe *Biochimique* avec ses thèmes, ses domaines et ses sous domaines.

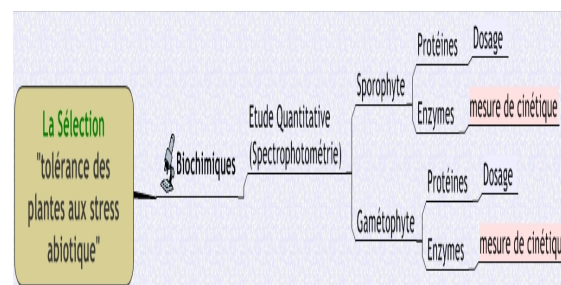
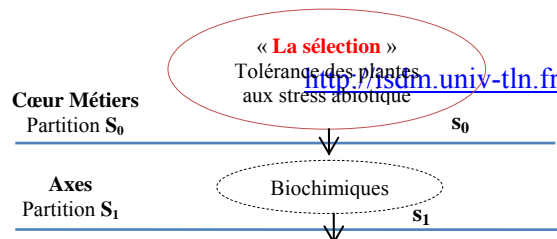


Figure. 4 – Cartographie des domaines génétique des plantes : axe « *Biochimique* »

La figure 4 illustre le graphe d'induction obtenue en utilisant les partitions S_0 (partition cœur métier) = $\{s_0\}$, S_1 (partition axes) = $\{s_1\}$, S_2 (partition thèmes) = $\{s_2\}$, S_3 (partition sous thèmes) = $\{s_3\}$, S_4 (partition domaines) = $\{s_4, s_5\}$ et S_5 (partition sous domaines) = $\{s_6, s_7\}$.



A partir de ces partitions nous déduisons sept règles de cartographie qui sont résumées dans la table 1 de la forme « *Si Condition alors Conclusion* ». Où *Condition* est une expression logique composée de conjonction que l'on nommera *Prémisse* et *Conclusion* la classe majoritaire dans le sommet décrit par la *condition*.

Figure. 5 – Construction des partitions S_0, S_1, S_2, S_3, S_4 et S_5 de la cartographie

Table 1. Base des connaissances

| Règles <i>i</i> | Si | Prémisse | Alors | Conclusion |
|-----------------|----|-----------|-------|--|
| Règle1 | si | { s_0 } | alors | {axes=Biochimique, s_1 } |
| Règle2 | si | { s_1 } | alors | {themes= Etude-qualitative, s_2 } |
| Règle3 | si | { s_2 } | alors | {sous-thèmes=Sporophyte, s_3 } |
| Règle4 | si | { s_3 } | alors | {domaines=Protéines, s_4 } |
| Règle5 | si | { s_3 } | alors | {domaines=Enzymes, s_5 } |
| Règle6 | si | { s_4 } | alors | {sous-domaines=Dosage, s_6 } |
| Règle7 | si | { s_5 } | alors | {sous-domaines=Mesure-cinétique, s_7 } |

Le moteur d'inférence du système cellulaire « *CIE* » utilise deux couches finies d'automates finis. La première couche, *CELFAIT*, pour la base des faits et, la deuxième couche, *CELREGLE*, pour la base de règles. Les états des cellules se composent de trois parties : *EF*, *IF* et *SF*, respectivement *ER*, *IR* et *SR*, sont l'entrée, l'état interne et la sortie d'une cellule de *CELFAIT*, respectivement d'une cellule de *CELREGLE*. L'état interne, *IF* d'une cellule de *CELFAIT* indique le rôle du fait : dans notre graphe $IF = 0$ correspond à un fait du type sommet « s_i », $IF = 1$ correspond à un fait du type *attribut=valeur* « $X_i = valeur$ » (Atmani & Beldjilali, 2007) (Abdelouhab & Atmani, 2008). Un automate cellulaire est une grille composée de cellules qui changent d'état dans des étapes discrètes. Après chaque étape, l'état de chaque cellule est modifié selon les états de ses voisins dans l'étape précédente. Les cellules sont mises à

jour d'une manière synchrone et les transitions sont effectuées, dans la théorie, simultanément. En appliquant des règles δ_{fait} et $\delta_{règle}$ simples et des transitions spécifiques, l'automate cellulaire *CIE* peut effectuer, d'une manière globale, des opérations complexes. Pour le voisinage des cellules du *CIE*, nous utilisons les deux matrices d'incidence d'entrée R_E et de sortie R_S qui représentent la relation entrée/sortie des faits (Atmani & Beldjilali, 2007). La figure 6 montre comment la base de connaissances extraite à partir du graphe (figure 5) est représentée par les couches *CELFAIT* et *CELREGLE*. Initialement, toutes les entrées des cellules dans la couche *CELFAIT* sont passives « $EF = 0$ », exceptées celles qui représentent la base des faits initiale « $EF(1) = 1$ ».

| Faits | EF | IF | SF | Règles | ER | IR | SR |
|---------------------------------|----------------|----|----|----------------|-----------------|----|----|
| s ₀ | 1 | 0 | 0 | R ₁ | 0 | 1 | 1 |
| axes=Biochimique | 0 | 1 | 0 | R ₂ | 0 | 1 | 1 |
| s ₁ | 0 | 0 | 0 | R ₃ | 0 | 1 | 1 |
| thèmes=Etude quantitative | 0 | 1 | 0 | R ₄ | 0 | 1 | 1 |
| s ₂ | 0 | 0 | 0 | R ₅ | 0 | 1 | 1 |
| sous-thèmes=Sporophyte | 0 | 1 | 0 | R ₆ | 0 | 1 | 1 |
| s ₃ | 0 | 0 | 0 | R ₇ | 0 | 1 | 1 |
| domaines=Protéines | 0 | 1 | 0 | | CELREGLE | | |
| s ₄ | 0 | 0 | 0 | | | | |
| domaines=Enzymes | 0 | 1 | 0 | | | | |
| s ₅ | 0 | 0 | 0 | | | | |
| sous-domaines=Dosage | 0 | 1 | 0 | | | | |
| s ₆ | 0 | 0 | 0 | | | | |
| sous-domaines =Mesure-cinétique | 0 | 1 | 0 | | | | |
| s ₇ | 0 | 0 | 0 | | | | |
| | CELFAIT | | | | | | |

Figure. 6 – Configuration initiale G_0 de l'automate cellulaire

2.3 – Voisinage et fonctions de transition

Dans l'opération suivante nous présentons dans la Table 2 les matrices d'incidence d'entrée R_E et de sortie R_S formulées comme suit :

✓ la relation d'entrée, notée $i R_E j$, est formulée par :

$\forall i \in [1, l], \forall j \in [1, r]$, si (le Fait $i \in$ à la prémisse de la règle j) alors $R_E(i, j) \leftarrow 1$.

✓ la relation de sortie, notée $i R_S j$, est formulée comme suit :

$\forall i \in [1, l], \forall j \in [1, r]$, si (le Fait $i \in$ à la conclusion de la règle j) alors $R_S(i, j) \leftarrow 1$.

Table 2 - Les deux matrices d'incidence d'entrée / sortie de la figure 6

| | R_E (relation d'entrée) | | | | | | | R_S (relation de sortie) | | | | | | |
|---------------------------------|---------------------------|----|----|----|----|----|----|----------------------------|----|----|----|----|----|----|
| | R1 | R2 | R3 | R4 | R5 | R6 | R7 | R1 | R2 | R3 | R4 | R5 | R6 | R7 |
| s ₀ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| axes=Biochimique | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| s ₁ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| thèmes=Etude quantitative | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| s ₂ | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| sous-thèmes=Sporophyte | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| s ₃ | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| domaines =Protéines | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| s ₄ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| domaines=Enzymes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| s ₅ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| sous-domaines=Dosage | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| s ₆ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| sous-domaines =Mesure-cinétique | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| s ₇ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Comme déjà signalé ci-dessus, la dynamique de l'automate cellulaire CIE , pour simuler le fonctionnement d'un Moteur d'Inférence, utilise deux fonctions de transitions δ_{fait} et $\delta_{règle}$ qui sont de la forme :

- La fonction de transition δ_{fait} pour effectuer une transition de t vers $t+1$:

$$(EF, IF, SF, ER, IR, SR)^{(t)} \rightarrow (EF, IF, \mathbf{EF}, \mathbf{ER} + (\mathbf{R}_E^T \cdot \mathbf{EF}), IR, SR)^{(t+1)}$$

- La fonction de transition $\delta_{règle}$ pour effectuer une transition de $t+1$ vers $t+2$:

$$(EF, IF, SF, ER, IR, SR)^{(t+1)} \rightarrow (EF + (\mathbf{R}_S \cdot \mathbf{ER}), IF, SF, ER, IR, \mathbf{ER})^{(t+2)}$$

Où la matrice R_E^T désigne la transposé de R_E et $\wedge ER$ la négation de ER .

En utilisant le principe booléen, la figure 7 ci-dessous présente l'état global des deux couches *CELFAIT* et *CELREGLE*, après *évaluation*,

sélection et filtrage en chaînage avant selon le mode *synchrone* : application de la première loi de transition δ_{fait} .

| <i>Faits</i> | <i>EF</i> | <i>IF</i> | <i>SF</i> | <i>Règles</i> | <i>ER</i> | <i>IR</i> | <i>SR</i> |
|---------------------------------|----------------|-----------|-----------|----------------|-----------------|-----------|-----------|
| S ₀ | 1 | 0 | 1 | R ₁ | 1 | 1 | 1 |
| axes=Biochimique | 0 | 1 | 0 | R ₂ | 0 | 1 | 1 |
| S ₁ | 0 | 0 | 0 | R ₃ | 0 | 1 | 1 |
| thèmes=Etude quantitative | 0 | 1 | 0 | R ₄ | 0 | 1 | 1 |
| S ₂ | 0 | 0 | 0 | R ₅ | 0 | 1 | 1 |
| sous-thèmes=Sporophyte | 0 | 1 | 0 | R ₆ | 0 | 1 | 1 |
| S ₃ | 0 | 0 | 0 | R ₇ | 0 | 1 | 1 |
| domaines=Protéines | 0 | 1 | 0 | | CELREGLE | | |
| S ₄ | 0 | 0 | 0 | | | | |
| domaines=Enzymes | 0 | 1 | 0 | | | | |
| S ₅ | 0 | 0 | 0 | | | | |
| sous-domaines=Dosage | 0 | 1 | 0 | | | | |
| S ₆ | 0 | 0 | 0 | | | | |
| sous-domaines =Mesure-cinétique | 0 | 1 | 0 | | | | |
| S ₇ | 0 | 0 | 0 | | | | |
| | CELFAIT | | | | | | |

Figure. 7 - Configuration obtenue avec δ_{fait}

La figure 8 présente l'état global des deux couches *CELFAIT* et *CELREGLE*, après *exécution* en chaînage avant selon le mode

synchrone : application de la deuxième loi de transition $\delta_{règle}$.

| <i>Faits</i> | <i>EF</i> | <i>IF</i> | <i>SF</i> | <i>Règles</i> | <i>ER</i> | <i>IR</i> | <i>SR</i> |
|---------------------------------|----------------|-----------|-----------|----------------|-----------------|-----------|-----------|
| S ₀ | 1 | 0 | 1 | R ₁ | 1 | 1 | 0 |
| axes=Biochimique | 1 | 1 | 0 | R ₂ | 0 | 1 | 1 |
| S ₁ | 1 | 0 | 0 | R ₃ | 0 | 1 | 1 |
| thèmes=Etude quantitative | 0 | 1 | 0 | R ₄ | 0 | 1 | 1 |
| S ₂ | 0 | 0 | 0 | R ₅ | 0 | 1 | 1 |
| sous-thèmes=Sporophyte | 0 | 1 | 0 | R ₆ | 0 | 1 | 1 |
| S ₃ | 0 | 0 | 0 | R ₇ | 0 | 1 | 1 |
| domaines=Protéines | 0 | 1 | 0 | | CELREGLE | | |
| S ₄ | 0 | 0 | 0 | | | | |
| domaines=Enzymes | 0 | 1 | 0 | | | | |
| S ₅ | 0 | 0 | 0 | | | | |
| sous-domaines=Dosage | 0 | 1 | 0 | | | | |
| S ₆ | 0 | 0 | 0 | | | | |
| sous-domaines =Mesure-cinétique | 0 | 1 | 0 | | | | |
| S ₇ | 0 | 0 | 0 | | | | |
| | CELFAIT | | | | | | |

Figure. 8 – Configuration obtenue avec $\delta_{règle}$

Nous considérons G_0 la configuration initiale de l'automate cellulaire illustrée par la figure 6 et, $\Delta = \delta_{règle} \circ \delta_{fait}$ la fonction de transition globale : $\Delta(G_0) = G_1$. Après l'application de la

loi de transition globale Δ nous obtenons les configurations G_1, G_2, G_3, G_4 , et enfin G_5 (voir figure 8).

| <i>Faits</i> | <i>EF</i> | <i>IF</i> | <i>SF</i> | <i>Règles</i> | <i>ER</i> | <i>IR</i> | <i>SR</i> |
|---------------------------|-----------|-----------|-----------|----------------|-----------|-----------|-----------|
| S ₀ | 1 | 0 | 1 | R ₁ | 1 | 1 | 0 |
| axes=Biochimique | 1 | 1 | 1 | R ₂ | 1 | 1 | 0 |
| S ₁ | 1 | 0 | 1 | R ₃ | 1 | 1 | 0 |
| thèmes=Etude-quantitative | 1 | 1 | 1 | R ₄ | 1 | 1 | 0 |
| S ₂ | 1 | 0 | 1 | R ₅ | 1 | 1 | 0 |
| sous-thèmes=Sporophyte | 1 | 1 | 1 | R ₆ | 1 | 1 | 0 |
| S ₃ | 1 | 0 | 1 | R ₇ | 1 | 1 | 0 |

Figure. 8 – Configuration finale G_5 obtenue après cinq itérations synchrones

On peut également utiliser R_S comme relation d'entrée et R_E comme relation de sortie pour lancer une inférence en chaînage arrière.

2.4 – Génération des règles

Sous réserve que notre échantillon soit représentatif de la population originelle nous pouvons déduire deux règles conjonctives de cartographie :

1. **Si** axes = « *Biochimique* »
et thème = « *Etude-quantitative* »
et sous-thème = « *Sporophyte* »
et domaines = « *Protéines* »
et sous-domaines = « *Dosage* »
alors classe majoritaire de s_6 .
2. **Si** axes = « *Biochimique* »
et thèmes = « *Etude-quantitative* »
et sous-thèmes = « *Sporophyte* »
et domaines = « *Enzymes* »
et sous-domaines = « *Mesure-cinétique* »
alors classe majoritaire de s_7 .

3 - CONCLUSION : LA CONVERGENCE

Une cartographie guidée par *fouille de données* possède des propriétés intéressantes et de nombreux avantages par rapport aux autres techniques de gestion de connaissances. Ce nouveau principe de cartographie constitue un outil très satisfaisant pour la représentation et la gestion des connaissances en adoptant une modélisation booléenne.

En effet, le résultat booléen de la carte ainsi obtenu, est perfectionné par un processus d'apprentissage automatique symbolique à base de graphe d'induction. Cette amélioration

se fait par l'automate cellulaire *CIE* qui va assister *SIPINA* dans le processus d'extraction de nouvelles connaissances tacites à partir de l'expérience passée « *sous forme de cas pratiques* » et assurer, par la suite, une contribution dans le processus général de création et de transfert des connaissances.

Les avantages de cette approche basée sur le principe de *CASI* peuvent être récapitulés comme suit:

- La représentation de la connaissance ainsi que son contrôle sont simples, sous forme de matrices binaires exigeant un prétraitement minimal.
- La facilité de l'implémentation des fonctions de transition qui sont de basse complexité.
- Les résultats de la cartographie sont simples pour être réorganisé par une fouille de données.
- Le système de cartographie est un modèle cellulaire composé d'un ensemble simple de fonctions de transition et de règles de production, qui permettent non seulement de décrire le problème actuel mais d'établir également une fonction de classification pour la prévision.
- La matrice d'incidence, R_E , facilite la transformation des liens entre partitions dans des expressions équivalentes booléennes, qui nous permet d'utiliser l'algèbre de Boole élémentaire pour examiner d'autres simplifications.

BIBLIOGRAPHIE

- Abdelouhab FZ, Atmani B. (2008). « *Intégration automatique des données semi-structurées dans un entrepôt cellulaire* », Troisième atelier sur les systèmes décisionnels, octobre 2008, Mohammadia – Maroc, p. 109-120.
- Atmani B, Beldjilali B. (2007). Knowledge Discovery in Database : Induction Graph and Cellular Automaton, Computing and Informatics Journal, Vol.26, N°2 P: 171-197.
- Aubertin G. (2004). « *Cartographier les connaissances critiques : une démarche stratégique pour l'entreprise* ». In Management des connaissances en entreprise, Ed. Lavoisier.
- Barroso A, Ricciardi R. (2003). « *Knowledge domains cartography of the radiopharmacy center of IPEN – a case study* ». Nuclear and Energy Research Institute (IPEN). Brazil. P.3-6.
- Benamina M, Atmani B. (2008). « *WCSS: un système cellulaire d'extraction et de gestion des connaissances* », Troisième atelier sur les systèmes décisionnels, Mohammadia – Maroc, p. 223-234.
- Boughzala I. (2007). « *Ingénierie de la collaboration : théories, technologies et pratiques* ». Edition : Lavoisier, Paris, P :203-216.
- Chabot J.L. (2006). « *Transfert de savoir en hydro-québec perspective et stratégie* ». Colloque annuel de cerfio atelier N°3.
- Ermine J-L. (2006) « *Introduction au Knowledge Management* ». From Journées d'études INFORUM'06. From information Management to Knowledge Management.
- Ermine J-L, Boughzala I. (2005). « *Using Cartography to Sustain Inter-Generation Knowledge Transfer: The M3C Methodology* ». In 2nd International Conference on Intellectual Management, Knowledge Management and Organisational Learning. American University in Dubai. U.A.E.
- Ermine J-L. (2007). « *Méthode d'alignement stratégique pour la gestion des connaissances* », Plus d'information : [20070920-INT-méthode-gestion-connaissances.pdf](#), p. 11-20
- Fayad-lameche F-Z, Yahia N. & Lachheb F. (2007). « *Tolérance au froid chez la tomate (Lycopersicum esculentum Mill.)* ». Etude comparée du développement du gamétophyte mâle et du sporophyte. Acta Bot. Gallica, 2007, 154 (2), p. 251-263.
- Guelain C. (2007). « *Cartographie des connaissances critiques au service de la recherche de la SNCF* ». Plus d'information : [230-cartographie_sncf.pdf](#).
- Karina M. (2008). « *Cartographie des connaissances (Knowledge Map)* ». Plus d'Information: [http://www.daretoshare.ch/KnowledgeMapCalend arF-kompr-korrigiert.pdf](#).
- Lavinia C. (2006). « *La cartographie des connaissances critiques* ». Plus d'Information: [Dossierdumoiss64ccc.pdf](#).
- Matta N, Ermine J.L. (2001). « *knowledge capitalization with a knowledge engineering approach : the mask method* ». IJCAT'2001, knowledge management and organisational memory workshop. International Joint Conference on Artificial Intelligence, seattle, Etats-Unis, 4-10 août 2001.
- Nhien-an le-khac, Lamine M. (2007). « *Knowledge Management for Distributed Data Mining Environments* ». In : International journal of computational intelligence, vol 4; n°3, p.3-4.
- Nonaka I, Takeuchi H. (1997). « *La connaissance créatrice, la dynamique de l'entreprise apprenante* », p. 2-6.
- Prax J-Y. (2007). « *Le manuel du Knowledge Management* », une approche de 2^{ème} génération. Edition, Dunod. Paris. P.219- 230.
- Rakotomalala R. (1997). « *Graphes d'induction* », Thèse pour l'obtention du Diplôme de Doctorat, Université Claude Bernard-Lyon 1.
- Rejean R. (2004) « *La cartographie des connaissances : votre réservoir de savoirs est-il bien plein ?* », Réseau CEFRIO, p. 7-9, juin 2004.