

L'intelligence artificielle : applications documentaires et bibliothéconomiques

Autor(en): **Jacquesson, Alain**

Objektyp: **Article**

Zeitschrift: **Arbido-R : Revue**

Band (Jahr): **5 (1990)**

Heft 1

PDF erstellt am: **22.09.2024**

Persistenter Link: <https://doi.org/10.5169/seals-771743>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern. Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

informazioni del settore della comunicazione scientifica. Potranno inoltre fungere da istanza di traduzione per richiedenti di informazioni nelle categorie di utenza specifiche e, soprattutto, non specifiche, e in questa veste appoggiare i fruitori nella formulazione e nella risoluzione di problemi d'informazione. L'autore esamina, a questo proposito, i presupposti a livello di tecnica procedurale necessari per un servizio bibliotecario di informazioni efficiente, prospetta gli sviluppi possibili nel campo della tecnica e della metodologia dell'informazione negli anni Novanta, e deriva da queste tendenze una serie di criteri per la formazione degli specialisti della conoscenza bibliotecaria. Una nuova prospettiva funzionale nell'ambito del settore bibliotecario europeo richiede pertanto un'estensione del curriculum formativo per comprendere nuove nozioni di tecnica, organizzazione, metodo e psicologia dell'informazione.

Anschrift des Autors:

Ralph Schmidt
Informationswissenschaftler
Viktoriastrasse 3-5
D - 7500 Karlsruhe 1

L'intelligence artificielle

Applications documentaires et bibliothéconomiques

Alain Jacquesson

L'auteur introduit la notion d'intelligence artificielle dans le contexte informatique. Son article présente certains domaines de l'intelligence artificielle pouvant intéresser les sciences de l'information, notamment en ce qui concerne la représentation des connaissances et les systèmes experts.

Trois applications (catalogage, documentation médicale, documentation juridique) sont ensuite présentées en insistant sur les représentations des connaissances utilisées (respectivement les règles de production, les «frames» et Planner).

Der Autor gibt eine Einführung in die Grundbegriffe der künstlichen Intelligenz im Bereich der Informatik. Der Artikel stellt gewisse Domänen der künstlichen Intelligenz vor, die in erster Linie für die Informatikwissenschaften von Interesse sind, namentlich was die Wissensrepräsentation und die Expertensysteme angeht. Drei Anwendungen (Katalogisierung, medizinische und juristische Dokumentation) werden vorgestellt, indem sie sich auf die Wissensrepräsentation beziehen, wofür die Produktionsregeln der Rahmen und die Programmiersprache Planner verwendet werden.

L'autore introduce la nozione di intelligenza artificiale nel contesto dell'informatica. L'articolo prosegue con la presentazione di alcuni campi dell'intelligenza artificiale suscettibili di interessare anche le scienze dell'informazione, specie per quanto concerne la rappresentazione delle conoscenze e i sistemi esperti (Expert Systems). Successivamente, vengono presentate tre applicazioni (catalogazione, documentazione medica, documentazione giuridica) con costante riferimento alla rappresentazione delle conoscenze utilizzate (le regole di produzione, i «frame» e il linguaggio Planner).

1. Introduction

L'intelligence artificielle est une discipline de l'informatique dont l'intitulé même fut à l'origine de nombreuses controverses. Si le terme «artificiel» est généralement compris dans un même sens par chacun, il existe autour du concept d'«intelligence» un flou sémantique qui n'a pas aidé à la compréhension et à l'acceptation généralisées de cette discipline. A cause de cette ambiguïté naquirent des accusations de mécanismes, voire d'anthropomorphismes. Il est aujourd'hui clair que les réalisations actuelles de l'intelligence artificielle n'ont, en aucun cas, un aspect «humain»,

mais qu'au contraire, elles s'expriment dans le cadre de compétences extrêmement spécialisées. Quelles que soient leurs très grandes limites, il est également acquis qu'elles sont utiles à l'homme. De nos jours, l'intelligence artificielle est devenue une discipline adulte, régulièrement enseignée dans les départements d'informatique.

On admet que cette branche de l'informatique est née en août 1956, date à laquelle se sont réunis à Dartmouth, Etats-Unis, un certain nombre de chercheurs de grande renommée (Minsky, Shannon, Newell, Simon notamment) qui se proposaient de réaliser des programmes informatiques doués d'intelligence. Le terme «artificial intelligence» est né à l'occasion de cette rencontre et, en dépit de ses défauts, a été adopté peu à peu par le monde scientifique.

En ce qui concerne plus particulièrement les bibliothèques et la documentation, Linda C. Smith publiait, il y a maintenant dix ans, le premier article de synthèse sur l'utilisation de l'intelligence artificielle dans les sciences de l'information. Ce texte laissait déjà entrevoir de nombreuses possibilités d'application de cette discipline aux procédures utilisées dans les bibliothèques et les systèmes documentaires. L'un des points importants de cette synthèse mettait en évidence, pour nos professions, l'intérêt des travaux sur la représentation des connaissances.

Aujourd'hui, il existe dans le monde plusieurs systèmes experts, les réalisations actuelles les plus performantes de l'intelligence artificielle, spécialement destinés aux sciences de l'information.

2. L'intelligence artificielle

Dans notre article, la présentation de l'intelligence artificielle est volontairement limitée aux aspects relatifs à notre domaine de travail.

Les objectifs de l'intelligence artificielle sont extrêmement ambitieux, puisqu'ils visent «à découvrir comment des programmes d'ordinateurs peuvent s'acquitter de fonctions normalement effectuées par l'intelligence humaine». On distingue deux niveaux de traitement: le premier fonctionne par *augmentation*; dans ce cas, l'ordinateur assiste l'utilisateur dans sa démarche intellectuelle. Le second niveau fonctionne par *délégation*; dans ce cas, le système prend lui-même les décisions en utilisant des critères programmés. A partir de ces premières définitions, on comprend facilement que la profondeur d'intelligence appliquée va varier selon les environnements pratiques: des raisonnements les plus simples à d'autres qui demandent de grandes capacités de connaissance (knowledge) et de puissantes possibilités d'inférences.

Que recouvre le terme «artificiel» dans ce contexte? Nous dirons tout d'abord qu'il s'agit de quelque chose de *synthétisé* par l'homme qui ressemble à la réalité par

un ou plusieurs côtés, mais ne colle pas parfaitement à cette réalité. Nous dirons enfin qu'un objet artificiel peut être défini en termes de *fonction*, de *but* ou d'*adaptation*.

2.1 La représentation des connaissances

Nous limiterons, dans le cadre de la documentation et de la bibliothéconomie, l'approche de l'intelligence artificielle à son aspect cognitif, ce qui pose immédiatement le problème de la représentation des connaissances. Dans un environnement documentaire traditionnel, ce concept recouvre les produits de l'indexation, de l'analyse documentaire, de la classification, etc., en fait de l'analyse de contenu dans son sens le plus large. Dans le strict domaine de la bibliothéconomie, il s'agira, par exemple, de la représentation procédurale d'un code de catalogage.

On caractérise généralement la représentation des connaissances en trois degrés de complexité croissante:

A) Les systèmes monadiques

Il s'agit d'unité de connaissance simple et indivisible. Dans un cadre documentaire, il s'agit typiquement des systèmes d'indexation et de classification traditionnelles: mots-clés, vedettes matières, indices CDU, etc.

B) Les systèmes structurels

Les systèmes structurels exploitent également les relations qui existent entre les éléments, les unités de connaissance individuelles. On les représente généralement sous forme de graphes dans lesquels les nœuds constituent les éléments simples et les arcs les relations. La notion de hiérarchie peut y être introduite au moyen de la relation d'inclusion. Dans le domaine documentaire traditionnel, des langages tels Syntol ou Precis illustrent cette direction. Dans un cadre linguistique, on citera la «grammaire des cas» de Fillmore, que plusieurs auteurs (notamment Courrier, 1977) examinèrent dans un cadre documentaire.

C) Les systèmes épistémiques

Il s'agit naturellement des systèmes les plus élaborés. On donne une description procédurale, un modèle général valable pour un ensemble de situations. Ces systèmes permettent d'identifier de nouveaux objets qui, au contraire d'une liste fermée de mots-clés par exemple, n'étaient pas connus a priori du système.

Les systèmes épistémiques permettent non seulement d'identifier de nouveaux objets, mais également de générer des objets nouveaux. Il s'agit là de l'une des «compétences» de l'intelligence.

De tels systèmes n'ont pas encore été utilisés dans le domaine documentaire. A notre connaissance, les seuls résultats probants se situent dans la démonstration des nouveaux théorèmes.

2.1.1 Les réseaux sémantiques

Dès les années cinquante, on envisagea d'utiliser l'ordinateur pour effectuer des travaux de type linguistique: systèmes de question/réponse en langage naturel, traduction automatique, etc., et naturellement systèmes de recherche documentaire utilisant le langage naturel. Cette époque est considérée comme l'âge d'or de la syntaxe avec les travaux de Y. Bar Hillel, Z. Harris et surtout N. Chomsky. Mais rapidement, dans un diagnostic sévère, Bar Hillel démontre que la linguistique computationnelle ne peut pas être réduite à un dictionnaire et à un jeu de règles de grammaire, si complets soient-ils; le rapport Bar Hillel eut pour conséquence de

faire couper, aux Etats-Unis, presque tous les crédits alloués à la recherche en linguistique computationnelle. De fait, il montrait déjà que les difficultés rencontrées n'allaient pouvoir être résolues qu'au moyen d'outils d'ordre extra-linguistique et que les solutions à envisager nécessitaient des bases de connaissances notamment sémantiques.

L'arrivée dans le domaine documentaire des unitermes, de l'indexation post-coordonnée et des thésaurus avait permis d'utiliser dans un premier temps la puissance brute des ordinateurs. Certains groupes de recherche en sciences de l'information ont ensuite tenté de formuler de nouvelles structures de représentation des connaissances, notamment en cherchant de l'aide du côté de la linguistique; comme nous l'avons dit plus haut, cette tendance fut illustrée par des réalisations telles que les langages documentaires du type Syntol (Gardin, 1962) ou Precis (Austin, 1974). C'est en utilisant d'une part les procédures développées par la linguistique computationnelle, et d'autre part les techniques avancées de représentation des connaissances développées en intelligence artificielle, que les résultats les plus encourageants ont dès lors été enregistrés.

Les *réseaux sémantiques* sont une forme de représentation destinée à encoder une information uniquement descriptive, en omettant toute référence à un plan d'action précis. La «mémoire» est un ensemble de nœuds interconnectés, chacun représentant un élément conceptuel. Cette mémoire prend donc la forme générale d'un graphe orienté. Ce type d'organisation inclut la gestion des redondances, ce qui permet à un concept d'être défini avec les termes d'un ou plusieurs autres. Un concept peut être représenté par un mot, un mot composé, une phrase, un paragraphe ou même une forme de texte plus importante (par exemple, un article de loi, une définition formelle, un brevet, etc.) définissant un concept plus ou moins général. Des formes élaborées de réseaux sémantiques permettent de traiter des propriétés d'attributs-valeurs (modificateurs du type «adjectifs» ou «adverbes»), ainsi que la notion hiérarchique exprimée par la relation «Elément de». Les réseaux sémantiques permettent donc d'intégrer le mécanisme d'*héritage des propriétés* qui est une forme fruste de raisonnement: on peut faire hériter certains nœuds des propriétés des nœuds auxquels ils sont reliés comme pour tout passage du général au particulier.

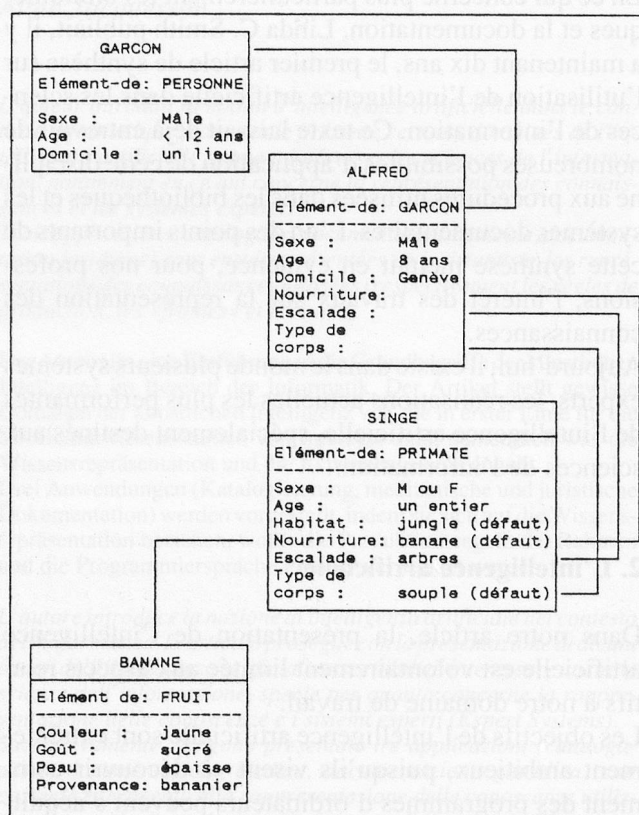
C'est à la fin des années soixante qu'un chercheur de l'Université Carnegie-Mellon, M. Ross Quillian, mit au point le modèle des réseaux sémantiques; mais, limité par les capacités de stockage des ordinateurs de l'époque, il abandonna l'intelligence artificielle pour se tourner vers la sociologie. Cependant, le modèle reste particulièrement intéressant et a été souvent repris.

On aura naturellement noté que le concept des réseaux sémantiques est extrêmement proche de celui des *thésaurus* et plus particulièrement de ceux qui sont représentés sous la forme de *schémas fléchés*.

2.1.2 Les «frames»

En 1974, Marvin Minsky du MIT, à la suite de différents travaux portant notamment sur la psychologie de la mémoire, émit l'hypothèse que l'esprit humain interprétait toute situation sous forme de structures de données qu'il appela «frames», terme dont il n'existe pas de traduction française satisfaisante (treillis, construction ou même objet structuré).

Chaque objet ou concept est un ensemble structuré de connaissances décrit sous forme de cases énumérant les objets et leurs attributs. Dans l'exemple qui suit (Ill. 1), on observera tout d'abord les *relations hiérarchiques* «élément de». Par exemple, ALFRED est «élément de» GARÇON, qui lui-même est «élément de» PERSONNE.



Ill. 1

Ensuite, on notera les *relations «objet-attribut»*: un garçon est de sexe «mâle», il a «moins de 12 ans». Ces informations recouvrent ce que Jean-Claude Gardin appelait les relations paradigmiques: elles sont admises par tous. Chaque attribut peut être décrit selon différentes facettes: valeurs possibles, intervalles, valeurs par défaut ou même procédures.

Certaines cases sont vides au moment de l'ouverture du «frame»; leur contenu dépend du contexte dans lequel le «frame» est utilisé; cela ressemble à ce que Gardin appelait les relations syntagmatiques. Un garçon possède un domicile que l'on ne connaît pas a priori.

Enfin, il existe des valeurs par défaut qui sont attribuées à certaines cases: on attribue une valeur par défaut lorsque tout laisse penser (Common Sense Knowledge) qu'il s'agit de circonstances courantes et que l'on ne possède pas d'informations contradictoires. Ainsi, un singe vit dans la jungle, à moins que l'on ne dise explicitement qu'il est élevé dans un zoo; cela signifie que les valeurs par défaut peuvent en tout temps être changées.

Les «scripts» utilisent le même principe que les «frames»; ils ne décrivent pas des objets, mais des actions simples. Par exemple, l'action de «manger» nécessite un certain nombre de pré-requis (assiette, services, nourriture, etc.) avant de pouvoir être déclenchée.

Ce type de représentation des connaissances s'est révélé assez performant lorsqu'il était appliqué à des systèmes de question-réponse dans des micro-domaines très typés. Il est illustré par les travaux de Wilks, Winograd ou Charniak. L'aspect original des recherches de ce dernier est constitué par l'inclusion de conditions logiques (Si..., alors...) dans les «frames». Quant à Schank, il propose donc dès 1977 des «scripts», une variante qui autorise la manipulation d'objets conceptuels ou d'actions dans un type d'activité donné.

2.1.3 Les règles de production

Les règles de production sont une autre forme de représentation des connaissances de nature déclarative. Chaque règle est une parcelle de connaissance de la forme:

Si conditions alors conclusions (coefficient)

qui signifie que si la ou les conditions sont vérifiées, alors on peut en tirer une conclusion, ou appliquer une nouvelle règle de production. Dans certains cas, la règle est accompagnée d'un coefficient de vraisemblance, ce qui permet de tester certaines hypothèses sans s'enfermer dans une certitude absolue. Dans les systèmes les plus simples, les règles sont exprimées à l'aide de la logique des propositions. Dans d'autres, c'est la logique des prédicats du premier ordre qui est employée, ce qui permet l'usage de variables et de quantificateurs.

Un des développements les plus récents dans le domaine de la représentation des connaissances est constitué par les objets structurés (tels qu'ils sont employés dans les langages orientés objets – LOO – du type Smalltalk). L'univers de travail (représentation des connaissances et manipulation) est constitué d'un ensemble hiérarchique de classes génériques et d'instances d'objets, munis des mécanismes d'héritage de propriétés. Chaque objet est décrit, comme un frame, par des attributs, mais il comprend aussi un ensemble de procédures assurant la manipulation des informations encapsulées dans l'objet. On peut attacher à un objet un ensemble de règles de production chargées de résoudre un sous-problème.

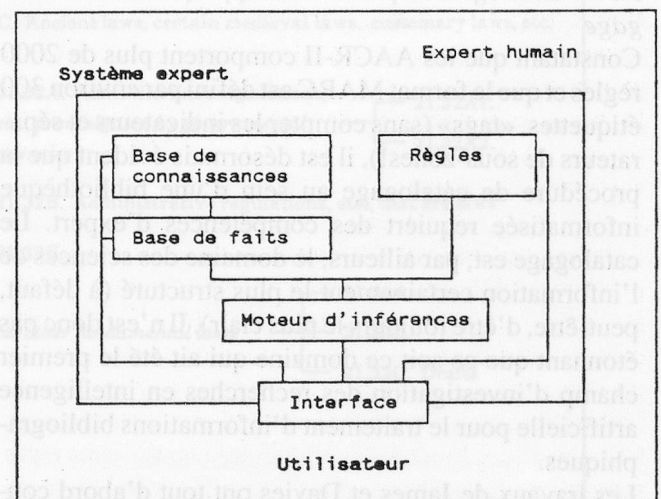
En examinant les structures de données que sont les réseaux sémantiques et, plus particulièrement, les frames, on est frappé par l'analogie certaine qui existe entre les représentations des connaissances retenues par la recherche en intelligence artificielle d'une part, et les outils des systèmes traditionnels d'analyse documentaire que sont les thésaurus ou les schémas fléchés d'autre part. Comme le prévoit Smith, c'est certainement un des points de convergence les plus prometteurs entre l'intelligence artificielle et les sciences de l'information. Nous reviendrons plus loin sur ce point.

2.2 Les systèmes experts en sciences de l'information

Les systèmes experts sont nés au milieu des années soixante, au moment où l'on prenait conscience que les techniques de raisonnement universel n'avaient débouché sur rien de vraiment concret. Les recherches se sont donc tournées vers des programmes qui allaient posséder des informations détaillées permettant de répondre à des questions spécifiques dans des domaines précis. Edward Feigenbaum, professeur à Stanford, est un des pères des systèmes experts; il résumait ainsi la situation: «Nous nous sommes aperçu qu'il valait mieux être bien informé qu'astucieux».

Diplômé en psychologie cognitive, Feigenbaum s'est penché sur la manière dont les chimistes analysent les composés organiques. Il mit au point le système DENDRAL qui intégrait, outre des listes de molécules, des règles du type «si..., alors...». Mais on s'accorde à reconnaître comme premier système expert MYCIN, destiné à diagnostiquer certaines maladies infectieuses du sang. Composé de 500 règles «Si..., alors...», il faisait, dès 1970, aussi bien qu'un vrai médecin.

Les systèmes experts (Ill. 2) sont donc des programmes conçus pour fournir à la fois les connaissances et le raisonnement d'un expert humain dans un domaine bien délimité. Ils se composent de deux parties: d'une part une *base de connaissances*, et d'autre part un



Ill. 2

moteur d'inférences. La base de connaissances est constituée de faits et de règles formelles ou empiriques. Le moteur d'inférences manipule l'information de la base de connaissances et intègre les données extérieures propres à la situation que l'on cherche à analyser et qui constituent la base de faits. Le moteur d'inférences détermine dans quel ordre faire les associations et les inférences.

La constitution d'une base de connaissances se fait en collaboration avec un expert humain, afin de traduire sous forme de données manipulables par une machine l'expertise humaine. On introduit également les diverses associations auxquelles se livre l'expert humain lorsqu'il est confronté à une situation donnée. Dans certains cas, les associations peuvent prendre la forme de réseaux sémantiques ou de «frames». Mais le plus souvent, les systèmes experts reposent sur les règles de production: on reformule le savoir humain sous forme de règles du type «Si..., alors...».

3. Quelques applications documentaires et bibliothéconomiques

L'intelligence artificielle ne pouvait manquer d'intéresser toute personne préoccupée par les développements informatiques des systèmes documentaires ou de bibliothèques. Aujourd'hui, plusieurs recherches visant à créer des systèmes experts pour nos domaines ont vu le jour tant aux Etats-Unis qu'en Europe, que ce soit dans les aspects les plus traditionnels de la bibliothéconomie ou les plus avancés en analyse de contenu. Par exemple, des résultats encourageants ont été enregistrés dans le domaine du catalogage. En Angleterre et en Suède, deux groupes ont tenté de formaliser les règles de catalogage anglo-américaines (AACR-II) sous forme de systèmes experts.

3.1 Les règles de production appliquées au catalogage

Constatant que les AACR-II comportent plus de 2000 règles et que le format MARC est défini par environ 300 étiquettes, «tags» (sans compter les indicateurs et séparateurs de sous-zones!), il est désormais évident que la procédure de catalogage au sein d'une bibliothèque informatisée requiert des compétences d'expert. Le catalogage est, par ailleurs, le domaine des sciences de l'information certainement le plus structuré (à défaut, peut-être, d'être toujours le plus clair). Il n'est donc pas étonnant que ce soit ce domaine qui ait été le premier champ d'investigation des recherches en intelligence artificielle pour le traitement d'informations bibliographiques.

Les travaux de James et Davies ont tout d'abord consisté à réécrire des règles de catalogage AACR-II sous forme de système de production:

SI l'œuvre est une monographie **ET SI** l'œuvre ne possède qu'un seul auteur personnel
ALORS la vedette est constituée par l'auteur personnel (Règle AACR 21.4A)

La même méthode est appliquée pour l'assignation des identificateurs MARC:

SI l'entrée principale est un auteur personnel **ET SI** son nom consiste en un nom de famille et en un prénom ou de ses initiales
ALORS l'identificateur MARC «100.20» doit être assigné

Comme le font remarquer les auteurs, il n'est pas facile d'exprimer toutes les règles AACR-II sous forme de règles de production. Par exemple, la règle AACR-II 22.2C2 demande au catalogueur de «préférer le nom prédominant si un auteur utilise à la fois son nom réel et des pseudonymes». Pour résoudre de tels cas, le système devrait connaître la liste des publications d'un auteur donné ainsi que ses noms de plume. Dans de tels cas, la décision finale est naturellement rendue à l'utilisateur.

D'un autre côté, le système est à même d'examiner, à chaque stade de la saisie, les règles qui doivent être prises en considération. Pour cela, il utilise le graphe des relations entre règles. Il évite ainsi l'écueil de la présentation linéaire d'un code de catalogage et utilise avantageusement le cheminement logique du catalogueur.

Un contrôle de cohérence, tant pour les choix effectués par le système que pour ceux opérés par l'utilisateur, est fait à différents stades de la saisie. De façon très classique, ces contrôles sont effectués au moyen de *tables de décision*.

Les auteurs de ce système expert, réalisés à l'Université d'Exeter en Angleterre, ont naturellement constaté que les règles AACR-II ne couvrent pas tous les cas rencontrés dans leur bibliothèque. Ils ont, pour résoudre de tels cas, utilisé des exemples:

SI une publication a un titre (comportant telle caractéristique...),
ALORS suivre l'exemple 6.

Une telle méthode pédagogique, même si elle peut paraître triviale, est connue sous le nom d'apprentissage par l'exemple (learning by example). Le niveau d'expertise du système peut tenir compte de particularités bien spécifiques: par exemple, en reconnaissant le préfixe d'un ISBN, le système en déduit qu'il s'agit selon toute vraisemblance d'un photocopie d'un département bien précis et propose une grille de saisie appropriée et les règles qui lui sont associées.

L'ensemble du dialogue avec l'utilisateur est effectué au moyen de menus. En tout temps, une «fenêtre» (au sens informatique) peut s'ouvrir sur l'écran afin de consulter les règles AACR-II qui s'appliquent à la situation en cours.

Ce système expert de catalogage a été tout d'abord écrit en Prolog, un langage de programmation bien adapté à l'intelligence artificielle, mais son utilisation s'est avérée trop lente et le système a été réécrit en Pascal.

C'est au Département d'informatique de l'Université de Linköping, en Suède, que Hjerpe et Olander ont, dès 1984, développé un projet appelé HYPERCATalog, et plus particulièrement le sous-système de catalogage ESSCAPE (Expert System for Simple Choice of Access Points for Entries). Ce système est basé sur des arbres de décisions portant sur les vedettes – les «points d'accès» aux catalogues – à adopter. Leur première approche fut de structurer sous forme d'arbres les règles AACR-II; des tentatives de formalisation des règles de catalogage avaient déjà eu lieu, notamment les travaux de Shaw (Using AACR2: a diagrammatic approach). Deux versions d'ESSCAPE ont été réalisées: la première utilise le système expert MYCIN que nous avons vu plus haut; l'ensemble des règles de catalogage est exprimé sous la forme de règles de production (Si..., alors,...). La seconde est basée sur le système EXPERT-TREE, un produit commercial réalisé à partir de EXPERT EASE; dans ce cas, le code de catalogage doit

être représenté, comme on l'a vu, sous la forme de structure d'arbre.

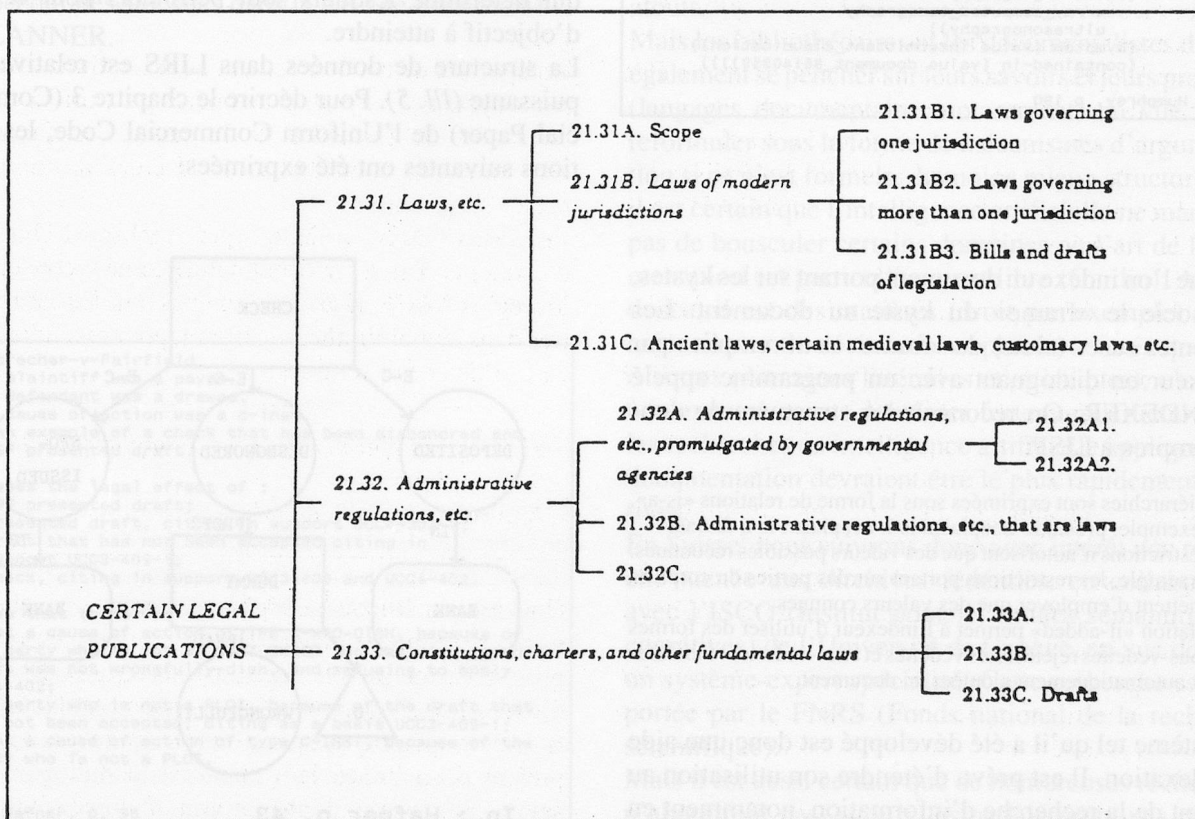
L'exemple ci-après (Ill. 3) présente un fragment de cet arbre de décision.

Les auteurs de ce système expert, destiné à des catalogueurs, ont noté les difficultés suivantes:

1. il est fort complexe de formaliser un code de catalogage comportant parfois des contradictions et des manques de clarté;
2. le rôle des exemples et des choix déjà faits est extrêmement important, en conséquence un système expert de catalogage de «deuxième génération» devrait avoir accès à un fichier d'autorité, comme dans les systèmes de catalogage informatisé les plus performants actuellement;
3. dans ses deux premières versions, ESSCAPE n'inclut que très peu d'expertise humaine; le pas suivant sera d'inclure au moyen de procédures de «knowledge engineering» le compétence des catalogueurs chevronnés, notamment pour le traitement des cas spéciaux.

3.2 Les «frames» appliqués à la documentation médicale

La National Library of Medicine (NLM) à Bethesda a toujours été à la pointe des nouvelles technologies et méthodes dans le domaine de la documentation médicale: le système documentaire MEDLINE est un des systèmes informatisés traditionnels les plus performants à ce jour. Dans le domaine de l'intelligence artificielle, cette bibliothèque développe un projet d'aide à l'indexation faisant partie du «Programme de classification et de recherche automatisées» (ACRP – Auto-



Ill. 3

mated classification and Retrieval Program). On rappellera que l'analyse de contenu est faite dans MEDLINE au moyen de la liste de vedettes matières MeSH (Medical Subject Headings). Le projet d'aide à l'indexation est basé sur une représentation des connaissances qui utilise la technique des «frames». Une partie de la recherche a donc consisté à réécrire une partie de la liste MeSH sous forme de «frames». Il emploie pour cela un langage expérimental de représentation des connaissances appelé FRAME KIT basé sur LISP et développé à la Carnegie Mellon University. Le système «FRAME KIT» utilise le langage LISP. Il fonctionne à la NLM sur un miniordinateur VAX 11/780.

A la différence de MeSH, les «frames» de la base de connaissance contiennent non seulement les entités sémantiques, mais également des données factuelles. De plus, le système permet d'encoder certaines connaissances procédurales qui sont utilisées comme aides à l'indexation au moment de la saisie. Par ailleurs, toute la structure traditionnelle (voir, voir aussi, etc.) du thésaurus MeSH est également prise en compte. L'exemple ci-après (Ill. 4) présente un «frame» concernant l'analyse de kyste.

```

-> (ppf cys_85140959)
  (setq frame :
    (cyst-85140959 (is-a (value cyst))
      (body-part (value popliteal_artery))
      (procedure
        (value angiography
          x-ray_computed_tomography
          ultrasonography))
      (symptom (value intermittent_claudication))
      (contained-in (value document_85140959))))
  In : Humphrey, p 190

```

Ill. 4

Lorsque l'on indexe un document portant sur les kystes, on associe le «frame» du kyste au document. Les différentes cases (slots) du «frame» sont remplies par l'indexeur en dialoguant avec un programme appelé ASKINDEXER. On reconnaît les structures de données propres à LISP:

- les hiérarchies sont exprimées sous la forme de relations «is-a». Par exemple: prostatic-neoplasm is-a neoplasm is-a disease
- les restrictions n'autorisent que des valeurs possibles reconnues. Par exemple, les restrictions portant sur des parties du corps ne permettent d'employer que des valeurs connues
- la relation «if-added» permet à l'indexeur d'utiliser des formes de sous-vedettes rejetées, les vedettes et sous-vedettes reconnues étant automatiquement ajoutées au document.

Le système tel qu'il a été développé est donc une aide à l'indexation. Il est prévu d'étendre son utilisation au moment de la recherche d'information, notamment en présentant les «frames» à l'utilisateur final. Cette dernière opération devrait être faite en langage naturel et il

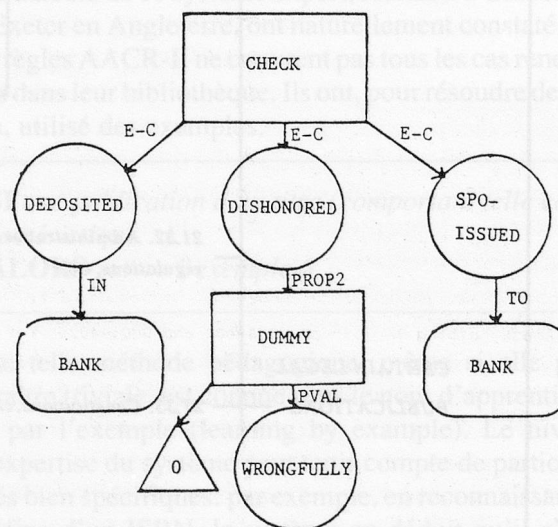
est prévu de construire le système d'analyse autour de la «grammaire des cas» de Fillmore.

3.3 Une représentation procédurale de documentation juridique

La thèse de doctorat de Carole D. Hafner a été présentée à l'Université de Michigan en 1978, puis reprise en vue de sa publication en 1981. Elle porte sur la création d'un système expérimental de recherche d'information juridique: LIRS (Legal Information Retrieval System). Le domaine juridique choisi par Hafner est la recherche de cas de jurisprudence en droit commercial, plus particulièrement celui qui a pour base l'«Uniform Commercial Code» aux Etats-Unis.

Le système LIRS est réalisée sous forme de base de données Micro-PLANNER. L'utilisation d'un langage de programmation tel que PLANNER a pour objectif de donner une plus large souplesse quant à la représentation des connaissances. On distinguera, sous PLANNER, trois niveaux de représentation des connaissances: le premier est le niveau sémantique traditionnel; le second concerne le traitement formel procédural; quant au troisième niveau, il permet de représenter le problème sous forme d'objectif à atteindre, sans pour autant préciser la manière d'atteindre cet objectif. PLANNER permet de manipuler ces trois niveaux sans rupture entre chacun d'entre eux. Plutôt que d'exprimer un fait, un «théorème», PLANNER exprime une séquence d'actions à entreprendre en vue d'atteindre l'objectif fixé. Les théorèmes sont stockés dans un index spécial qui détermine lesquels sont pertinents pour le type d'objectif à atteindre.

La structure de données dans LIRS est relativement puissante (Ill. 5). Pour décrire le chapitre 3 (Commercial Paper) de l'Uniform Commercial Code, les relations suivantes ont été exprimées:



In : Hafner p. 43

Ill. 5

- Relations simples (CVAL)
 - relations d'appartenance (set membership)
 - relations de constituants (constituents links)
- Relations semi-complexes (PVAL)
 - sous-classes (classification non-hiérarchique: un objet peut appartenir à plusieurs classes)
 - propriétés (au sens de l'héritage des propriétés)
 - binaire
- Relations complexes (RVAL)
 - relations de rôle (dans la situation)
 - relations conditionnelles
- Frontières floues (Fuzzy boundaries)
 - Certaines assertions (par exemple: rapide, tendu, etc.) sont difficiles à classer et la notion de «frontière floue» peut leur être attribuée.

Cette représentation des connaissances est manipulée par un langage de description des situations (LDS) qui permettra d'une part de décrire les situations juridiques envisagées, et d'autre part de formuler les recherches que l'on veut effectuer sur la base de données Micro-PLANNER. Le LDS fonctionne sur le principe de la «grammaire des cas» de Fillmore, où chaque type de concept génère les «cas» possibles.

Les résultats, pour la jurisprudence retrouvée, sont présentés sous la forme suivante: les parties, l'objet du litige et les considérants, le tout sous la grammaire et la sémantique Micro-PLANNER (III. 6).

Le système LIRS utilise une interface d'interrogation classique, qui permet de sélectionner les cas appropriés de jurisprudence en utilisant les relations sémantiques et procédurales stockées dans la base de données Micro-PLANNER.

```
Steinbrecher-v-Fairfield
The plaintiff was a payee.
The defendant was a drawee.
The cause of action was a C-inst.
is an exemple of a check that has been dishonored and
a NSF presented draft.
```

```
Discusses the legal effect of :
a NSF presented draft;
a presented draft, citing in support UCC4-303-2;
a draft that has not been accepted citing in
support UCC3-409-1;
a check, citing in support UCC3-409 and UCC4-402.
```

```
Decides that there was :
not a cause of action of TPE C-WRO-DISH, because of
the party who is not an ACCT-OWNER and the dishonored
draft was not wrongfully-dish, and refusing to apply
UCC4-402;
a party who is not a PLOI, because of the draft that
had not been accepted; citing as a basis UCC3-409-1;
not a cause of action of type C-INST, because of the
party who is not a PLOI.
```

In : Hafner, p. 96

III. 6

4. Perspectives et développements

L'intelligence artificielle est ainsi devenue une science adulte, et ses applications relevant de la documentation et des bibliothèques commencent à être désormais nombreuses et variées.

Mais les outils actuellement à notre disposition vont, selon les spécialistes du domaine, évoluer au cours des prochaines années. Il est raisonnable de penser que dans le futur nous disposerons des facilités suivantes:

- langages appropriés permettant d'exprimer et de manipuler des formes variées de représentation des connaissances, par exemple au moyen des langages orientés objets;
- interface entre ces langages et les systèmes de gestion de bases de données (SGBD);
- apparition de matériel informatique spécifique à l'intelligence artificielle; c'est déjà le cas avec quelques machines LISP ou PROLOG; les Japonais nous promettent des ordinateurs dits de la «Cinquième génération», directement conçus pour l'intelligence artificielle;
- architecture informatique parallèle ou neuronale.

En ce qui concerne les thèmes avancés, notamment par Jean-Paul Haton, nous avons noté plus particulièrement les aspects suivants:

- extraction automatique de connaissances; cet aspect doit naturellement être particulièrement surveillé dans le domaine de l'analyse automatique de contenu;
- capacité d'apprentissage automatique (learning machines);
- approche connexionniste ou neuro-mimétique.

Dans les domaines voisins de la documentation, il est évident que les progrès, lents mais continus, de la linguistique computationnelle ne peuvent être que des atouts.

Mais les bibliothécaires et les documentalistes devront également se pencher sur leurs savoirs et leurs pratiques (langages documentaires notamment) en vue de les reformuler sous la forme de mécanismes d'argumentation si ce n'est formels, du moins mieux structurés; car il est certain que l'intelligence artificielle ne manquera pas de bousculer certains domaines où l'art de l'argumentation est particulièrement libre (Gardin). A l'évidence, il est des matières (droit par exemple) où les mécanismes de décision sont plus structurés que dans d'autres (domaines littéraires et artistiques); c'est certainement dans ces domaines fortement structurés que les recherches en intelligence artificielle appliquée à la documentation devraient être le plus rapidement positives.

En Suisse, nous suivrons donc avec intérêt une recherche menée aux Archives fédérales, en collaboration avec l'ISCO (Institut pour les études sémantiques et cognitives) de l'Université de Genève, en vue de créer un système-expert spécialisé. Cette recherche est supportée par le FNRS (Fonds national de la recherche scientifique).

Mais il est aussi certain que de nombreuses recherches n'aboutiront pas, et qu'elles ne laisseront pas plus de traces que les analyses automatiques de contenu d'il y a vingt ans.

Références

- SMITH, Linda C. Artificial Intelligence in Information Retrieval Systems. *Information Processing and Management*, vol. 12 (1976), pp. 189–222.
- SMITH, Linda C. Artificial Intelligence Applications in Information Systems. *Annual Review of Information Science and Technology*, vol. 15 (1980), pp. 67–106.
- SMITH, Linda C. Artificial Intelligence and Information Retrieval. *Annual Review of Information Science and Technology*, vol. 22 (1987), pp. 41–77.
- COURRIER, Yves. L'indexation automatique : état de la question et perspectives d'avenir. *Documentation et bibliothèques*, juin 1977, pp. 59–72.
- JÜTTNER, Gerald und GÜNTZNER, Ulrich. *Methoden der künstlichen Intelligenz für Information Retrieval*. München : K.G. Saut, 1988.
- QUILLIAN, M.R. *Semantic memory*. Semantic Information Processing. Ed. by M. Minsky. Cambridge : MIT Press, 1986; pp. 227–270.
- WINOGRAD, Terry. *Frames and the declarative procedural controversy*. Representation and Understanding. New York : Academic Press, 1975.
- MINSKY, Marvin. *A framework for representing knowledge*. The psychology of computer vision. New York : McGraw-Hill, 1975; pp. 211–280.
- HATON, Jean-Paul et HATON, Marie-Christine. *L'intelligence artificielle*. Paris : Presses universitaires de France, 1989.
- SALTON, Gerard. *Automatic text processing : the transformation, analysis and retrieval of information by computer*. Reading, Mass : Addison-Wesley, 1989.
- DAVIES, Roy and JAMES, Brian. Towards an expert system for cataloguing : some experiments based on AACR2. In : *Program*, vol. 8 (1984), no 4, pp. 283–297.
- DAVIES, Roy. *Expert systems and cataloguing : new wine in old bottle?* Expert systems in libraries. Ed. by F. Gibb London : Taylor Graham, 1986.
- BURGER, R.H. Artificial intelligence and authority control. *Library resources and technical services*, octobre 1984, pp. 337–345.
- HJERPPE, R., OLANDER, B., MARKLUND, K. *Projet ESSCAPE – Systèmes experts appliqués au choix des points d'accès pour les entrées de catalogue : application de l'intelligence artificielle*. IFLA, Chicago, 1985. (163-CAT-4-F).
- JAMES, B. *Expert systems for library-cataloging functions*. Univ. Exeter, oct. 1983, 119 p. (PhD thesis).
- BORKO, Harold. Getting started in library expert systems research. *Information processing and management*, vol. 23 (1987), no 2, pp. 81–87.
- DREYFUS, Hubert L. *Intelligence artificielle : mythes et limites*. Trad. de l'éd. américaine de 1979. Paris : Flammarion, 1984.
- GIBB, Forbes (Editor). *Expert systems in libraries. Proc. of a conference of the Library Association Information technology group... Nov. 1985*. London : Taylor Graham, 1986.
- GARDIN, J.C. et al. *Systèmes experts et sciences humaines. Le cas de l'archéologie*. Paris : Eyrolles, 1987.
- HAFNER, Carole D. *An Information Retrieval System based on a computer model of legal knowledge. Revision of thesis (Ph.D.) Uni. of Michigan*. Ann Arbor : UMI Research Press, 1981.
- TODESCHINI, Claudio and FARREL, Michael P. An expert system for quality control in bibliographic databases. *Journal of the American Society for Information Science*, vol. 40 (1989), no 1, pp. 1–11.
- HJERPPE, Roland and OLANDER, Brigitta. Cataloging and expert systems : an AACR II as a knowledge base. *Journal of the American Society for Information Science*, vol. 40 (1989), no 1, pp. 27–44.
- HUMPHREY, Suzanne M. and MILLER, Nancy E. Knowledge-based indexing of the medical literature: the indexing aid project. *Journal of the American Society for Information Science*, vol. 38 (1987), no 3, pp. 184–196.
- BASSANO, C.J. Systèmes experts et stratégies de recherche de l'information. *L'informatique documentaire*, no 32, 1988, pp. 41–68.
- CROFT, W. Bruce. Artificial Intelligence and Information Retrieval. *Information processing and management*, vol. 23, 1987, no 4, pp. 249–382.

(Zusammenfassung auf deutsch)

Die künstliche Intelligenz ist eine Sparte der Informatik, welche besonders für die Informatikwissenschaften bezüglich der Recherchen über die Wissensrepräsentation von Interesse ist. Die semantischen Netze, die Rahmen und die Produktionsregeln sind Modelle der Wissensrepräsentation, welche speziell auf dem Gebiet der Inhaltsanalyse neue Perspektiven eröffnen. Die Expertensysteme sind zugleich die fortschrittlichsten Ausführungen im Bereich der künstlichen Intelligenz. Der Artikel stellt auch einige Anwendungen der künstlichen Intelligenz im Bereich der Informatikwissenschaften vor. Die Produktionssysteme sind gut übertragbar auf eine strukturierte Domäne, wie die der Formalkatalogisierung, was durch die englischen und schwedischen Teams gezeigt wird. Die Rahmen wurden an der National Library of Medicine in Bethesda anhand der Medizin-Dokumentation geprüft. Schliesslich erlaubte die Programmiersprache Planner eine auf drei Ebenen arbeitende Datenbank (semantisch, vorgehensspezifisch und vorgabenspezifisch) im Bereich der juristischen Dokumentation einzurichten. Der Schluss resümiert die gegenwärtig studierten Richtungen im Bereich der künstlichen und «anfälligen» Intelligenz, welche für die Informatikwissenschaften von Interesse sind. Er zeigt auch, dass wir unsere Argumentationen besser strukturieren müssen, besonders in den «weichen» (= nicht Exakte) Wissenschaften.

(Résumé en italien)

L'intelligenza artificiale è una disciplina dell'informatica che interessa più da vicino le scienze dell'informazione inerenti alle ricerche sulla rappresentazione delle conoscenze; le relazioni semantiche, i «frame» e le regole di produzione sono modelli di rappresentazione delle conoscenze che aprono nuove prospettive, in particolare nel campo dell'analisi dei contenuti. I sistemi esperti (Expert Systems) sono attualmente la realizzazione più avanzata nell'ambito dell'intelligenza artificiale. L'articolo prosegue con la presentazione di alcune applicazioni dell'intelligenza artificiale nel campo delle scienze dell'informazione. Ricerche ingle-

si e svedesi hanno dimostrato l' idoneità dei sistemi di produzione al trattamento di un campo strutturato come la catalogazione formale. I «frame» sono stati applicati alla documentazione medica nell' ambito di esperimenti condotti alla National Library of Medicine a Bethesda. Infine, il linguaggio Planner ha permesso di costituire una base di dati a tre livelli (semantico, procedurale e per obiettivi mirati) nel campo della documentazione giuridica.

In conclusione, l' autore riassume gli indirizzi di ricerca più recenti relativi all' intelligenza artificiale, suscettibili di interessare anche le scienze dell' informazione. Inoltre, constata la necessità di una migliore strutturazione delle nostre argomentazioni, specie nel campo delle scienze «non rigorose».

Adresse de l' auteur:

Alain Jacquesson, directeur
Bibliothèques municipales de la Ville de Genève
Place de la Madeleine
1204 Genève



DEWEY, Melvil. – Dewey decimal classification and Relative index. – Ed.20 / ed. by John P. Comaromi ... et al. . – Albany : Forest Press, 1989. – 4 vols. – ISBN 0-910608-37-7 : \$ 200.–

Les révisions de la Classification de Dewey se sont succédées depuis 1876 à des intervalles qui ont varié de 2 à 15 ans, la moyenne se situant autour de 10 ans. La 20^e édition, qui paraît 10 ans après la 19^e, est donc «dans les temps», et les organismes responsables de la mise à jour peuvent à juste titre se flatter de satisfaire au mieux les utilisateurs, dans les limites permises par un système hiérarchique, auquel le qualificatif de rigide est facilement attribué.

Pour chaque nouvelle édition, le «mode d'emploi» est révisé, l' introduction est retravaillée pour davantage de lisibilité et une compréhension plus facile. Dans les nouvelles versions sont également incorporés les progrès faits dans le domaine de la théorie de la classification en général.

Enfin, les modifications de contenu et de signification des indices, leur création ou leur développement, comme leur suppression, résultent d' une concertation entre des experts qualifiés et divers organismes, la Decimal Classification Division, rattachée à la Bibliothèque du Congrès à Washington, les Forest Press et le Decimal Classification Editorial Policy Committee (EPC), rattachés depuis 1988 à OCLC Online Computer Library Center.

Il faut noter aussi que chaque révision est basée sur l' expérience pratique de l' indexation de 500 000 à 800 000 unités bibliographiques, effectuée pour la Bibliothèque du Congrès, donc sur un contexte réel de littérature en plusieurs langues, et de plusieurs pays.

La 20^e édition tient ses promesses. Du point de vue du contenu, les modifications pour les praticiens ne seront pas une surprise, puisque les principales avaient déjà fait l' objet de publications intermédiaires, ou d' informations dans le périodique à parution irrégulière: *Decimal Classification Additions, Notes and Decisions*. Elles touchent principalement les classes

«004–006 data processing and Computer science»

«301–307 sociology»

D' autres modifications importantes touchent les subdivisions géographiques: «Japon, Colombie britannique, Afrique du Sud, Suisse», etc.