

La Connaissance et l'ingénierie des Connaissances

A. Dziri Ghouas Anissa & B. Yalaoui

Laboratoire R&D en IST CERIST

E-mail: adziri@mail.cerist.dz

E-mail: byalaoui@mail.cerist.dz

1- Introduction

La prolifération du texte en ligne et la disponibilité d'importantes bases de données textuelles, ont motivé la plupart des travaux sur l'interprétation des textes. Bien que des volumes massifs d'information soient disponibles à bas prix sous forme de textes libres, les usagers n'ont pas les capacités de tout lire et digérer cette masse d'information rapidement.

Souvent, l'utilisation efficace de l'information à partir du texte, exige que l'information soit mise dans une forme structurée et organisée. Actuellement, rechercher une information à partir de données même indexés et stockées dans des bases volumineuses, est un processus qui devient difficile et complexe aboutissant souvent à des problèmes de bruit et de silence.

Conçue plus pour répondre à des besoins de documentation, que pour aider à la résolution de problèmes, ce type de recherche classique se fait avec des moyens plutôt limités et sans aucune corrélation avec, par exemple, un système à base de connaissance ou d'aide à la synthèse. A ce titre, l'extraction des connaissances à partir des bases de données est une question actuelle qui se pose selon une intention stratégique basée sur des méthodes intelligentes d'accès à l'information.

Nous comptons à travers cette étude de donner un aperçu sur la connaissance, son contexte, ses formes, et l'évolution de la culture humaine vis-à-vis de ce concept. Puis on présentera le domaine de l'ingénierie des connaissances avec ses origines et ses applications, et son rapport avec les sciences de l'information.

2. La Connaissance

2.1 Tentatives de définition

La connaissance a de multiples sens, selon la discipline ou le contexte.

❶ Les philosophes sont les premiers à s'intéresser par l'étude de ce concept. Ces derniers cherchaient à travers ce concept la place réelle de l'homme dans le monde. Les philosophes considèrent que la connaissance est le résultat de l'appréhension de l'univers dans sa réalité première. D'après eux, quand l'homme acquiert une connaissance, il se construit une image très partielle du monde¹

❷ Les psychologues s'efforcent d'expliquer le rôle de la connaissance et le fonctionnement humain en tentant de trouver des outils de modélisation de cette donnée qui permettra d'analyser le comportement de son fonctionnement²

❸ Actuellement, de nombreux chercheurs provenant de disciplines différentes (informatique, logique, linguistique, automatique, etc.) se sont intéressés à la connaissance. Leurs efforts sont investis vers la capitalisation des connaissances en essayant d'automatiser le processus d'accès efficace à l'information soutenue avec un processus de raisonnement et/ou de décision.

On peut dire que la connaissance est un ensemble de concepts et de liens relationnels possibles, acquis à partir des informations disponibles sur divers supports et/ou des expériences vécues.

La connaissance est de plusieurs types et a plusieurs origines, elle peut être extraite sur :

- ↳ des sources d'information variées
- ↳ de l'expertise
- ↳ des explications
- ↳ des stratégies
- ↳ et des façons de faire

¹ J. D. Warnier.- L'homme face à l'intelligence artificielle"; Ed. Organisation, 1984.

² C. Jacquard-choulet.- Proposition d'un modèle cognitif axé sur l'auto organisation de la connaissance pour la résolution de problèmes.- Thèse de l'université de Besançon 1993.

Elle peut être donc donnée de façon :

- Explicite : sur des supports d'information de toute nature
- ou implicite : par les spécialistes et les experts d'un domaine

2.2 Cadre historique

Selon J.P.Micaelli¹, le concept de connaissance est passé par trois périodes distinctes :

1940 – 1950

Durant cette période est née la notion d'information telle qu'elle est connue aujourd'hui. A cette période on assimilait l'information à un flux et le savoir à un stock de données.

- La capitalisation des connaissances consistait à accroître le stock du savoir.

1960 – 1980

A cette période, les efforts étaient orientés sur le rôle du savoir scientifique et technique produit par la recherche et le développement industriel (R&D). La compétitivité se mesurait alors par rapport au volume de connaissance produit par l'activité R&D.

- La capitalisation des connaissances revient donc à affiner celle développée précédemment, en lui confiant en outre un caractère dynamique.

A partir de 1990

L'hypothèse selon laquelle l'accroissement du stock du savoir scientifique et technique suffit pour expliquer la croissance ou la compétitivité d'une nation est remise en cause et d'autres concepts ont été recherchés dont l'un est incontestablement celui de " compétence ".

La capitalisation qui émerge est de type " Actionaliste " c.a.d. que l'action vers la résolution des problèmes est déjà acquise, sans passer par la trajectoire classique de recherche de solution dans diverses sources. Cette nouvelle tendance rejette toute possibilité de mesure universelle et unique de la performance cognitive.

¹ J. P. Micaelli, "Les trois âges de la capitalisation de la connaissance", in J. M. Fouet, connaissance et savoir faire en entreprise, Ed Hermes, 1997.

2.3. Modélisation des connaissances : Entre l'approche des behavioristes et fonctionnalistes

Pour répondre à la question " les machines peuvent-elles penser ? ", Turing¹ a proposé de la remplacer par le jeu de l'imitation. Le principe de ce jeu est que la machine doit pouvoir être capable de convaincre un interlocuteur humain et que elle-même est un humain. Le problème qui se pose : pouvant nous construire des machines capables de réussir ce test ?

Ce test se base sur une conception " béhavioriste " du comportement intelligent, mais le béhavioriste seule ne peut pas rendre compte de l'intelligence d'un système. Autrement dit, si deux systèmes (une machine et un homme) produisent la même sortie, cela ne prouve pas qu'ils adoptent le même comportement intelligent. C'est les processus internes qui ont permis de produire la sortie en question qu'il faut considérer. Pour pallier aux déficiences de cette théorie, les philosophes proposent le fonctionnalisme.

Pour un " fonctionnaliste " se sont les relations causales d'un état qui sont à prendre en considération et pas seulement l'entrée et la sortie comme pour les béhavioristes. Les états mentaux sont identifiés sur la base de leur rôle fonctionnel, et sont considérés capables d'expliquer le comportement dont ils sont la cause².

Les fonctionnalistes distinguent plusieurs niveaux d'abstraction dont les systèmes sont organisés et décrits. Le fait de décomposer en niveaux d'abstraction amène à l'idée de décomposer en niveaux de traitement, aboutissant au niveau le plus inférieur pour une manipulation de symboles formels.

2.4 Principales sources de la connaissance

a) Sources implicites

➤ **Les experts** : Les experts sont incontestablement la source la plus importante des connaissances, ils détiennent une connaissance d'une importance stratégique. C'est une connaissance qui est validée et surtout renouvelée et enrichie à chaque nouvelle expérience et/ou information. Les principaux

¹ <http://www.geocities.com/SouthBeach/Marina/8256/turing.htm>

² Erica Valencia.- Représentation des connaissances : notes bibliographiques .
<http://www.limsi.fr/Individu/erika/representations.html>

outils mis en œuvre pour l'extraction de ce type de connaissances sont les questionnaires et les interviews. Ce type de connaissance, est intéressant dans la mesure où celle-ci est dynamique puisque des informations peuvent être fournies au fur et à mesure que le besoin est exprimé.

b) Sources explicites

➤ **Les documents** : Les documents sont les principaux moyens de sauvegarde et conservation systématique des connaissances. Ils sont susceptibles de contenir des connaissances complémentaires à celle des experts et souvent ce sont les seuls supports disponibles. L'extraction des connaissances à partir des données textuelles est plus complexe que l'extraction des connaissances à partir des experts, car avec les documents on perd toute la dynamique du dialogue avec les experts. Il faut donc extraire cette connaissance en focalisant sur sa pertinence.

➤ **Les bases de données** : On retrouve dans les organisations et les entreprises de très grandes masses de données, accumulées durant des années et couvrant toute l'activité de l'organisation : c'est la mémoire des entreprises. Ces données sont organisées, et très souvent sous forme de bases de données et de fichiers. Et qui représentent des sources formelles de connaissance.

➤ **Les documents vidéo** : Ces documents, qui font appel au sens visuel et auditif de l'être humain, sont de plus en plus utilisés dans les organisations dans divers domaines. Ils servent de support d'enseignement mais aussi, pour présenter des modes opératoires, des suivis d'expériences ou de présentation d'experts en action de diagnostic ...etc.

➤ **Les documents sonores** : Ces documents sont moins difficiles à analyser que les documents vidéo car ils ciblent un seul sens humain qui est l'audition. Ils sont utilisés pour mémoriser toute information sonore tel que l'échantillonnage de voix dans le cas d'expériences scientifiques, bruits caractéristiques d'outils, etc. Les documents sonores sont pris en charges par les techniques de traitement de signal avec de réelles possibilités de mémorisation, de traitement et de restitution fidèle de signaux sonores.

2.5 Les types de connaissances

Un individu utilise différents types de connaissances dans ses activités quotidiennes. Par exemple, il se sert de connaissances déclaratives comme " une bicyclette possède deux roues " pour définir les éléments de son environnement et aussi des connaissances procédurales comme " savoir rouler en bicyclette " pour accomplir certaines actions. Ces connaissances se distinguent par leur action et utilité.

" Les connaissances déclaratives permettent de répondre à la question " quoi ? " alors que les connaissances procédurales répondent à la question " comment ? "

Les connaissances déclaratives peuvent être divisées en deux types secondaires : à savoir les connaissances sémantiques et les connaissances épisodiques. La figure ci-dessous résume ces différents types de connaissances ¹

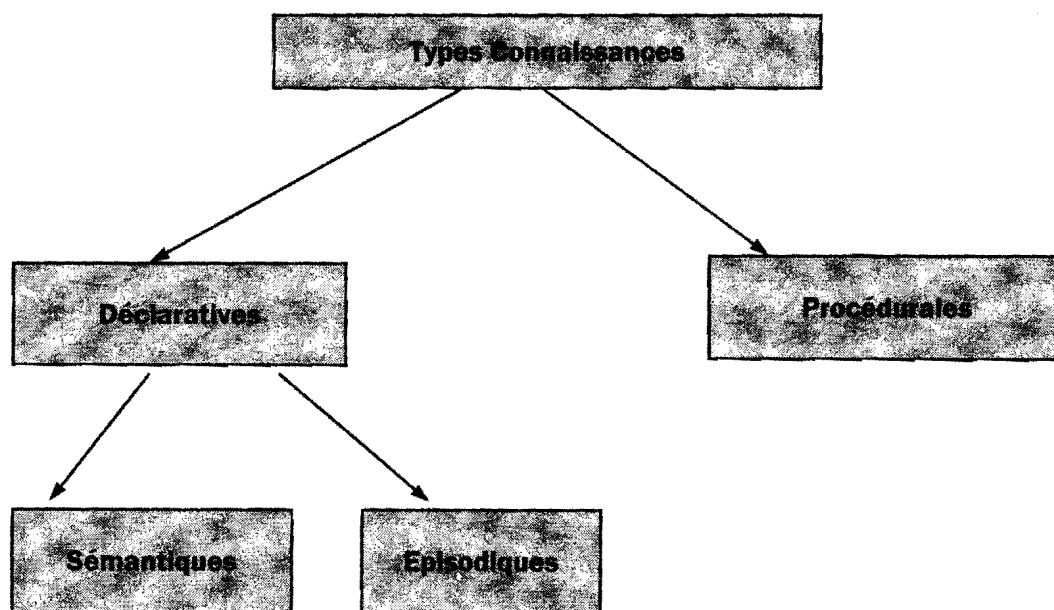


Figure 1 : Hiérarchie des types de connaissances

¹ Gilles Willet .- La Communication modélisée : Une introduction aux concepts, aux méthodes et aux théories. – Ottawa : Ed du renouveau pédagogique, 1992.- p.216

2.5.1/ Les connaissances déclaratives

Les connaissances déclaratives sont les connaissances qui portent sur les faits, les objets et les choses. Ces connaissances, relativement statiques, permettent à l'individu d'identifier ou de définir les éléments de son environnement.

Exemple : - Anne a des cheveux blonds (fait)

- Les oiseaux pondent des œufs (généralisation)

a) Les connaissances sémantiques et épisodiques

Une connaissance sémantique est une connaissance linguistique ou verbale. L'ensemble de ces connaissances constitue l'inventaire des mots et des significations qu'utilise un individu pour s'exprimer.

Une connaissance épisodique est une connaissance qui porte sur des événements ou des expériences personnelles. Elle se compose donc d'éléments d'information temporelles et contextuelles. Elle concerne soit l'individu en tant qu'acteur, soit l'individu en tant qu'observateur. Par exemple " le souvenir d'avoir rencontré une personne hier soir à la salle de lecture " est une connaissance épisodique.

2.5.2/ Les connaissances procédurales

Une connaissance procédurale est une connaissance qui porte sur une action ou sur une habilité cognitive ou perceptivo-motrice orientée vers un objectif : elle concerne, en définitive le " comment faire " ou le " savoir faire "

Exemple :

- savoir patiner
- savoir trouver la valeur de l'hypoténuse d'un triangle rectangle

La connaissance procédurale possède un caractère plus dynamique que la connaissance déclarative. Lorsqu'elle est activée son résultat ne consiste pas seulement en un retrait mais en une transformation d'information.

3. L'Ingénierie des connaissances " IC "

3.1. Présentation

Le terme d'Ingénierie de la connaissance " IC " (Knowledge Engineering) fut introduit par Feigenbaum (en 1977) dans le contexte de l'intelligence artificielle. Il définit l'IC comme l'art d'acquérir, de modéliser et de représenter la connaissance en vue de son utilisation par des ordinateurs qui accomplissent ainsi des tâches dites intelligentes.

D'une manière générale, " l'IC porte sur l'instrumentation technique des contenus pour les exploiter dans un cadre où ils sont mobilisés pour leur signification. Par conséquent elle porte sur la représentation informatique du sens pour que les systèmes construits puissent avoir leur comportement paramétré par cette représentation "¹. Le sens peut être le sens d'une activité ou d'un document (i.e. comment le système traite le document en fonction de ce qu'il signifie ?)

A notre sens, l'IC correspond à l'étude des méthodes et techniques permettant de manipuler et/ou de construire des systèmes informatiques utilisant le concept de la connaissance.

L'IC se base sur deux principes fondamentaux :

➤ **L'acquisition des connaissances**

Cette première action concerne l'acquisition des connaissances par l'intégration de techniques bibliométriques et d'ingénierie linguistique spécialisée dans le langage écrit.

➤ **L'organisation des connaissances**

Organiser la connaissance existante afin d'accroître l'efficacité de la compréhension humaine.

3.2. Rôle de l'ingénierie des connaissances

De part les concepts et méthodes mis en œuvre au regard des champs d'application, l'ingénierie des connaissances propose des méthodes et des outils tels que :

- La modélisation cognitive
- La modélisation conceptuelle
- Des langages de modélisation et de représentation
- Des méthodes de travail sur corpus
- etc.

¹ <http://ic2002insa-rouen.fr/pagemenu.html>

pour des domaines comme :

- L'acquisition des connaissances à partir de textes
- La recherche de l'information sur le WEB
- La gestion et la capitalisation de connaissances en entreprise
- La définition d'indicateurs de gestion et leur mise en œuvre dans des systèmes d'information.

4. Les sciences de l'information et l'IC

La nouvelle tendance de l'ingénierie de connaissance est de formaliser et de représenter la connaissance sans sujet connaissant, c'est-à-dire, celle qui est enfouie dans les données en IST (publications scientifiques, littérature grise, documentation technique ou clinique, brevets) ou celles détenus par les experts.

Brookes¹ soutenait l'idée que la science de l'information a besoin d'une théorie de la connaissance objective plutôt que subjective, il rappelle une distinction entre le document et la connaissance. Ce ne sont pas des entités identiques. Il faut donc distinguer entre l'ingénierie documentaire et l'ingénierie de la connaissance écrite. Selon le même auteur, la connaissance est une structure de concepts liés par leurs relations et l'information est une petite partie de cette structure.

En effet, notre état de connaissance sur un sujet donné, à un moment donné, est représenté par une structure de concepts liés par leurs relations : " c'est notre image du monde "2. Lorsque nous constatons une déficience ou une anomalie de cet état de connaissance, nous nous trouvons dans un état anormal de connaissance³. Nous essayerons d'obtenir une ou des informations qui vont annuler cette anomalie. Il en résultera un nouvel état de connaissance. C'est ce que Brookes⁴ a voulu schématiser et représenter sous la forme de ce qu'il a appelé l'équation fondamentale des sciences de l'information :

¹ B.C. Brookes, " The foundations of information science ", Journal of Information Science, 2, 1980

² K.E. Boulding, " The image : Knowledge in life and society ", Ann Arbor, University of Michigan Press, 1956

³ N.J. Belkin, " Anomalous states of knowledge as a basis for information retrieval ", Canadian Journal Information Science, 5, 1980

⁴ Y. F. Le Coadic.- La Science de l'information.- Paris : PUF. – collection (Que sais-je ?), 1994

$$K(S) + \partial K = K(S+\partial S)$$



qui exprime le passage d'un état de connaissance $K(S)$ à un nouvel état de connaissance $K(S+\partial S)$ par l'apport d'une connaissance ∂K extraite d'une information ∂I , ∂S indiquant l'effet de cette modification ¹.

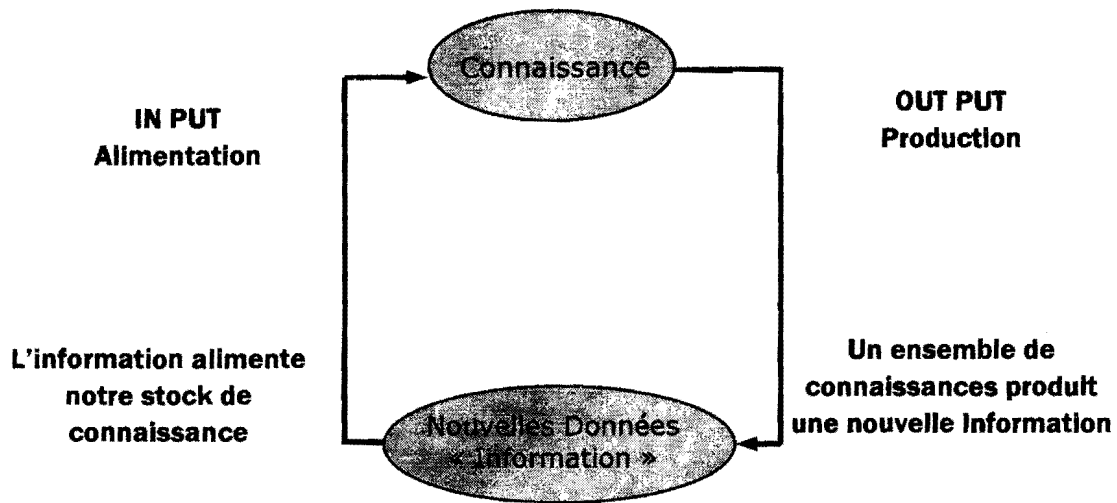


Figure 2 : Relation Information/ Connaissance

Cette étroite relation qui existe entre Information et Connaissance permet implicitement l'interférence entre les sciences de l'information et l'ingénierie des connaissances. le lien existe aussi lorsqu'il s'agit de se rapprocher de la connaissance contenue dans les documents sous la forme de langage écrit. Dans ce sens l'extraction des connaissances et la recherche d'information se définie comme la visualisation du contenu d'une base de données utilisant des méthodes bibliothéconomiques et bibliométriques qui se rapportent à différentes opérations :

¹ Yves François Le Coadic.- La science de l'information. Op cit.

- Traitement automatique du langage naturel pour la recherche, l'indexation, et la représentation du savoir : ces opérations sont fondées sur les analyses linguistiques qui représentent des méthodes de traitement local du groupe nominal permettant de faire un repérage des termes et des variantes de termes. Ces analyses s'appuient sur le fait que dans un texte scientifique et technique, l'information se focalise de manière privilégiée dans les " groupes nominaux "

- Classification automatique ou cartographie du contenu cognitif et factuel d'une base de données : Etablir des catégories qui rassemblent des unités d'information ayant des caractéristiques communes fournissant une représentation visualisation d'une structure de classes

- Accès et recherche de l'information : Par le raffinement et/ou l'extension des structures des langages d'indexation de type thesaurus, exploitation d'autres structures d'organisation des connaissances : Ontologies ; traitement des requêtes ; techniques de filtrage.

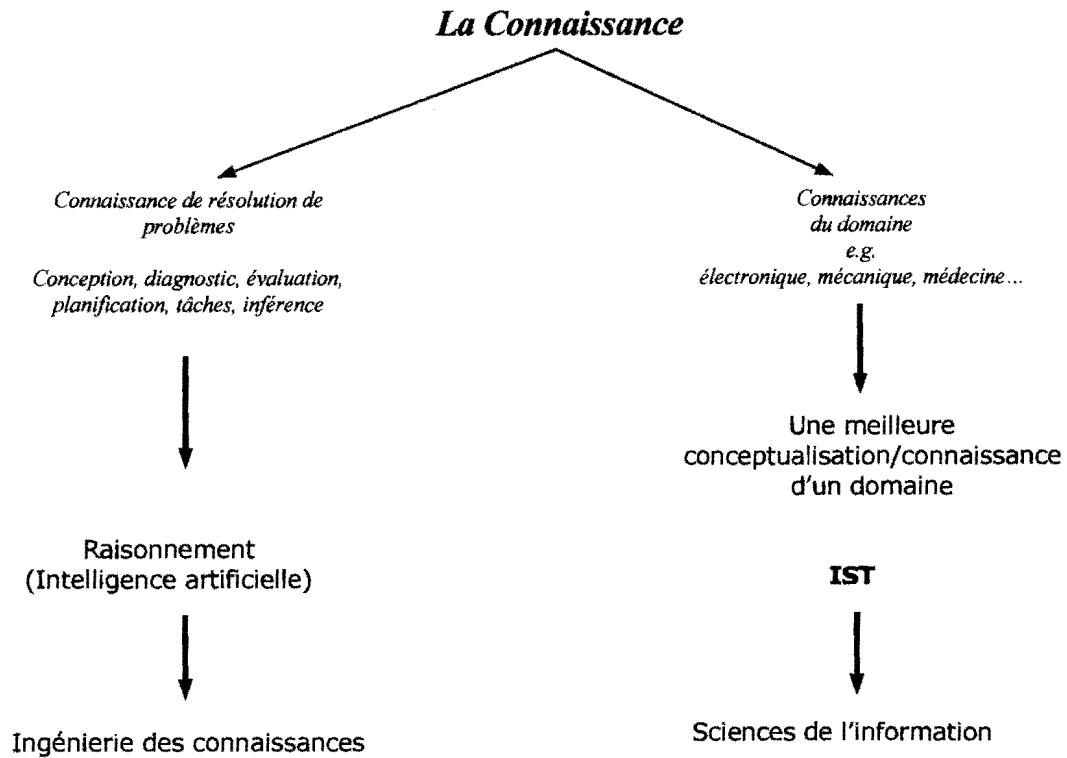
A retenir que La connaissance et son processus d'extraction et de représentation exploitent profondément des méthodes et des techniques utilisées en sciences de l'information, telle que : l'indexation ; bibliométrie, outils linguistiques.

Dans le cas général la connaissance peut être utilisée :

- Pour avoir une image conceptuelle d'un objet, d'un domaine ou d'un processus (i.e. avoir une image du monde qui nous entoure)

- Pour faire des raisonnements et prendre des décisions

En sciences de l'information, il s'agit de la première utilisation. En effet, l'IC à travers ces modèles conceptuels proposés et avec sa nouvelle tendance à élaborer des techniques et méthode de travail en corpus ouvre un champ d'application remarquable pour l'application en IST. Ainsi l'IC intervient en IST pour aider dans la structuration et la recherche de l'information de plus, par l'utilisation du concept de la connaissance elle introduit une nouvelle génération de systèmes appelés systèmes experts de deuxième génération ou encore systèmes à base de connaissances.



Conclusion

La connaissance, s'implique de plus en plus dans les grandes organisations, comme donnée immatérielle et stratégique pour des actions de développement. Sa gestion implique une forte liaison entre les sciences de l'information, l'informatique (Intelligence artificielle) et les sciences cognitives. Car celle ci ne dérive pas seulement de sources explicites (documents, bases de données, tableaux...etc) mais concerne également l'expertise tacite, la perspicacité et l'expérience des experts et spécialistes d'un domaine.

Bibliographie

Brookes, B. C. - The Foundations of information science. –In: Journal of Information Science, 2, 1980

Boulding, K.E. - The image : Knowledge in life and society. – In: Ann Arbor, University of Michigan Press, 1956

Belkin, N. J. - Anomalous states of knowledge as a basis for information retrieval . – In: Canadian Journal Information Science, 5, 1980

CHARLET, J. - M.ZACKLAD, G.KASSEL, D.BOURIGAUT.- Ingénierie des connaissances : Evolutions récentes et nouveaux défis. – Paris : Ed. Eyrolles, 2000

Coadic, Y. F. - La science de l'information. – Paris : Presses Universitaires de France, 1994. - Collection " Que sais-je "

Jacquard-choulet.- Proposition d'un modèle cognitif axé sur l'auto organisation de la connaissance pour la résolution de problèmes.- Thèse de l'université de Besançon; 1993

Micaelli, J.P. - Les trois ages de la capitalisation de la connaissance.- In : J. M. Fouet, connaissance et savoir faire en entreprise, Ed Hermes, 1997.

Polanco, X. - Extraction et modélisation des connaissances : une approche et ses techniques (EMCAT).- In : Organisation des connaissances en vue de leur intégration dans les systèmes de représentation et de recherche de l'information, Université Charles de Gaulles, Lille3 , 1999

Vink, D. - La connaissance, ses enjeux et ses institutions. In : connaissance et savoir-faire en entreprise, J. M. Fouet, ed HERMES, Paris, 1997.

Warnier J.D. - L'homme face à l'intelligence artificielle.- Ed. Organisation, 1984.

<http://ic2002insa-rouen.fr/pagemenu.html>

<http://www.geocities.com/SouthBeach/Marina/8256/turing.htm>

<http://www.limsi.fr/Individu/erika/representations.html>

