

Quels sont les processus de réflexion qui conditionnent un raisonnement d'expert ?

informatique et gestion

Techniques innovantes en informatique

Luc Rubiello

HERMES

Chapitre 2

L'acquisition des connaissances

La question philosophique de l'impact des ordinateurs sur les activités humaines, dès qu'ils outrepassent leur rôle de machines de stockage, de traitement et de transmission des données nous amène à prendre conscience que les problèmes de l'expertise et de la connaissance viennent de cette constatation : quels sont les processus de réflexion qui conditionnent un raisonnement expert et comment peut-on en faire une représentation satisfaisante en vue d'un codage pour une utilisation automatisée ?

Cette question centrale, clé des sciences cognitives, contient en elle toutes les autres questions que nous aborderons. Les SBC ne doivent pas être mesurés uniquement en termes de puissance de moteur ou de nombre de règles mais bien sur la qualité (ou tout simplement sur l'existence) de l'expertise contenue. L'apparition et l'accroissement de nouveaux besoins qui peuvent être couverts par les SBC ne font que rendre le problème plus urgent : comment capturer les connaissances ? Les connaissances extraites sont-elles correctes et utiles ?

2.1. Trois approches principales

Quels que soient la méthode ou l'outil utilisés durant la phase d'acquisition des connaissances, on peut définir trois grandes approches qui sont :

- l'interview des experts,
- l'acquisition par interaction,
- l'acquisition par induction.

Dans la pratique, les trois approches sont utilisées concurremment par les ingénieurs de la connaissance et parfois ces derniers ne font pas la distinction car il y a toujours à l'origine un dialogue ingénieur de la connaissance-expert. Les approches sont pourtant à la base de la différenciation des outils que l'on peut rencontrer.

2.1.1. *L'interview des experts*

L'interview de l'expert semble un passage obligé dans tout projet SBC.

La faible capacité de l'ingénieur de la connaissance à être un "bon communicant", facteur trop souvent ignoré, est la cause directe des échecs constatés. L'ingénieur de la connaissance doit également être capable de conceptualiser le domaine étudié d'après les énoncés de l'expert. Cette dualité de compétences ne s'acquiert pas facilement et met l'accent sur la formation inégale ou inexistante des ingénieurs de la connaissance.

2.1.2. *L'acquisition par interaction*

L'acquisition par interaction fait appel à une participation directe de l'expert pour l'expression des règles sans pour autant renoncer totalement à la présence d'un ingénieur de la connaissance. En fait, les produits existants se placent tout au long du spectre possible, depuis les simples aides à la conceptualisation (pour les ingénieurs de la connaissance) jusqu'aux outils de représentation visuelle.

2.1.3. *L'acquisition par induction*

L'acquisition par induction fait référence à l'apprentissage par l'exemple ou l'observation. C'est un thème récurrent du "machine learning". Dans un processus d'induction, l'expert seul (et c'est là l'intérêt principal de l'approche) utilise des exemples et des figures de cas et les saisit directement dans le logiciel. Les règles peuvent être dérivées automatiquement par le système, suite à la généralisation des exemples entrés. L'usage de ces logiciels inductifs s'accroît surtout lorsque le prototypage est utilisé comme modèle de développement. Ingénieur de la connaissance et experts les utilisent alors comme "faciliteurs".

2.2. Les modes d'acquisition : manuel ou automatique

Depuis le début de ce chapitre, nous avons surtout évoqué le mode d'acquisition qui est encore dominant, à savoir l'expert comme source et l'ingénieur de la connaissance comme transfert de la connaissance, et qui est essentiellement manuel.

Il faut également s'interroger sur les autres modes d'acquisition des connaissances qui présentent une variation continue du "tout manuel" vers le "tout automatique". On peut discerner cinq modes principaux, selon l'origine de l'expertise et la manière d'acquérir cette expertise :

Source	Mode
experts du domaine	ingénieur de la connaissance
données existantes	ingénieur de la connaissance
experts du domaine	éditeur de connaissances
experts du domaine	outil inductif
textes et documents	programme automatique d'acquisition

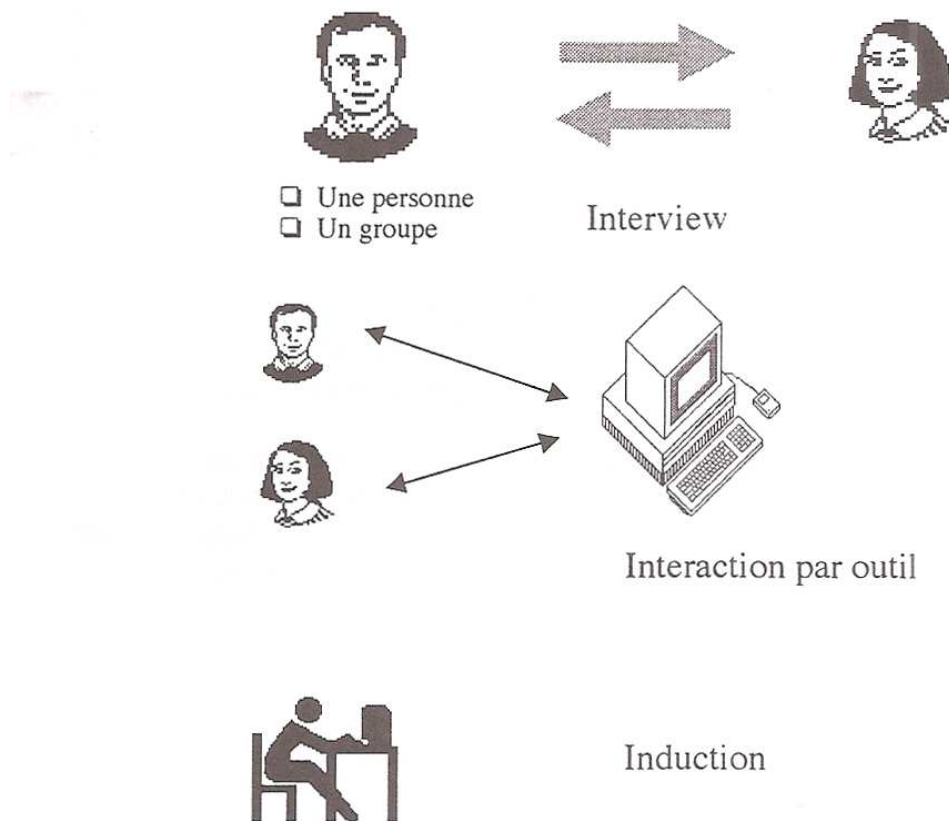


Figure 2.1. Les différentes approches pour l'acquisition des connaissances

Nous séparons les deux premiers modes, qui sont manuels dans leur nature, des trois suivants qui présentent un début d'automatisation.

2.2.1. Deux modes d'acquisition manuels

Les deux modes manuels s'expliquent d'eux-mêmes : un ingénieur de la connaissance, armé de crayons et de papier, entreprend l'acquisition de

connaissances avec un expert ou des sources d'informations telles que des documentations utilisateurs, documentations systèmes, vidéos...

2.2.2. Trois modes d'acquisition automatiques

Les autres modes font appel à l'automatisation de l'acquisition ou du moins à une réduction du rôle de l'ingénieur de la connaissance dans le processus en faveur de l'expert du domaine.

Le troisième mode fait référence à des programmes qui seraient des éditeurs de connaissances directement liés à la base du SBC final. Tout générateur de systèmes experts proposant une interface expert évoluée peut prétendre appartenir à cette catégorie. C'est là que se trouve la majorité des outils d'acquisition qui sont toujours spécifiques à un générateur de systèmes experts et à son formalisme de représentation.

Le quatrième mode est celui de l'induction que nous avons décrit en tant qu'approche possible au paragraphe précédent. Il semble toutefois que ces outils ne puissent convenir que pour des "petits" SBC, tant que le mécanisme d'inférence des règles restera strictement limité à l'acceptation d'exemples dans un format précis.

Le dernier mode n'est là que pour illustrer les promesses du futur où il y aurait vraiment acquisition et apprentissage par un programme intelligent. Cette notion d'apprentissage rejoint la théorie des réseaux de neurones. Un réseau neuronal acquiert plusieurs états stables simplement par présentation des données représentatives du cas. Les états stables deviennent alors les solutions vers lesquelles le réseau convergera lors de l'utilisation.

Les réseaux de neurones sont encore très primitifs et sont à l'opposé des SBC : ils ne contiennent aucune connaissance abstraite d'un domaine mais uniquement des données. Il y a donc un grand pas à franchir avant que la réalité d'un programme automatique d'acquisition ne soit effective. En attendant, l'acquisition des connaissances restera principalement une activité manuelle ou partiellement automatisée. La dispersion des outils et la présence majoritaire des méthodes livresques montrent bien que les obstacles prévisibles (d'ordre psychologique et cognitif) ne seront pas rapidement résolus. Il ne faut pas en conclure à un ralentissement des recherches. Celles-ci restent très actives et concernent tous les modes décrits.

Depuis 1986, date du premier colloque sur l'acquisition des connaissances de Banff (Alberta, Canada), de nombreux prototypes ont été conçus, ayant surtout pour principe d'aider l'ingénieur de la connaissance. Ils sont donc orientés vers l'édition plutôt que vers l'acquisition. Ils sont représentatifs des recherches allant dans la

direction du poste de travail de demain : le terme d'atelier de génie cognitif est également utilisé.

2.3. De l'expertise au code

Trois approches et cinq modes d'acquisition sont une première description du sujet. Il convient également de présenter les principales phases qui mènent de l'expertise à la base de connaissances codée. L'acquisition des connaissances recouvre spécifiquement les activités d'interviews et d'interactions entre ingénieur de la connaissance et experts.

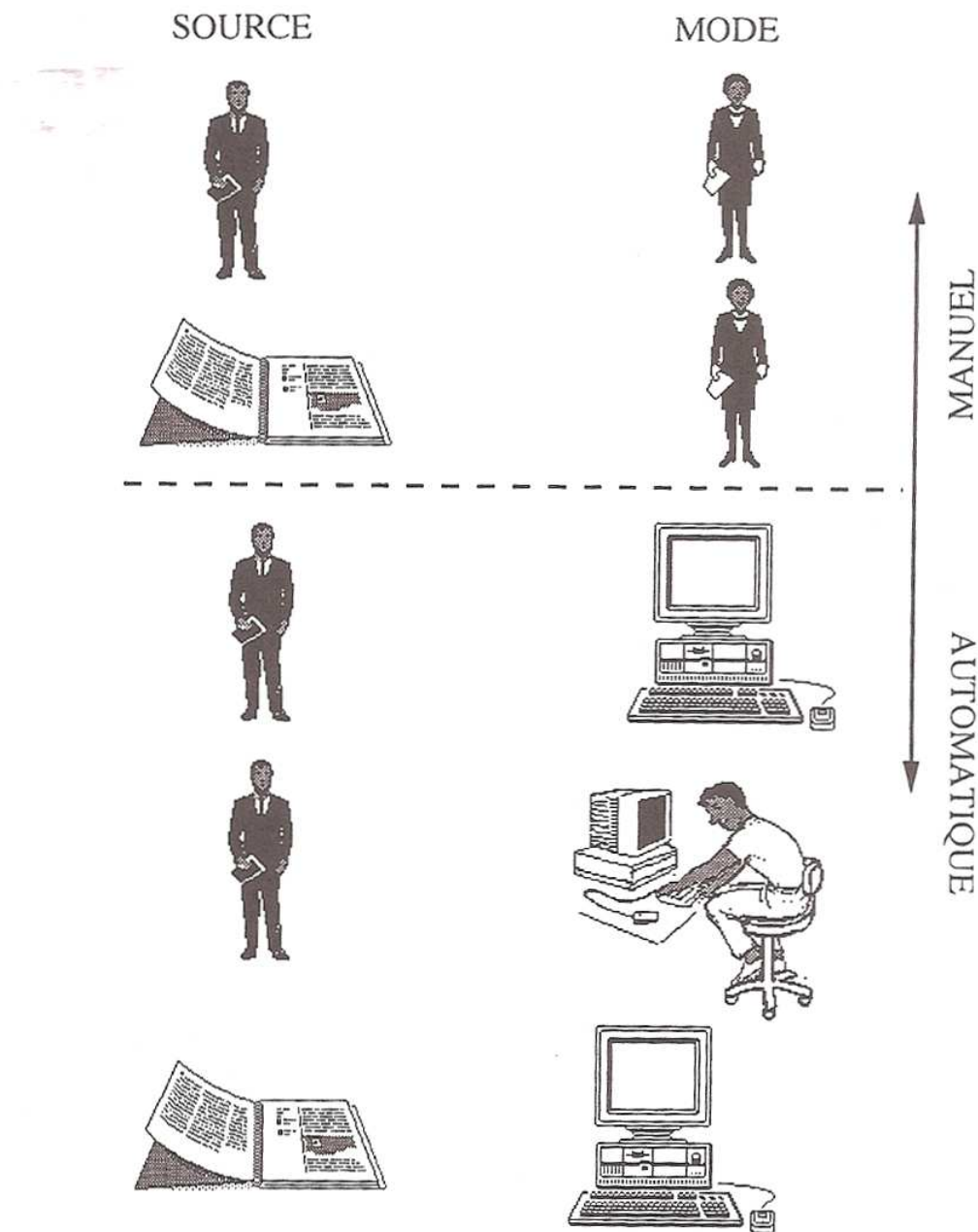


Figure 2.2. Modes d'acquisition manuels ou automatiques

Les cinq phases, qui sont maintenant acceptées et reprises dans toutes les méthodes proposées aux Etats-Unis, sont la première manifestation d'une approche structurée de l'acquisition des connaissances :

- l'identification,
- la conceptualisation,
- la formulation,
- l'implémentation,
- la validation.

Ces cinq phases nous conduisent de l'expertise au code contenu dans le SBC.

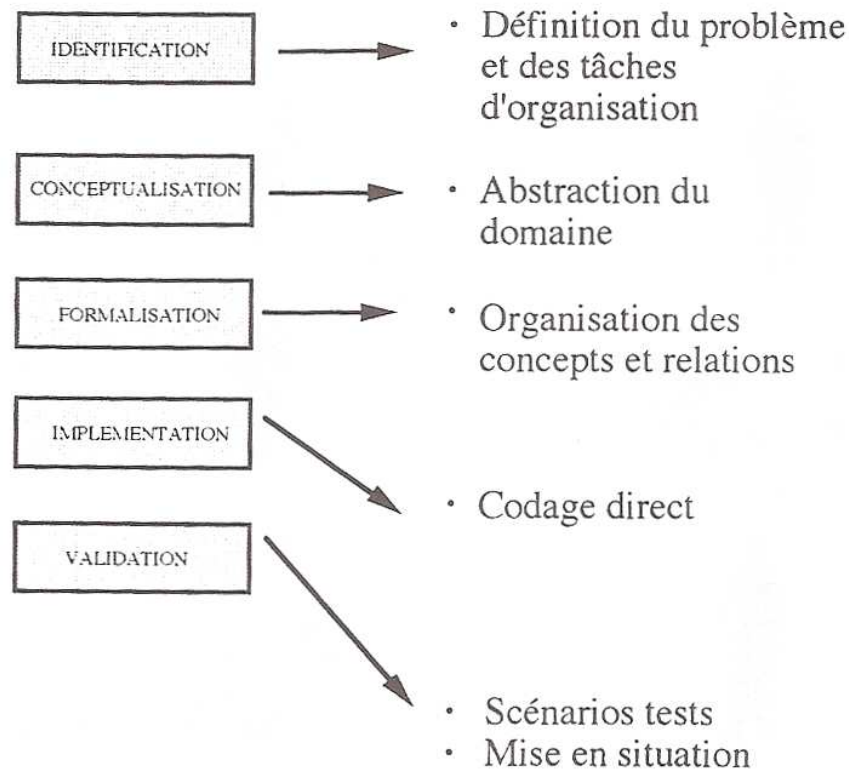


Figure 2.3. Les phases d'acquisition

2.3.1. L'identification

L'identification est la phase de description du problème à résoudre, des objectifs, des ressources et des contraintes inhérentes au projet. Elle correspond à la phase de spécifications dans les modèles de développement de logiciels. Durant cette phase, les tâches d'organisation (trouver les experts, définir les procédures de documentation et la forme utilisée pour la représentation des règles...) sont prédominantes.

Les techniques d'interviews non structurées et d'analyse de tâches sont couramment utilisées pendant l'identification. A l'issue de la phase, l'ingénieur de la connaissance aura identifié :

- les limites du domaine à explorer,
- les fonctions et les entrées/sorties qui apparaissent dans les tâches de l'expert,
- les types de connaissances manifestées par l'expert,
- les performances attendues du SBC.

2.3.2. La conceptualisation

La conceptualisation est souvent décrite comme la phase centrale de l'acquisition car elle semble la plus difficile à mener. Il s'agit de trouver les moyens et les objets pour décrire de façon abstraite le domaine (par définition le conceptualiser). Il s'agit d'une véritable quête pour trouver les bons concepts, les attributs, leurs valeurs et les hiérarchies représentatives du domaine. Cette phase n'est jamais facile même pour des domaines apparemment "conceptualisés" depuis longtemps (i.e. des domaines sur lesquels on a beaucoup écrit et théorisé) car ce qui constitue un concept ou une relation applicables au domaine dépend de l'objectif poursuivi.

Le but de la phase de conceptualisation est de produire des descriptions de catégories ou classes d'objets et de relations. Les techniques pertinentes sont connues sous le nom de graphe conceptuel, grille d'évaluation ou de répertoire, arbre hiérarchique... La plupart des outils informatisés se concentre sur l'aide à la conceptualisation du domaine d'expertise.

Il est pourtant illusoire de vouloir entièrement automatiser une activité aussi intellectuelle que celle d'identifier et de définir des concepts. L'apport principal de ces outils est de proposer un cadre de développement qui évitera les deux erreurs classiques commises pendant cette phase :

- itération excessive de la part de l'ingénieur de la connaissance qui a du mal à discerner à quel moment la conceptualisation devient suffisante pour entreprendre un début de formalisation ;
- tentative de cerner complètement le domaine avant de tester les schémas développés alors qu'il est important de confronter l'expert avec un prototype contenant les premiers concepts.

2.3.3. Formalisation et implémentation

Ces deux phases peuvent être regroupées. La formalisation est la phase de structuration et d'organisation des concepts et relations à l'aide d'un mécanisme de représentation formelle. Il s'agit en fait, plus que de sélectionner un formalisme (par exemple décider que les concepts peuvent être représentés par des prédicats logiques), d'identifier un modèle de traitement analogue à celui de l'expert et de le mettre en œuvre sous forme symbolique. Concevoir une représentation de concepts peut être un immense défi surtout lorsque ces concepts font intervenir des notions de

temps, d'incertitude et de croyances. Au cours de la phase d'implémentation, les connaissances sont mises sous une forme exploitable par l'outil de développement choisi. Ces deux phases sont souvent confondues car elles sont la partie la plus visible du processus de développement : la construction de la base de connaissances et le développement effectif du SBC.

C'est une erreur de conclure que l'acquisition des connaissances s'arrête là et que le développeur se substitue à l'ingénieur de la connaissance. Les tâches attribuées à ce dernier sont d'évaluer le mode de représentation choisi et de fournir la documentation détaillée qui donnera le lien entre le contenu de la base et la source de l'expertise. Durant ces deux phases, les connaissances acquises auprès des experts sont mises sous une forme opérationnelle, c'est-à-dire exécutable par un programme. Auparavant, les connaissances étaient sous une forme non-opérationnelle (discours, mots, écrits...) familière pour l'expert mais non implémentable sur machine.

Il est délicat d'affirmer qu'un domaine quelconque d'expertise soit directement opérationnel, soutenu par une théorie existante qui puisse inclure les connaissances de l'expert (le domaine des jeux est l'exception à cette règle). Aussi, la mise sous une forme opérationnelle conduit le plus souvent à une formulation étrangère à l'expert. Cela explique la relative pauvreté des outils automatisant les phases de formalisation et d'implémentation car tous sont basés sur l'existence d'un domaine déjà opérationnel.

L'approche directe "Extraire et Coder" qui peut servir de méthode de développement pour les "petits" SBC présente les mêmes défauts, issus de cette assumption. En ignorant les phases d'identification et de conceptualisation, l'ingénieur de la connaissance extrait directement des règles qu'il suppose opérationnelles et des faits qui ne seront jamais séparés en concepts et objets du domaine. En résumé, le codage, suite à l'énonciation de la connaissance par l'expert, fait la supposition d'un domaine déjà conceptualisé et d'un savoir opérationnel, ce qui est généralement faux. De plus, le choix de l'outil de développement dès le début du projet met encore plus de contraintes sur l'ingénieur de la connaissance qui finit par filtrer les interviews de l'expert pour ne garder que les règles directement implémentables sur l'outil choisi.

2.3.4. La validation

Cette phase est un passage obligé pour tout SBC devant être utilisé dans un contexte professionnel. Les tests n'ont de sens que s'ils valident en détail chaque phase antérieure et s'ils servent à revoir les défauts constatés à chaque étape et non pas uniquement l'implémentation. Il faut alors évaluer la formulation des concepts, l'efficacité des techniques d'acquisition et la validité de la formalisation. C'est la raison d'être de l'aspect itératif des méthodes d'acquisition des connaissances.

L'utilisation de cas et de scénarios tests est la voie couramment utilisée pour vérifier qu'un SBC donne les résultats que l'on attend de lui. La mise en situation devant les experts, avec les notions de compétition et de jeu que cette démarche induit, aboutit à un premier test du système. L'art de tester un système à base de connaissances est comparable à celui requis pour tester un système informatique complexe, avec en supplément la nécessité de prendre en compte ce qui fait l'originalité d'un SBC : la possibilité de donner des réponses approchées.

Après avoir décrit le paysage à plusieurs dimensions (approches – modes – phases) de l'aide à l'acquisition des connaissances, nous allons maintenant présenter les principales techniques qui s'y rapportent.