

INTRODUCTION

L'ingénierie des connaissances : acquis et nouvelles perspectives

1

Régine Teulier

CNRS, Centre de recherche en gestion, UMR 7655,
1, rue Descartes, 75005 Paris
teulier@ext.jussieu.fr

Jean Charlet

Mission de recherche en sciences et technologies de l'information médicale,
Assistance Publique / Hôpitaux de Paris,
91, bd de l'Hôpital, 75634 Paris cedex 13
& INSERM ERM 202
Jean.Charlet@spim.jussieu.fr

Pierre Tchounikine

Laboratoire d'informatique de l'université du Maine (LIUM),
avenue Laennec, 72085 Le Mans cedex 9
Pierre.Tchounikine@lium.univ-lemans.fr

1 OBJECTIF GENERAL DE L'OUVRAGE

L'objectif général de cet ouvrage est de proposer un état représentatif des travaux récents de l'ingénierie des connaissances en rassemblant une sélection d'articles produits par cette communauté de recherche et présentés à la conférence d'Ingénierie des connaissances (IC) au cours des trois années 1999, 2000, 2001. Placée sous l'égide du GRACQ (Groupe de recherche en acquisition des connaissances – <http://www.irit.fr/GRACQ/>), cette conférence annuelle est un lieu d'échange et de réflexion de la communauté française sur les problématiques de l'ingénierie des connaissances.

L'ouvrage précédent du même type¹, qui regroupait des articles produits entre 1995 et 1998, ayant rencontré un écho très positif, nous avons souhaité rendre compte des évolutions du domaine depuis sa parution en 2000 : plusieurs années s'étant écoulées, certaines orientations sont apparues, d'autres se sont raréfiées.

L'ouvrage rassemble des articles issus d'un double processus de sélection. Cent soixante articles avaient été soumis, quatre-vingts ont été acceptés (relus pour leur grande majorité par trois lecteurs) par les comités de programme des trois conférences IC (dont le taux habituel de sélection des articles est de un sur deux). Sur ces quatre-vingts articles présentés aux conférences IC'99 (École polytechnique, Palaiseau), IC 2000 (IRIT, Toulouse), IC 2001 (INPG, Grenoble), vingt et un ont été choisis par les trois présidents de conférence, coordinateurs de cet ouvrage, comme représentatifs de l'état des travaux et des avancées dans les différents domaines de l'ingénierie des connaissances.

2 L'INGENIERIE DES CONNAISSANCES, ISSUE DE L'INTELLIGENCE ARTIFICIELLE

L'ingénierie des connaissances est une discipline en émergence et il faut peut-être consacrer quelques lignes à une lecture de cette émergence. Dans les années 1980, l'intelligence artificielle bénéficie d'une certaine ouverture et d'une notoriété qui dépasse soudainement le cadre des laboratoires. Les capacités des machines augmentant d'une manière importante et les premières vagues de l'informatisation étant maîtrisées, il devient possible d'envisager des applications industrielles en écho aux recherches académiques, dont certaines peuvent dès lors être infléchies. La diffusion des systèmes experts, technologie phare de cette double vague, permet de produire des applications sur des micro-mondes. Cette tentative, qui utilise toutes les

¹ CHARLET J., ZACKLAD M., KASSEL G. & BOURIGAULT D. (éd.), *Ingénierie des connaissances, évolutions récentes et nouveaux défis*, Eyrolles, Paris, 2000.

techniques de l'intelligence artificielle et vise à l'appliquer à de nombreux domaines, connaît des réussites diverses. Beaucoup de ces techniques sont ensuite intégrées, absorbées par l'informatique en entreprise ou par celle des composants logiciels, d'autres connaissent moins de débouchés, même si les recherches continuent en laboratoire.

Depuis, une partie de la recherche en intelligence artificielle s'est poursuivie sur des sujets plutôt formels, en vue de représenter des types de raisonnement complexes et de mieux gérer des connaissances imprécises ou contradictoires dans des environnements incertains. Quant aux chercheurs qui ont persévéré en dépit de la difficulté de modéliser les connaissances et les processus cognitifs en situation réelle, ils ont construit en France un axe de recherche d'abord intitulé « acquisition de connaissances », puis « ingénierie des connaissances », et, au niveau européen ou nord-américain, « *knowledge acquisition* » ou « *knowledge engineering* ». Si l'on considère que les travaux présentés dans des conférences comme « Reconnaissance des formes et intelligence artificielle » (RFIA)² ou « European Conference of Artificial Intelligence » (ECAI) sont représentatifs des thèmes de recherche de l'intelligence artificielle (apprentissage automatique, représentation des connaissances, démonstration automatique, langages formels, décision et incertitude, diagnostic, etc.), il apparaît que les échanges avec l'ingénierie des connaissances sont restreints. Finalement, peu d'avancées récentes en intelligence artificielle sont utilisées en ingénierie des connaissances, et réciproquement. Il ne faut pas pour autant en déduire que le lien avec l'intelligence artificielle serait maintenant seulement historique. Cet éloignement n'est peut-être qu'apparent et conjoncturel, probablement accentué par le fait que le défi de l'ingénierie des connaissances (modéliser les connaissances dans des situations réelles) exige de diversifier les collaborations disciplinaires, par exemple avec la terminologie, l'ergonomie ou la gestion, des sciences humaines et sociales (SHS), selon la terminologie consacrée.

L'une des façons de présenter l'évolution de l'ingénierie des connaissances est de considérer que, de la modélisation de méthodes de résolution de problème et de la simulation en machine de raisonnements humains, puis de savoir-faire d'experts, on est passé à l'ambition d'assister un utilisateur. Cette mutation des ambitions et des objectifs a pris quelques années. Les travaux se focalisant sur des méthodes d'ingénierie et d'assistance à l'utilisateur ont en effet pris le pas sur la définition de méthodes et d'outils formels tels qu'on les retrouve en intelligence artificielle. Cependant, les thèmes de recherche de l'ingénierie des connaissances et de l'intelligence artificielle convergent parfois à nouveau, comme c'est le cas actuellement avec la représentation des connaissances pour le Web sémantique.

Dès son origine, l'ingénierie des connaissances a choisi de se confronter au problème concret des systèmes informatiques opérationnels. Cette orientation lui a permis de poser le problème de la modélisation des connaissances et des processus

² Site de la conférence RFIA 2004 tenue en janvier 2004 à Toulouse : <<http://www.laas.fr/rfia2004/>>.

cognitifs à assister ; c'est là que résident son identité comme champ de recherche autonome et son apport principal. Elle s'est attaquée de front à la question des connaissances en usage, de leur partage et de leur diffusion, là où d'autres spécialités de l'intelligence artificielle, comme les réseaux de neurones, ont retenu des hypothèses fondamentalement différentes, à savoir que les connaissances sont difficiles et trop coûteuses à représenter. Leurs modélisations ayant toujours un domaine de validité restreint et contextualisé, elles sont peu sujettes à production scientifique. Le fait que l'ingénierie des connaissances ne prétende pas simuler l'intelligence, mais produire un artefact qui fasse sens dans un contexte d'usage, la situe sur un autre paradigme représentationnel. Elle centre ses travaux sur la modélisation non opérationnelle des connaissances, puis sur leur implémentation en machine : modélisation de la tâche à réaliser par le couple opérateur/système, pour mieux en définir les rôles respectifs et les interactions ; modélisation du processus de résolution de problème mis en œuvre par le système pour réaliser des parties de cette tâche globale ; modélisation des connaissances du domaine, plus ou moins en lien avec la représentation du processus de raisonnement.

3 UN CHAMP DISCIPLINAIRE ORIGINAL

L'ingénierie des connaissances s'impose donc comme une discipline jeune, en quête de ses propres fondements théoriques (l'artefact plongé dans un système d'usage et associé à un schème d'utilisation, le paradigme représentationnel des inscriptions numériques de connaissance³, l'engagement ontologique), avec son domaine d'utilisation et ses liens avec d'autres disciplines. Parler d'une ingénierie de la connaissance est une affirmation forte : c'est dire que la connaissance peut s'appréhender et se travailler comme un produit, c'est-à-dire qu'elle est non seulement modélisable, mais aussi qu'un ensemble de dispositifs professionnels concernant cette modélisation permet de garantir des procédés, des résultats, des évaluations de réalisation. L'objectif global du champ disciplinaire est bien de fournir progressivement aux entreprises les éléments d'une gamme de méthodes, d'outils, de références, permettant de réaliser les applications du traitement de la connaissance dans les activités humaines. L'ingénierie des connaissances vise à proposer des concepts, méthodes et techniques permettant de modéliser, de formaliser, d'acquérir des connaissances dans les organisations dans un but d'opérationnalisation, de structuration ou de gestion au sens large.

Plonger le système à base de connaissances et ses interfaces dans le monde de l'utilisateur est un enjeu essentiel de l'ingénierie des connaissances. Le caractère local des ontologies⁴, l'orientation de la modélisation des connaissances par rapport

³ Voir à ce sujet le mémoire d'habilitation à diriger des recherches de Bruno Bachimont.

⁴ Alors que les premières ontologies envisagées ont été initialement construites en intelligence artificielle pour faciliter la réutilisation et l'interopérabilité entre systèmes, indépendamment de la tâche réalisée par

aux tâches de l'utilisateur, trouvent leur prolongement et une traduction plus tangible pour l'utilisateur à travers leur concrétisation dans l'application qui va les manipuler. C'est à travers l'insertion dans un poste de travail et dans ce lien avec l'activité que la modélisation des connaissances prend tout son sens, que les connaissances modéli-sées peuvent s'insérer dans une nouvelle activité, redéfinie avec l'outil. L'artefact devient instrument lorsque l'acteur l'associe à un schème d'utilisation⁵; ce faisant, il se l'approprie et l'artefact est instrumentalisé. Cette étape ne va pas toujours sans problème. C'est une partie importante et difficile des tâches d'ingénierie que de travailler sur la modification de l'activité que va entraîner l'outil et sur les conditions de l'insertion dans les usages.

L'objectif consistant à construire un système à base de connaissances ayant un comportement compréhensible et acceptable par l'utilisateur passe par une modélisation à un niveau d'abstraction pertinent qui fait sens pour les différents acteurs impliqués dans sa conception (ingénieurs cogniticiens, experts métiers, utilisateurs, etc.). En phase d'utilisation du système, le modèle est rendu opérationnel de manière à ce que l'utilisateur s'approprie le comportement du système et puisse interagir avec lui. Ce niveau de modélisation pertinent est le « niveau des connaissances », dont l'importance et les articulations avec les niveaux informatiques ont été mis en évidence par Newell⁶. Cela suppose aussi que soient appréhendés le fonctionnement et l'utilisation d'un système à base de connaissances par rapport à son intégration dans une organisation. L'ingénierie des connaissances doit permettre de modéliser et d'explicitier le fonctionnement cognitif de l'utilisateur ou du collectif dans sa relation au système à base de connaissances.

Parce que la connaissance implique une interprétation, on ne peut la modéliser sans se préoccuper de son contexte d'utilisation, des interactions et coopérations d'acteurs dans lesquels elle est utilisée, et des contextes organisationnels de ceux-ci. C'est ce qu'affirment aussi bien les épistémologies constructivistes que d'autres approches théoriques, comme celle de la cognition située : les connaissances sont avant tout connaissances pour l'action, elles sont situées dans un contexte physique, socialisées dans un réseau d'acteurs, interprétées grâce à leurs capacités autant perceptives que de raisonnement. Elles prennent place et s'enrichissent dans les organisations qu'ils construisent et qu'elles font évoluer en permanence. Ces mêmes connaissances sont transmises comme des informations destinées à être, *in fine*, interprétées par un hu-

main dans son interaction avec l'artefact – c'est-à-dire le système à base de connaissances – construit.

eux, les expériences ont montré que les ontologies les plus pertinentes pour des applications étaient spécialisées et influencées par la tâche considérée (on parle d'« ontologie régionale »). C'est à ce type d'ontologie que nous faisons allusion.

⁵ Rabardel P., *Les Hommes et les technologies*, Armand-Colin, Paris, 1995.

⁶ NEWELL A., « The Knowledge Level », *AI Journal*, 19 (2), 1982.

La démarche modélisatrice de l'ingénierie des connaissances s'inscrit donc toujours en situation réelle. Et cela ne manquera pas de frapper le lecteur à travers la lecture des travaux présentés ici : les projets qui ont servi à dégager une production scientifique sont toujours des réalisations pour des acteurs donnés dans des situations de travail impliquant un ensemble de contraintes. Les exemples jouets ont quasiment disparu. Ainsi, l'ingénierie des connaissances trouve matière à recherche partout où l'on travaille sur l'instrumentation technique des connaissances pour leur exploitation dans un cadre où elles sont mobilisées avec leur signification et leur interprétation. Par ailleurs, elle trouve son champ d'application là où l'on veut modéliser les connaissances, les processus cognitifs, les processus de coopération et les savoir-faire d'un métier.

Produire des connaissances scientifiques sur une ingénierie, c'est produire à la fois des connaissances pour l'action et les fondements théoriques de ces connaissances, les outils conceptuels et les méthodologies qui constituent le cadre de leur production. Les sciences de l'ingénierie ont ceci de particulier qu'elles se fondent sur les méthodes, sur l'art de la conception, qu'elles s'intéressent à l'intelligence des procédés, intelligence humaine en œuvre dans les réalisations, se situant ainsi parmi les modélisations constructivistes de systèmes complexes. Nous affirmons que mettre en place une réelle démarche d'ingénierie, reproductible et proposant des instrumentations techniques – c'est-à-dire des méthodes et des outils – favorisant la dynamique des connaissances dans l'organisation est réalisable (et cette production collective le confirme) et pertinent. La recherche est alors souvent « *bottom up* », partant de la complexité de situations réelles et cherchant, tout en apportant une réponse concrète aux cas traités, à faire œuvre scientifique en faisant apparaître un problème générique, puis scientifique, et en développant des méthodes et des outils utilisables dans d'autres situations.

4 DE RECENTES EVOLUTIONS

L'ingénierie des connaissances est un champ disciplinaire jeune. Néanmoins, plusieurs années de travaux permettent de constater que la communauté des chercheurs de ce domaine a pu accumuler des travaux et constituer des acquis sur lesquels un savoir-faire collectif est transmissible.

Les recherches sur les **ontologies**, thématique très active depuis 1998, ont accumulé des résultats souvent repris, tant sur les langages pour la représentation des connaissances qu'elles contiennent que sur l'évaluation de leur apport à des applications aussi variées que la gestion des connaissances des entreprises, la recherche d'information sur le Web ou la gestion documentaire. Au niveau international, la standardisation des langages (en lien avec le W3C), la réutilisation ou l'intégration d'ontologies, ainsi que l'étude de leurs propriétés formelles, sont les thèmes actuels des recherches. Au niveau national, l'accent a été mis dès le début sur l'outillage,

théorique, technique et méthodologique, de leur construction. Les résultats établis par ces travaux, menés en collaboration avec des linguistes ou des philosophes, soulignent l'intérêt qu'il y a à s'appuyer sur des textes pour construire des ontologies ayant un ancrage terminologique. On dispose désormais d'outils automatisant des parties de la chaîne de traitement, des critères d'identification des connaissances à retenir ou de validation d'une ontologie sur un domaine et pour une application donnés.

L'**acquisition** des connaissances auprès d'experts fait actuellement l'objet de moins de travaux. Cette phase de la construction d'un système à base de connaissances est aujourd'hui mieux maîtrisée : un ensemble de techniques a été proposé et validé pour des situations variées d'utilisation avec des objectifs différents, tant par des chercheurs de l'ingénierie des connaissances que par des chercheurs d'autres champs disciplinaires et des praticiens.

De même, après avoir beaucoup focalisé les travaux de la communauté, notamment européenne, les **méthodes** de développement de systèmes ont accumulé beaucoup de résultats. Aujourd'hui, un chef de projet peut choisir entre plusieurs méthodes de développement qui détaillent les différents composants d'une application et le cycle de vie du projet. Ces méthodes ont été appliquées, validées et comparées sur plusieurs projets. Cependant, la maîtrise du coût et de la durée de ces projets doit pouvoir être encore améliorée. Pour cela, certaines méthodes favorisent la réutilisation de fragments de modèles lorsqu'elle est possible. À cette fin, et c'est là un autre type de résultat bien établi, elles proposent des principes de réutilisation et des bibliothèques de modèles de méthodes de résolution associés à des types de tâches qui se retrouvent dans différentes applications.

Les **langages** de formalisation, qui permettent de rendre les modèles opérationnels, d'effectuer les opérations nécessaires d'agrégation ou de vérification de cohérence, sont actuellement l'objet de travaux plus dispersés. Les plus récents sont mis au point en collaboration avec des chercheurs de la communauté « représentation des connaissances », en particulier les langages de représentation d'ontologies pour le Web sémantique.

Si l'ingénierie des connaissances compte parmi ses thèmes traditionnels des réflexions sur les langages de représentation des connaissances, sur la modélisation des méthodes de résolution de problème et l'adaptation d'approches cognitives pour accéder à des savoir-faire d'experts, ces thèmes ont été relativisés au cours de ces trois années. Les **évolutions** des thèmes de recherche se dessinent assez bien à travers les conférences successives. Ces dernières années ont vu l'approfondissement des approches communes à l'ingénierie des connaissances, l'épistémologie, la conception ou l'ergonomie. Les échanges avec les chercheurs en gestion ayant accumulé des travaux sur les organisations et les connaissances dans ces organisations, parfois sans utiliser le terme de connaissances (apprentissage, représentations pour l'action, compétences, etc.) ont permis d'éclairer l'environnement organisationnel des systèmes à base de connaissances.

L'ingénierie éducative et la modélisation du travail coopératif sont des thèmes voisins abordés dans des communautés spécifiques, mais qui rejoignent des préoccupations générales de modélisation des connaissances. Enfin, de nouveaux thèmes, comme le Web sémantique et l'aide à la conception, apparus en 2001 dans plusieurs articles, semblent très prometteurs.

5 STRUCTURE DE L'OUVRAGE

L'ouvrage est structuré en quatre parties. Les articles regroupés dans chaque partie présentent certaines caractéristiques communes que nous avons reliées à ce qui nous apparaissait comme une grande thématique. Ce regroupement, non explicité par les auteurs, relève d'une « lecture » des travaux de la communauté et n'a pour seul but que de faciliter la mise en perspective de ces travaux. Certains articles auraient pu se trouver dans l'une ou l'autre partie, et d'autres regroupements auraient été possibles.

5.1 Analyse de corpus et construction d'ontologies

La dynamique créée autour du thème des ontologies, en accroissement ces dernières années, continue à rassembler les travaux de plusieurs équipes très actives. En France, une grande partie de ces recherches s'intéressent à la construction des ontologies à partir de textes. L'analyse terminologique d'articles des dernières conférences d'ingénierie des connaissances présentée au chapitre 2 confirme cette tendance. Ces recherches bénéficient de collaborations avec le traitement automatique des langues, la linguistique de corpus et la terminologie, laquelle se trouve renouvelée et stimulée par ses échanges avec l'ingénierie des connaissances.

Le chapitre 2, de D. Bourigault et J. Charlet, part des résultats fournis par un outil de traitement automatique, LEXTER, qui organise l'ensemble des candidats termes extraits d'un corpus en un réseau terminologique. La construction d'index à partir de ce réseau se décompose alors en quatre étapes : sélection des entrées, structuration de l'index, filtrage des occurrences, choix du niveau des renvois. Le choix des entrées reste subjectif et constitue une tâche d'ingénierie de la connaissance.

Le chapitre 3, de N. Aussenac-Gilles *et al.*, identifie les étapes, outils et méthodes à appliquer pour dégager un modèle conceptuel du domaine à partir de l'analyse d'un corpus en utilisant des outils de traitement automatique des langues. Pour tester cette méthode, les auteurs construisent une ontologie des outils de l'ingénierie des connaissances à l'aide des logiciels TERMINAE, LEXTER et CAMELEON. Les étapes identifiées sont la constitution du corpus, l'extraction automatique des termes, l'étude linguistique (validation des candidats termes, puis étude des relations), la normalisation (élimination et regroupement des termes, puis structuration des

concepts), enfin la formalisation ou la traduction des concepts et relations sémantiques dans le langage du système à base de connaissances.

Le chapitre 4, de E. Morin et C. Jacquemin, a pour objectif d'offrir une meilleure complémentarité entre acquisition sémantique et acquisition terminologique. Les auteurs proposent une méthode pour la structuration automatique de termes polylexicaux qui s'appuie, non sur des régularités syntaxiques, mais sur des liens entre mots simples extraits automatiquement à partir de corpus. Les deux principaux modes d'expansion utilisés des relations entre termes simples vers des relations entre termes complexes sont le transfert et la spécialisation. Une acquisition de liens d'hyponymie à partir de corpus de textes est proposée en combinant plusieurs approches.

Le chapitre 5, de B. Bachimont, s'intéresse à la numérisation des contenus audiovisuels et à l'informatisation de leurs traitements (stockage, catalogage, indexation et publication des documents audiovisuels). À partir de deux paradigmes informatiques, l'ingénierie documentaire et l'ingénierie des connaissances, il propose une instrumentation de l'indexation audiovisuelle autorisant la description des documents audiovisuels (autour du format MPEG-7) en contrôlant tout au long de la chaîne documentaire la sémantique des descripteurs mobilisés à l'aide d'une ontologie.

Le chapitre 6, de C. Reynaud *et al.*, traite du processus de représentation d'une ontologie du domaine dans un serveur d'information (PICSEL) qui détermine les sources d'information pertinentes pour répondre à une requête utilisateur. La première étape est guidée par le langage et les fonctionnalités du serveur d'information. La seconde étape, qui vise à affiner et réorganiser les connaissances représentées à la première étape, est guidée par la façon dont les fonctionnalités du serveur sont mises en œuvre.

5.2 Sciences et technologies de l'information et de la communication

La contribution de l'ingénierie des connaissances aux sciences et technologies de l'information et de la communication s'affirme de plus en plus forte ; c'est pourquoi nous avons choisi d'en faire une thématique spécifique. La contribution de l'ingénierie des connaissances aux recherches des divers champs de l'informatique est importante et se développe, ce qui dénote la généralisation de l'évolution d'une société de l'information vers une société de la connaissance. Concernant le Web sémantique, nous pouvons déjà constater que l'ingénierie des connaissances apporte, par ses réflexions et ses méthodes (ancrage linguistique, représentation des connaissances pour les ontologies, utilisation d'ontologies pour l'indexation, etc.), des solutions originales et efficaces. La contribution de l'ingénierie des connaissances est également importante pour le traitement des documents. Le fait de traiter les documents en fonction de leur contenu implique la mise en œuvre de

connaissances. Dans le domaine éducatif, l'émergence du terme EIAH (pour « Environnement informatique pour l'apprentissage humain ») dénote l'accent mis sur les coopérations homme-machine et sur les activités des apprenants à travers des environnements informatiques spécialisés (activités collectives, etc.). De plus en plus de travaux abordent les problématiques de l'ingénierie de ces EIAH, participant à ce mouvement général pour l'élaboration des méthodes d'ingénierie. Le raisonnement à partir de cas prend part au même mouvement par ses travaux de modélisation des connaissances sur la classification des cas en populations ordonnées.

Dans le chapitre 7, F. Gandon et R. Dieng-Kuntz proposent un système multi-agents dédié à une mémoire d'entreprise : COMMA. Ils utilisent le paradigme des systèmes multi-agents, adapté au déploiement d'une configuration logicielle au-dessus d'informations distribuées dans l'intranet d'une entreprise, ainsi qu'une ontologie qui permet une communication de haut niveau au sein de ce nouveau type de système d'information. La solution proposée s'apparente donc à un « Web sémantique » à l'échelle de l'entreprise. Les étapes de la conception de l'ontologie ainsi que l'utilité des scénarios d'utilisation sont présentées en détail. Le formalisme RDF(S) et les feuilles de style permettent de naviguer dans l'ontologie et la mémoire au travers des niveaux terminologiques, intensionnels et extensionnels.

Dans le chapitre 8, qui se situe dans la perspective de la mise à disposition de grandes bases de données à travers le Web, M.-S. Segret *et al.* s'intéressent à l'extraction et à l'intégration d'informations semi-structurées dans les pages Web. L'objectif est d'identifier les sites qui traitent d'un même service. Le problème est de lier les deux types de données (structurées et textes) et de les décrire à l'aide du même formalisme, qu'ils proviennent ou non du même site. La solution retenue s'appuie sur les ontologies et les agents informationnels.

B. Fuchs et A. Mille proposent dans le chapitre 9 un cadre de modélisation pour décrire un modèle générique du raisonnement à partir de cas au niveau connaissance, en s'appuyant sur le concept de tâches de raisonnement avec deux modèles de tâches complémentaires : un modèle de décomposition de tâches et un modèle de spécification des tâches. Les modèles de tâches mettent en évidence les différentes catégories de connaissances impliquées, les mécanismes d'inférence mis en œuvre et la structuration hiérarchique des tâches en sous-tâches. Un modèle générique de la phase de remémoration du raisonnement à partir de cas est décrit, et CREEK, système d'aide au diagnostic, est vu comme une spécialisation du modèle générique de la remémoration.

Dans le chapitre 10, F. Tort *et al.* comparent ingénierie des besoins et ingénierie des connaissances en partant des méthodologies EKD et CommonKADS. Ils font émerger les complémentarités et les coopérations éventuelles. EKD traite du changement organisationnel, pour lequel cette méthode propose des modèles de scénario alternatifs. La méthode CommonKADS modélise des processus à granularité plus fine. Elle propose une étude organisationnelle qui est une étude

d'opportunité pour déterminer les tâches à forte composante de connaissances. C'est cette complémentarité d'échelle qui paraît la plus prometteuse pour des études ultérieures en organisation.

Dans le chapitre 11, J. Bouaud et B. Séroussi constatent que la faible observance des guides de pratiques s'explique par leur nature de formalisation hors contexte de notions médicales contextuelles. Ils proposent une démarche de modélisation et d'opérationnalisation de l'expertise (Oncodoc) à la fois formelle (les connaissances décisionnelles sont représentées par un arbre de décision qui impose un parcours) et informelle (elle n'est pas exécutée par un programme, mais destinée à être lue sous forme d'hypertexte). Le médecin évalue lui-même l'appréciation à porter sur chaque critère constituant un nœud de l'arbre. Ce système a été validé sur l'expertise concernant le cancer du sein dans un service hors cadre de conception.

Dans le chapitre 12, qui aborde l'annotation de documents informels à l'aide de représentations formelles, J. Euzénat note que le type d'annotation possible est varié et conduit à des traitements diversifiés. Il répertorie huit questions qui peuvent guider le développement d'un système cohérent. Elles concernent la forme et l'objet des représentations retenues, la nécessité d'utiliser de la connaissance indépendante du contenu des documents (ontologie, connaissance de contexte) et le statut du système résultant (système à base de connaissances ou éléments de connaissance distribués).

5.3 Gestion des connaissances

Le thème « Gestion des connaissances » regroupe des articles dont les auteurs ont pour objectif de gérer explicitement les connaissances dans une organisation. Ils proposent différentes méthodes, utilisant des techniques allant des livres de connaissances aux ontologies pour décrire et structurer les connaissances d'un domaine. Ils s'intéressent à différents problèmes rencontrés dans les organisations comme la veille technologique, la gestion de code ou la conception et la gestion d'une équipe de recherche.

Le chapitre 13, de G. Kassel *et al.*, propose un système de gestion des connaissances vu comme un système documentaire et fondé sur un système à base de connaissances multitâches. Son architecture logicielle est présentée, ainsi qu'une méthode de développement de ce système à base de connaissances. La méthode inclut une méthode de construction d'ontologie qui distingue l'ontologie conceptuelle de l'ontologie computationnelle. Un exemple de construction d'ontologie est donné, ONTO-ORG, constituée de plusieurs sous-ontologies, dont une ontologie de documents.

Le chapitre 14, de S. Moisan et J.-L. Ermine, aborde la chaîne des connaissances pour la gestion de codes spécialisés de bas niveau par des professionnels non informaticiens, de leur conception et de leur développement jusqu'à leur utilisation. Ils proposent deux nouvelles techniques utiles dans cette chaîne : d'une part, des livres

de connaissances, descriptifs de toutes les connaissances et savoir-faire nécessaires au traitement des codes ; d'autre part, des outils de pilotage de code construits à partir des livres de connaissances, intégrant des connaissances sur l'utilisation des codes et aidant à la réalisation des tâches métier.

Dans le chapitre 15, S. Mahé présente le système PUMEO, réalisé dans le contexte de l'industrie électrotechnique, dans lequel il choisit, plutôt qu'une formalisation des connaissances, une modélisation de leur contexte d'utilisation, à travers des règles de notification dans un agent automatisé. Cela permet d'activer des échanges entre les personnes et d'augmenter la réutilisation et le partage des connaissances, notamment en sollicitant automatiquement par mail les participants aux mêmes thèmes d'étude.

Dans le chapitre 16, M. Lewkowicz et M. Zacklad proposent un logiciel de structuration des échanges (MEMO-NET) qui s'ajoute à un logiciel d'aide à la résolution de problème en mode distribué et facilite l'accès et le partage des connaissances. Pour un travail collectif à distance, cette assistance gère plus efficacement à la fois les interdépendances liées à la ressource partagée qu'est le temps de parole et celles liées aux représentations partagées de l'espace de problème, améliorant ainsi les échanges par rapport à un mode d'échange faiblement structuré, comme un forum ou une réunion à distance.

Dans le chapitre 17, P. Benhamou *et al.* proposent une méthode pour optimiser les interactions du patrimoine de connaissances de l'entreprise avec le système de veille scientifique, technique ou économique. Testée sur deux cas concrets, au Commissariat à l'énergie atomique et chez Renault, cette méthode part d'une modélisation du domaine et permet de dégager des axes stratégiques de recherche d'information afin d'organiser la veille en trois phases : correspondance entre l'organisation et son environnement, recueil d'information, puis création de connaissances.

5.4 Modélisation de l'activité

Le thème « Modélisation de l'activité et de l'organisation » rassemble des travaux où l'activité proprement dite des acteurs est au cœur des études, au-delà des textes et des termes utilisés dans ces textes. Les auteurs intéressés par ce thème se donnent pour objectif d'assister l'activité, d'en concevoir une nouvelle à partir de l'utilisation d'un nouvel instrument, ou de catégoriser en grands types une activité propre à un métier.

Dans le chapitre 18, R. Teulier et N. Girard proposent une réflexion sur le lien entre activité, connaissances et organisation. Elles soulignent l'intérêt, pour l'ingénierie des connaissances, de continuer à s'intéresser à l'activité. Les connaissances étant systématiquement mobilisées pour l'action et situées dans l'environnement physique et social de l'acteur, la construction d'un système à base de connaissances destiné à être instrumentalisé dans la pratique d'un utilisateur doit intégrer une certaine modélisation de l'activité. Les approches complémentaires de différents champs

disciplinaires sont comparées et situées par rapport à la démarche modélisatrice de l'ingénierie des connaissances.

Le chapitre 19, de G. Prudhomme *et al.*, discute d'une nouvelle classe d'outils d'aide à la conception. En s'appuyant sur des travaux empiriques sur la coopération de la conception, il propose une évolution vers des outils intégrant les connaissances locales, notamment celles créées par des acteurs de métier différent qui confrontent leurs règles métier et créent des connaissances nouvelles. Ces connaissances émergent souvent dans l'activité de simulation, mais aussi aux interfaces entre conception et simulation.

Le chapitre 20, de M.-L. Betbeder et P. Tchounikine, s'intéresse à la conception d'une ACCA (activité médiatisée collective dans un contexte d'apprentissage) avec un objectif de capitalisation des connaissances. L'alternance des phases de travail individuelles et collectives, synchrones et asynchrones, présentes ou à distance de la communauté d'apprenants est particulièrement étudiée et discutée. Les considérations à prendre en compte lors de la conception de tels systèmes sont identifiées avec l'objectif de mise en œuvre d'une démarche méthodologique.

Le chapitre 21, de N. Girard, en combinant la théorie des représentations prototypiques de Rosch et la technique des grille répertoire de Kelly utilisée en acquisition des connaissances, opère une classification des comportements techniques d'agriculteurs utilisant leur territoire et leur exploitation pour une gestion des pâturages. La formalisation de prototypes de comportement ainsi que la démarche associée sont détaillées et les aspects méthodologiques d'une catégorisation de tels « objets » sont discutés.

Enfin, dans le chapitre 22, F. Darses donne un ensemble de recommandations appuyant les modèles de l'expertise de conception et les systèmes à base de connaissances d'assistance à la conception dite « créative », et visant les spécificités cognitives de l'activité de conception : disposer de bases de simulation associées à tous les niveaux d'élaboration de la solution, faire élaborer simultanément les différents points de vue sur l'objet, permettre une modification itérative des pondérations des contraintes, aider à la construction des critères d'évaluation et tracer les contraintes, documenter des bases de cas et, enfin, permettre la construction opportuniste de la solution. Les spécificités de la conception collective sont également abordées.

6 INDEXATION ET REALISATION DU CÉDÉROM

Comme pour le livre précédent¹, la constitution de l'index a été l'occasion de mettre en œuvre une méthode spécifique permettant d'expérimenter de nouveaux services d'indexation. De la même façon que précédemment⁷, nous avons travaillé à partir du corpus des textes des vingt et un articles constituant l'ouvrage. Ce travail a été réa-

lisé dans un contexte entièrement numérique, c'est-à-dire à partir de fichiers numériques informatiques, le but étant de constituer la collection des articles de l'ouvrage en un ensemble de fichiers HTML que l'utilisateur peut consulter via un navigateur. Ce travail tire parti des expériences acquises par les auteurs :

- La constitution d'un index pour le livre sur l'ingénierie des connaissances, regroupant trente-cinq articles des années 1995-1998, effectuée par D. Bourigault et J. Charlet⁷. Plutôt que de faire appel aux auteurs des articles, ce travail innovait en exploitant les résultats fournis par un outil de traitement automatique des langues, l'analyseur syntaxique de corpus Syntex⁸. Si le repérage des candidats termes nous est rapidement apparu comme un point de départ fondamental, l'expérience nous a permis d'identifier un certain nombre de difficultés, à commencer par la quantité de candidats termes repérés et la difficulté à trouver un critère autre que quantitatif pour décider lesquels retenir ;
- Le développement d'un système de constitution d'index par T. Aït El Mekki et A. Nazarenko, le système IndDoc⁹. Tirant parti des enseignements du travail précédent, l'équipe du LIPN (laboratoire d'informatique de Paris-Nord) a proposé une réflexion et une nouvelle architecture de constitution d'index. Cette architecture considère un index comme une ressource, constituée à partir d'un corpus que des outils permettent d'ébaucher (index ébauche), que l'utilisateur complète (index source) et qui peut ensuite être visualisée.

Ainsi, que ce soit la première expérience, vis-à-vis de la complexité de l'index construit, ou la seconde, vis-à-vis de la complexité des fonctions attendues, tout concourait au développement d'un index numérique permettant de naviguer dans une collection d'articles numériques. Nous avons donc décidé de monter un projet – le projet CEDERILIC¹⁰ – qui vise à associer à un livre « papier » un cédérom proposant une indexation des articles et un accès direct à ceux-ci. Le projet comporte cinq étapes principales (voir figure 1) :

- La transformation des articles du format d'origine (RTF) vers le format XML ;
- L'enrichissement de ce format selon plusieurs contraintes (visualisation, indexation) ;
- Le traitement du corpus ainsi constitué par SYNTAXE afin d'obtenir une liste de candidats termes ;

Mise en forme : Puces et numéros

Mise en forme : Puces et numéros

⁷ BOURIGAULT D. & CHARLET J., « Construction d'un index thématique de l'ingénierie des connaissances », dans [cet ouvrage](#).

⁸ BOURIGAULT D. & FABRE C., « Approche linguistique pour l'analyse syntaxique de corpus », in *Cahiers de grammaire*, n° 25, université Toulouse-Le Mirail, 2000, p. 131-151.

⁹ AÏT EL MEKKI T. & NAZARENKO A., « Comment aider un auteur à construire l'index d'un ouvrage ? », in Colloque international sur la fouille de texte, *Tunis, 2002*, p. 141-157.

¹⁰ CEDERILIC, pour « Cédérom pour indexer le livre IC », est un projet soutenu par France Télécom. En dehors de la forte activité de recherche suscitée par le projet, le soutien a principalement consisté dans le support aux réunions de travail et le financement du stage de DESS d'intelligence artificielle de Baruk Toledano, qui a réalisé les programmes de transformation et d'enrichissement des fichiers.

- La constitution de l'index à partir de ces termes, grâce au logiciel INDDOC ;
- La réalisation finale des fichiers à visualiser et de l'interface de navigation.

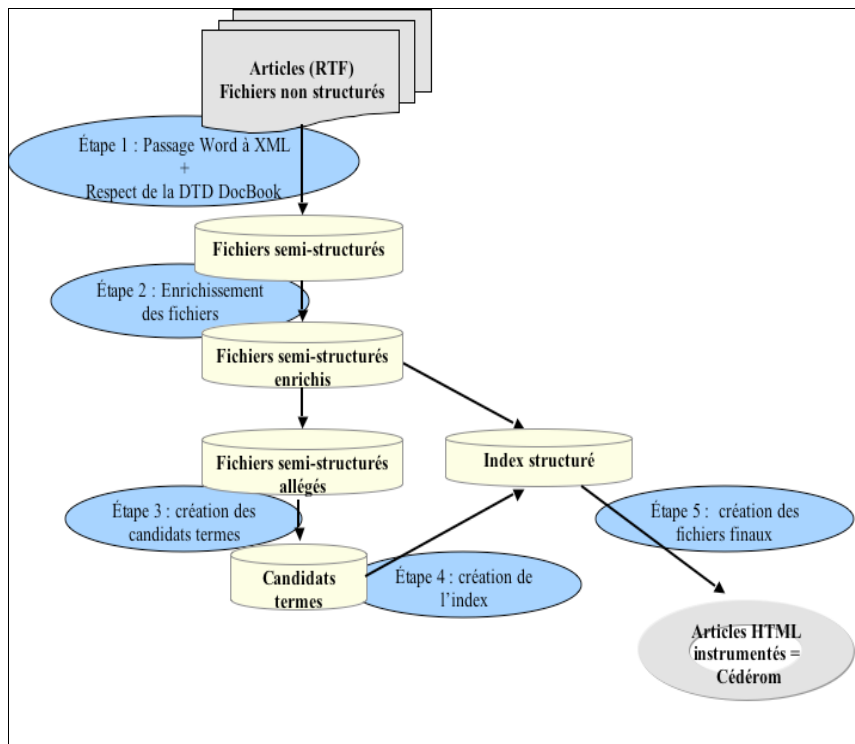


Figure 1 – Étapes de la constitution du cédérom.

Le cédérom ainsi construit est inséré dans ce volume et permet de naviguer dans les articles via un index extrêmement riche (plus de 1 000 entrées).

Nous sommes conscients du caractère expérimental du travail réalisé, mais espérons qu'il satisfera les lecteurs. Un premier article relatant les tenants et les aboutissants de cette expérience a été publié à la Conférence internationale sur le document électronique 2004 (CIDE .7)¹¹.

¹¹ CHARLET J., AÏT EL MEKKI T., BOURIGAULT D., NAZARENKO A., TEULIER R. & TOLEDANO B., *CEDERILIC : constitution d'un livre et d'un index numériques. Actes de la conférence internationale sur le document numérique (CIDE .7), La Rochelle, juin 2004.*

7 CONCLUSION

Si nous avons consacré du temps à la mise en valeur de ces travaux, c'est qu'ils nous semblent constituer un apport notable et spécifique au niveau international, même si l'on peut regretter que la communauté française ne sache pas suffisamment les rendre visibles dans les publications anglophones. C'est aussi que le lien avec les ingénieurs utilisant ce type de concept, de méthode et d'outil, et le lien avec les chercheurs de disciplines connexes, doivent rester vivants et s'actualiser en permanence. Autant de raisons, donc, pour avoir fait l'effort de donner une certaine visibilité à ces travaux.

Nous espérons que les lecteurs qui n'ont pas l'occasion de participer aux rencontres de la communauté « ingénierie des connaissances » pourront ainsi mesurer les avancées de la communauté et participer au débat scientifique qui nous passionne.

7.1.1 Remerciements

Nous tenons à remercier l'ensemble des membres du comité de programme pour leurs relectures de la première sélection des articles. Nous remercions également le bureau du GRACQ et, plus particulièrement, Nathalie Aussenac-Gilles, et tous ceux qui ont animé la communauté pendant ces trois années. Ils ont su susciter, solliciter, avec d'autres, d'intéressantes recherches et communications. Nous remercions également Catherine Garbay pour la préface qu'elle a accepté de rédiger pour cet ouvrage et, plus globalement, pour l'attention fidèle et exigeante qu'elle porte à nos travaux depuis plusieurs années. Merci enfin à Catherine Maksud, qui a assumé un travail long et fastidieux de réalisation des figures, ainsi qu'à toute l'équipe qui a travaillé avec nous à la constitution de l'index et du cédérom (Touria Aït El Mekki, Adeline Nazarenko, Didier Bourigault et Baruk Toledano).