

Chapitre 3 : Les systèmes à base de connaissances



Pr. Mustapha BOURAHLA, Département
d'Informatique, Université de M'Sila, Contact :
mustapha.bourahla@univ-msila.dz

Table des matières



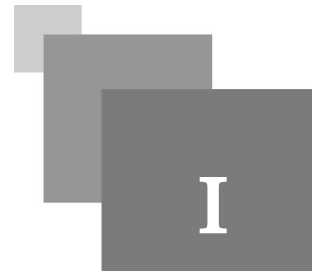
Introduction	3
I - Définitions	4
II - Architecture des systèmes à base de connaissances	6
III - Ingénierie des connaissances	10
IV - Domaines d'application	12
V - Exercices sur le troisième chapitre	13
Conclusion	15
Solutions des exercices	16

Introduction



L'IA est la branche de l'informatique qui consiste à concevoir des systèmes intelligents, c'est-à-dire qui soient capables de produire un raisonnement proche de celui de l'être humain. Ces systèmes intelligents sont en général à base de connaissances.

Définitions



Différence entre donnée, information et connaissance

- Une donnée transporte l'information (un signal non interprété).
- L'information est une interprétation de la donnée (par exemple, la donnée "!" transporte l'information (interprétation) "point d'exclamation"),
- La connaissance utilise l'information dans le cadre d'actions, dans un but précis (par exemple, écrire un "!" pour marquer une exclamation en fin de phrase).
- Les actions peuvent être la prise de décisions, la création de nouvelles informations, etc.

Base de connaissances

- C'est un ensemble de connaissances dédié à un domaine particulier (c'est un ensemble de faits et de règles

Système à base de connaissances

- Un système à base de connaissances (SBC) est un programme construit pour modéliser les compétences de résolution de problèmes des humains.
- C'est à dire, un système informatique fonctionnant avec une base de connaissances explicite sur un sujet donné.

Système expert

- Système informatique permettant de résoudre les problèmes dans un domaine d'application déterminé à l'aide d'une base de connaissances établie à partir de l'expertise humaine.
- C'est à dire un système expert classique est un cas particulier d'un système à base de connaissances (tout système informatique utilise de la connaissance, mais dans un système à base de connaissances, celle-ci est représentée de façon explicite.).

C'est un logiciel jouant le même rôle qu'un expert humain, c' est-à- dire capable, à partir de ses connaissances spécialisées, en appliquant des méthodes d'inférence, de fournir des conseils relatifs à son domaine d'expertise.

Un expert doit :

1. Posséder un savoir spécialisé qui regroupe de vastes connaissances dans un domaine donné. Ce savoir ne se limite pas à des connaissances factuelles, mais il comprend un réseau de concepts, de règles, de méthodes qui forment l'expertise.
2. Savoir utiliser les connaissances efficacement. Il ne suffit pas à l'expert de posséder de vastes connaissances, il doit pouvoir les utiliser rapidement et avec succès. Pour cela, il doit savoir poser les bonnes questions au client de façon à repérer les connaissances qui s' appliquent dans son cas.
3. Connaître ses limites de façon à ne pas s' occuper de questions hors du cadre de sa compétence.

4. Fournir des explications appropriées sur les informations qu'il donne ou les solutions qu'il propose. Généralement, ce n'est pas l'expert qui décide, mais celui qui demande conseil. Il est donc un conseiller qui doit être capable de convaincre le client que sa solution est bonne.

La consultation d'un expert est donc essentiellement un dialogue où celui-ci pose des questions de plus en plus précises pour identifier les connaissances dont il a besoin et où il explique les raisons de ses recommandations. Dans un tel dialogue, le client fournit des faits et pose des questions pour bien comprendre le sens de la solution proposée par l'expert. À la fin, il lui revient de prendre la décision finale

Les inférences

- Avec un moteur d'inférence on peut déduire des faits à partir de faits initiaux et des règles.
- Deux approches de base :
 1. A partir de ce qu'on veut trouver et remonter vers les faits (chaînage arrière) ;
 2. A partir des faits et aller vers ce qu'on veut trouver (chaînage avant).
- La structure classique des systèmes experts utilisaient toujours un ensemble de règles de production.

Mécanismes de raisonnement

- Pour les inférences, la plupart des systèmes à base de connaissances (systèmes experts) existants reposent sur des mécanismes de logique formelle et utilisent le raisonnement déductif.
- Pour l'essentiel, ils utilisent le syllogisme : si P est vrai (fait ou prémisse) et si on sait que $P \rightarrow Q$ (règle) alors Q est vrai (nouveau fait ou conclusion).
- Les plus simples des systèmes experts s'appuient sur la logique des propositions (dite aussi « logique d'ordre 0 »).
- Dans cette logique, on n'utilise que des propositions, qui sont vraies, ou fausses.
- D'autres systèmes s'appuient sur la logique des prédicats du premier ordre (dite aussi « logique d'ordre 1 »), que des algorithmes permettent de manipuler aisément.

Architecture des systèmes à base de connaissances



Architecture

- Un système expert est un outil capable de reproduire les mécanismes cognitifs d'un expert, dans un domaine particulier.
- Plus précisément, un système expert est un logiciel capable de répondre à des questions, en effectuant un raisonnement à partir de faits et de règles connus.
- Il peut servir notamment comme outil d'aide à la décision.
- Dans tout système à base de connaissances ou système expert, on retrouvera à la base les composantes suivantes :
 1. Une base de connaissances composée d'une base de faits et une base de règles ;
 2. Un moteur d'inférence : Le moteur d'inférence est capable d'utiliser faits et règles pour produire de nouveaux faits, jusqu'à parvenir à la réponse à la question posée.
 3. En plus à une interface

1. Base de connaissances

- La base de connaissances est propre au domaine traité et, contrairement aux bases de données, n'est pas limitée aux connaissances de type factuel, aux données.
- Elle rassemble tous les types de connaissances pertinentes pour le domaine considéré : description des objets et de leurs relations, règles à appliquer pour résoudre un problème, méta-connaissances permettant de choisir quelles règles appliquer, etc.
- D'autre part, le système conserve à tout moment dans sa mémoire de travail, une base des faits connus.
- Grâce à elle, le moteur d'inférence peut choisir les éléments de sa base de connaissances; par exemple, les règles à utiliser en fonction des faits disponible, de manière efficace et adaptée au problème posé.
- La base de faits s'enrichit au fur et à mesure que le moteur d'inférence déduit de nouveaux faits en appliquant des connaissances aux faits déjà connus.

2. Moteur d'inférence

- Le moteur d'inférence est le programme qui construit les raisonnements en puisant ses matériaux dans la base de connaissances et la base de faits.
- En examinant la base de faits, il détecte les connaissances intéressantes pouvant s'appliquer à certains faits, les enchaîne et construit un plan de résolution.
- Il déduit de nouveaux faits de ceux fournis au départ ou en cours d'interaction par l'utilisateur.

- Indépendant du domaine, le moteur d'inférence rassemble les mécanismes de raisonnement qui vont exploiter la base de connaissances.

3. Interface

- À ces deux composantes essentielles s'ajoutent des modules d'interface eux aussi indépendants du domaine de connaissance.
- Grâce à eux, l'expert peut accéder facilement à la base de connaissances, la modifier en rectifiant une information inutile ou erronée, ou en ajoutant une précision.
- L'utilisateur peut quant à lui suivre le raisonnement du système dans un langage qui lui est naturel, poser des questions, demander des explications, sans avoir besoin d'acquiescer une connaissance approfondie des systèmes experts ou de l'informatique.

Séparation entre les connaissances spécialisées et les règles

- Séparation de la connaissance et du raisonnement : les systèmes à base de connaissances, tout comme les systèmes experts, sont fondés sur une séparation nette entre les connaissances et les méthodes d'inférence.
- Ainsi les connaissances, au lieu d'être imbriquées dans la structure du programme sont considérées comme des données interchangeables, susceptibles d'être corrigées, mises à jour et exploitées par des programmes (les moteurs d'inférence) qui, eux, simulent les mécanismes de raisonnement et assurent l'interface avec les utilisateurs.
- L'idée de distinguer les connaissances spécialisées des règles agissant sur cet ensemble a permis plusieurs progrès importants dans certains domaines de l'intelligence artificielle.
- La technologie des systèmes experts a notamment profité de cette distinction en permettant le développement de moteurs d'inférence indépendants d'un domaine.
- Ces logiciels contiennent une représentation des connaissances générales nécessaires au traitement de règles et de faits et permettent à un utilisateur de se concentrer uniquement sur la construction d'une base de connaissances spécialisées.

Inférence et résolution de problèmes

- Résoudre des problèmes est une des fonctions fondamentales de l'intelligence artificielle qui exige une bonne représentation des connaissances et des méthodes d'inférence dont l'application permettra de trouver la solution.
- On construit ainsi une base de connaissances qui, au départ, contiendra une description de l'état initial. La base de connaissances contiendra également des opérations exprimées par des règles permettant de transformer cet état initial en d'autres états qui, petit à petit, nous rapprocheront de l'état final ou solution.
- **Heuristique** : Technique consistant à apprendre petit à petit, en tenant compte de ce que l'on a fait précédemment pour tendre vers la solution d'un problème.

Exemple : Le problème des seaux à remplir

-
- Prenons le problème suivant : deux seaux A et B contiennent respectivement 9 et 4 litres.
 - Une personne a un tonneau de douze litres d'huiles ; elle veut en donner 3 litres à un ami. Pour les mesurer, elle n'a que ces deux seaux A et B, A contient 9 litres et B contient 4 litres. Comment doit-elle opérer pour avoir les 3 litres dans le seau A?

- Peut-on obtenir exactement 3 litres dans le contenant A, uniquement en remplissant ou en vidant les seaux ou en les versant l'un dans l'autre?

On peut représenter les connaissances nécessaires pour résoudre ce problème à l'aide d'états et d'opérateurs comme suit :

- les états sont de la forme [A] [B] où A et B sont le nombre de litres de chaque récipient; par exemple, [4] [5] signifie que le récipient A contient 4 litres et le récipient B, 5 litres.
- les opérateurs sont représentés par les règles SI... ALORS de transformation entre états permises par l'énoncé du problème tel qu'indiqué au dessous.

```

1 Remplir(A) = Si [A][B] Alors [9][B]
2 Remplir(B) = Si [A][B] Alors [A][4]
3 Vider(A) = Si [A][B] Alors [0][B]
4 Vider(B) = Si [A][B] Alors [A][0]
5 Verser(B dans A) = Si [A][B] et A+B <= 9 Alors [A+B][0]
6 Verser(B dans A) = Si [A][B] et A+B > 9 Alors [9][A+B-9]
7 Verser(A dans B) = Si [A][B] et A+B <= 4 Alors [0][A+B]
8 Verser(A dans B) = Si [A][B] et A+B > 4 Alors [A+B-4][4]
    
```

Signification des opérateurs

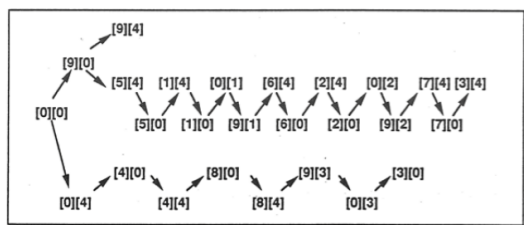
- La présence de deux opérateurs pour le versement tient compte du fait que le calcul sera différent si on peut verser complètement un récipient dans un autre ou seulement une partie.
- Par exemple, l'avant-dernier opérateur signifie que si le total des deux récipients est inférieur ou égal à 4 litres, on peut verser tout le contenu de A dans B, lequel contient alors A + B litres, alors que A est devenu vide.
- Dans le cas contraire, l'autre opérateur affirme que l'on pourra remplir B jusqu'à 4 litres et que dans A, on ajoutera B - 4 litres à A qui contiendra alors A + B - 4 litres.

Inférence

- Pour utiliser ces connaissances, le programme devra contenir un moteur d'inférence qui « instancie » les variables A et B c'est-à-dire leur donne des valeurs particulières correspondant à un état déjà obtenu.
- Après cette opération, le moteur peut appliquer la règle de modus ponens et utiliser ces opérateurs comme des énoncés en logique propositionnelle.

Résolution du problème

- Le problème se présente donc comme une suite d'états obtenue en appliquant une suite d'opérateurs (instanciés correctement) à partir de l'état initial jusqu'à l'état final soit [0][0]; puis remplir A [9][0]; puis vider A dans B [5][4].
- Le problème « obtenir 3 litres dans A » n'est pas aussi simple et nécessite d'essayer divers chemins en appliquant tous les opérateurs possibles à une situation avant de trouver un chemin-solution



- La résolution de ce problème est un graphe (arbre) : les nœuds représentent les états et les arcs représentent les opérations comme en témoigne la figure.

Démarche de résolution

- Cette démarche peut être faite par un moteur d'inférence capable d'instanciation de variables.
- À partir d'un fait initial [0][0], on applique systématiquement les règles permises (les opérateurs) en déduisant toujours des faits nouveaux jusqu'à déduire la solution, un fait du type [3][B].
- On s'aperçoit qu'il y a plusieurs façons de parcourir l'arbre des possibilités. Notamment, l'ordre dans lequel on applique les opérateurs est critique.
- Ici, on perd beaucoup de temps à développer la « branche » [5][4], alors que la « branche » [0][4] conduit deux fois plus rapidement à une solution.
- L'étude des méthodes les plus efficaces de recherche dans un arbre des possibilités sert à la fois à trouver de bonnes méthodes de déduction ou d'inférence et de bonnes méthodes de résolution de problèmes.
- L'inférence, tout comme la résolution d'un problème, sont tous deux des mécanismes de recherche dans un arbre des possibilités qui peuvent être décrits ainsi :
 - En partant d'une situation initiale connue (les faits), atteindre une des situations finales (fait-conclusion, fait-solution) désirées en appliquant une suite finie d'opérateurs à la situation initiale ou aux situations déjà obtenues de celle-ci.

Ingénierie des connaissances



Définition

- L'ingénierie des connaissances est la gestion des connaissances au sein d'une organisation.
- Elle fait référence à l'ingénierie de systèmes complexes intelligents incorporant beaucoup de connaissances tels les systèmes à bases de connaissances et systèmes experts.

Dimensions

- L'exploitation des connaissances passe par cinq opérations : **identification, création, stockage, partage et utilisation.**
- L'ingénierie des connaissances se concentre sur l'identification, la création, le stockage et la mise à disposition des connaissances afin de rester neutre face aux outils de partage et d'utilisation.

Identification

- L'identification des connaissances consiste en l'identification des connaissances critiques pour une organisation.
- Elle peut s'effectuer de manière individuelle sur base de questionnaires, d'interviews, etc., ou de manière collective en identifiant les départements, services, équipes.

Collecte

La collecte peut consister en un transfert de connaissances ou en un partage de connaissances, le tout en structurant correctement l'information :

- Transfert de connaissances :
 - le tutorat ;
 - le parrainage.
- Partage de connaissances :
 - les communautés de pratique ;
 - l'inter-vision (partant d'un problème) ;
 - le focus groupe (petits groupes);
 - le retour d'expérience.

Structuration

- La structuration consiste à organiser les connaissances collectées en catégories, éventuellement en plusieurs niveaux hiérarchiques, pour en faciliter l'accès et la consultation.
- La structuration est une étape nécessaire pour pouvoir arriver à l'exploitation des données.

Exploitation

- La collecte et la structuration des données sont bien distinctes dans l'approche de la gestion des connaissances.
- Néanmoins, la collecte et la structuration ne suffisent pas pour parler d'une approche complète de gestion de connaissances.
- En effet, cette approche ne s'achève que lorsqu'on pourrait utiliser les données collectées et stockées pour atteindre les objectifs de l'organisme concerné.
- Ainsi, l'exploitation des données, connue aussi sous l'expression « Extraction des Connaissances à partir des Données » (ECD) (en anglais Knowledge Data Discovery, KDD) consiste à relier et à interpréter les faits pour en déduire des résultats et des conséquences utiles.
- L'extraction des connaissances à partir des données peut être réalisée suivant plusieurs méthodes, selon le domaine d'application et la nature des données brutes. Dans ce cadre, on distingue diverses méthodes

Méthode : Exploration de données

- L'exploration de données (aussi appelé fouille de données ou encore datamining pour les anglophones), est la pratique (par des moyens automatiques ou semi-automatiques) de la recherche et de l'exploration de grands ensembles de données ayant pour résultat la découverte de motifs significatifs et de règles.
- Pour faire cela, le data mining utilise des techniques informatiques empruntées à la statistique, et la reconnaissance de motifs récurrents dans de grandes masses de données récolté par un système d'information.
- Les résultats des analyses Data Mining ont pour but de connaître le comportement d'un usager, inférer puis prévoir son comportement.
- Ainsi, ce sont les résultats du datamining qui génèrent de la connaissance pour l'organisation (Knowledge Discovery). En entreprise, le data mining s'utilise principalement en:
 - Marketing: la connaissance du client, son comportement plus précisément est essentiel pour: mieux cibler les clients, les fidéliser, améliorer les forces de vente, améliorer la relation client.
 - Gestion des risques: principalement dans le secteur bancaire et assurantiel. Il s'agit de connaître des clients ou partenaires à risque et mesurer ce risque. Cette information est indispensable pour la banque ou l'assurance car constitue le cœur de son métier.

Méthode : intelligence artificielle

- L'intelligence artificielle peut se définir comme un traitement automatique des données basé sur des règles universelles pour la prise de décisions satisfaisantes au cas traité.
- Les méthodes d'ECD qui se basent sur l'intelligence artificielle ont évolué considérablement depuis l'apparition du web 2.0 et depuis le progrès de l'informatique notamment face à la multiplication des sources de données (Web, ERP, gestion de la relation client, etc.)

Domaines d'application



IV

Pourquoi utiliser un système à base de connaissances

- Pour remplacer un expert par un programme
- Automatiser une tâche routinière nécessitant un expert
- Besoin d'une expertise dans un environnement hostile
- Assister un expert
- Améliorer la productivité
- Gérer la complexité

Les domaines d'applications

- Santé et IA,
- Agriculture et IA,
- Énergie hydrocarbure,
- énergie solaire + IA,
- Environnement (catastrophes naturelles (inondation, séisme, etc),...) + IA,
- Administration et IA,
- Défense nationale et IA,
- Transport et IA
- industrie et IA

Santé et IA : MYCIN

- Il s'attaque au problème du diagnostic et du traitement des maladies infectieuses du sang.
- Sa base de connaissances comprend à peu près 400 règles qui mettent en relations des états possibles et des interprétations correspondantes.
- Lorsqu'il résout un problème, MYCIN compare les conditions d'application d'une règle aux données disponibles ou encore, il demande des données additionnelles au médecin.
- Le cas échéant, il essaie de déduire la vérité ou la fausseté d'une condition à partir d'autres règles.

Exercices sur le troisième chapitre

V

Exercice : Donnée, Information et Connaissance

[solution n°1 p.16]

Donner la classe de

ahmed est une personne

ahmed

si ahmed est une personne alors il est humain

donnée	information	connaissance

Exercice : L'objectif du langage de données

[solution n°2 p.16]

Complétez les trous du texte suivant :

Un système à base de connaissances est composé de trois composants :

1. Une
2. Un
3. Une

Exercice : Base de connaissances

[solution n°3 p.16]

Soit la base de connaissances suivante :

$$\{pere(ahmed, mohamed), pere(mohamed, farid), pere(X, Z) \wedge pere(Z, Y) \rightarrow grandPere(X, Y)\}$$
Donner le graphe ET/OU pour le but (question) $grandPere(X, Y)$

Remplir les trous suivants

L'exploitation des connaissances passe par cinq opérations : , , , et .

Conclusion



Ce troisième chapitre a présenté un aperçu sur les systèmes à base de connaissances (SBC)



Solutions des exercices



> Solution n°1

Exercice p. 13

Donner la classe de

donnée	information	connaissance
ahmed	ahmed est une personne	si ahmed est une personne alors il est humain

> Solution n°2

Exercice p. 13

Complétez les trous du texte suivant :

Un système à base de connaissances est composé de trois composants :

1. Une base de de connaissances
2. Un moteur d'inférence
3. Une interface

> Solution n°3

Exercice p. 13

Soit la base de connaissances suivante :

$$\{pere(ahmed, mohamed), pere(mohamed, farid), pere(X, Z) \wedge pere(Z, Y) \rightarrow grandPere(X, Y)\}$$
Donner le graphe ET/OU pour le but (question) $grandPere(X, Y)$ le noeud initial : $grandPere(X, Y)$, le noeud suivant $pere(X, Z)$ ET $pere(Z, Y)$, le noeud suivant $pere(ahmed, mohamed)$ OU $pere(mohamed, farid)$ à continuer

> Solution n°4

Exercice p. 14

Remplir les trous suivants

L'exploitation des connaissances passe par cinq opérations : identification, création, stockage, partage et utilisation.