

**enssib**

Ecole Nationale Supérieure des Sciences  
De l'Information et des Bibliothèques

**DEA**  
**Sciences de**  
**L'Information et de la**  
**Communication**

**Option :**  
**Systèmes d'information**  
**documentaire**

**MEMOIRE DE**  
**DEA**

**Apport des nouvelles**  
**technologies dans les systèmes**  
**d'Information et de**  
**Communication: orientations**  
**et perspectives**

**POULET Lionel**

Sous la direction de :

Monsieur BOUCHÉ Richard  
Monsieur BOUZIDI Laïd

SEPTEMBRE 1997

Université Lumière  
Lyon 2

Ecole Nationale Supérieure  
Des Sciences de l'Information  
et des Bibliothèques

Université Jean Moulin  
Lyon 3

BIBLIOTHEQUE DE L'ENSSIB



8027060

**enssib**

**Ecole Nationale Supérieure des Sciences  
De l'Information et des Bibliothèques**

**DEA  
Sciences de  
L'Information et de la  
Communication**

**Option :  
Systèmes d'information  
documentaire**

**MEMOIRE DE  
DEA**

**Apport des nouvelles  
technologies dans les systèmes  
d'Information et de  
Communication: orientations  
et perspectives**

**POULET Lionel**

Sous la direction de :  
Monsieur BOUCHÉ Richard  
Monsieur BOUZIDI Laïd

SEPTEMBRE 1997

*L'intelligence organise le monde en s'organisant elle-même.*

PIAGET 1937

## **Remerciements**

L'aide et la pertinence de Monsieur BOUCHÉ Richard, professeur à l'ENSSIB et de Monsieur BOUZIDI Laïd, Maître de conférence à l'Université Jean Moulin LYON 3, ont été très importantes et appréciées et ont grandement contribué à l'aboutissement de ce mémoire. Je leur adresse mes vives remerciements.

J'adresse toute ma reconnaissance à l'ensemble du personnel de l'ENSSIB, de l'Université Lumière LYON 2 et de l'Université Jean Moulin LYON3 pour leur dévouement à l'égard des étudiants de ce DEA.

Lionel POULET

Sous la direction de  
Richard BOUCHÉ      Laïd BOUZIDI  
ENSSIB                      LYON 3

**Résumé :**

Le système d'information et de communication est actuellement la clé de voûte de l'entreprise. Il permet de collecter, de traiter, de communiquer toutes les informations stockées, archivées et générées par l'entreprise. Toutes ces informations sont capitales pour sa survie et représentent son histoire et son savoir-faire.

Quelles sont les orientations et les perspectives de l'apport des nouvelles technologies dans les différentes phases de gestion de cette information?

Les systèmes dits d'aide à la décision et les systèmes de Gestion Electronique de Documents répondent à cette multitude d'informations et permettent de mettre en avant de nouvelles pistes de recherche.

**Descripteurs :** Systèmes Interactifs d'Aide à la Décision (SIAD) ; Gestion Electronique de Documents (GED) ; Aide à décision ; Systèmes d'Information et de Communication (SIC) ; Nouvelles technologies.

**Abstract :**

At the present time, the information and communication system is the keystone of the company. It allows the collection, the processing and the communication of all the data stored, recorded and generated by the company. All this data is its history and know-how and is therefore of capital importance in order to everlast.

What are the trends and prospects of the new technologies in the different stages of the data processing?

The so-called decision and electronic processing systems of the documents meet the multitude of informations and opens new tracks of searches.

**Keywords :** Decision Support Systems (DSS) ; Electronic Documents Management (EDM) ; Decision aid ; Information and Communication Systems (ICS) ; New technology.

## Table des matières

REMERCIEMENTS.....	2
TABLE DES MATIÈRES.....	4
INTRODUCTION.....	6
LES SYSTÈMES DITS D'AIDE À LA DÉCISION (SIAD).....	11
CONTEXTE D'AIDE À LA DÉCISION.....	12
ORIGINES.....	12
DÉFINITION.....	13
ARCHITECTURE.....	13
<i>Modèles de coopération homme-machine à base d objets multimédia</i> .....	14
<i>Modèles de connaissances multi-agents</i> .....	14
CONCLUSION.....	16
LES SYSTÈMES DE GESTION ELECTRONIQUE DE DOCUMENTS (GED).....	18
LA GED EN COMPLÈMENT DES APPLICATIONS INFORMATIQUES.....	19
LES DIFFÉRENTS CATEGORIES ET SOLUTIONS GED.....	21
<i>La GED Administrative</i> .....	21
<i>La GED Bureautique</i> .....	21
<i>La GED COLD</i> .....	22
<i>La GED Documentaire</i> .....	22
<i>La GED Technique</i> .....	23
INFRASTRUCTURE GÉNÉRALE D'UN SYSTÈME GED.....	23
L'ACQUISITION DES DOCUMENTS DANS UNE SOLUTION GED.....	24
L'ACQUISITION DES INDEX DANS UN SYSTÈME GED.....	24
<i>OCR ou reconnaissance optique des caractères</i> .....	26
<i>Le code à barres pour l'indexation automatique de documents numérisés</i> .....	27
LES TECHNIQUES ET MÉTHODES D'INDEXATION.....	27
STOCKAGE ET CONSERVATION DES DOCUMENTS NUMÉRISÉS DANS UN SYSTÈME GED.....	29
RECHERCHE, RESTITUTION, DIFFUSION DES INFORMATIONS ET DES DOCUMENTS DANS UN SYSTÈME GED.....	30
OUTILS D'ÉCHANGE, WORKFLOW INTÉGRÉS AUX SOLUTIONS GED.....	31
L'APPLICATION DE GESTION ELECTRONIQUE DE DOCUMENTS: "MULTIGEST" ©.....	32
AVÈNEMENT DES NOUVELLES TECHNOLOGIES.....	35
INTRANET/INTERNET.....	36
TÉLÉCOMMUNICATIONS - LA RÉVOLUTION NUMÉRIQUE.....	38
<i>Les différents modes de communication et les usages</i> .....	38
GROUPWARE.....	39
DATA WAREHOUSE.....	40
MULTIMÉDIA.....	41
EDI: INGÉNIERIE COLLABORATIVE.....	43
<i>Contexte de l'ingénierie collaborative</i> .....	43
CONCLUSION.....	45
IMPACT DES NOUVELLES TECHNOLOGIES SUR LES SYSTÈMES DITS D'AIDE À LA DÉCISION ET SUR LE SYSTÈMES DE GESTION ÉLECTRONIQUE DE DOCUMENTS.....	46
SYSTÈME INTERACTIF D'AIDE À LA DÉCISION "AVANCÉ".....	47
<i>Présentation</i> .....	47
<i>Sous-système A: La modélisation des connaissances de l'utilisateur</i> .....	50
<i>Sous-système B: Base de données</i> .....	51
<i>Sous-système C: Base de décisions</i> .....	52

SYSTÈME DE GESTION ELECTRONIQUE DE DOCUMENTS .....	52
<b>CONCLUSION.....</b>	<b>55</b>
<b>GLOSSAIRE .....</b>	<b>57</b>
<b>BIBLIOGRAPHIE.....</b>	<b>66</b>

## **Introduction**

*"Il était fatal que le développement des nouvelles technologies ronge cette culture en différé de l'écrit. La photographie et la montée en puissance corrélative de l'image, la radio, le téléphone, la télévision ou les immenses prouesses du multimédia débitées par les autoroutes de l'information, ne peuvent que modifier les régimes de la mémoire." (BOUGNOUX, 1995).*

L'information est partout présente dans notre existence. Les messages les plus variés ne cessent de nous parvenir de tous les points du globe, sous des formes de plus en plus diversifiées, au point que certains s'inquiètent de l'ampleur de ce flot où il devient de plus en plus malaisé de discerner l'essentiel de l'accessoire. Cet ouragan d'informations en provenance simultanée de tous les fronts place l'homme moderne dans une situation embarrassante. Dès lors, la nécessité d'accroître encore le potentiel de diffusion, de communication, de collecte des faits et des documents antérieurs apparaît impérieuse. La mémoire humaine est insuffisante, le savoir consigné dans les livres ne peut servir à quiconque si la consultation des ouvrages et au sens le plus large les différents supports d'informations, de plus en plus nombreux n'est pas rendue plus aisée par quelque méthode plus pratique et rapide que la recherche dans une bibliothèque. Les services d'information utilisant des mémoires électroniques se sont considérablement multipliés; déjà, dans de multiples applications, les procédés traditionnels d'écriture ont été abandonnés.

L'ordinateur n'est que le prolongement logique de la téléphonie, de la radiophonie, de la vidéophonie; son emploi s'impose dès qu'il importe de trier cette multitude de sons, de phrases, d'images, ce qui explique que la théorie de l'information doit ses acquisitions les plus remarquables aux recherches menées par une entreprise de télécommunications.

L'usage a conféré au mot "information" un sens extrêmement étendu. Dire que l'information est partout présente signifie simplement que nous pouvons accroître nos connaissances par l'observation et l'analyse de tout objet ou de tout phénomène.

En premier lieu, le contenu d'un message échappe à l'observation et ne peut être

défini que d'une manière relative, comme un apport de renseignements. Dès lors que le message est reçu par plusieurs personnes il apparaît que la même forme peut véhiculer plusieurs contenus qui dépendent entre autre des connaissances antérieures des récepteurs.

Un caractère commun à tous les messages est qu'ils constituent une représentation appauvrie de l'objet décrit, mais mettant par contre plus nettement en lumière les similitudes entre cet objet et certains autres.

Dans un inventaire aussi bref de la nature et des qualités de l'information, la notion de récepteur est à peine esquissée. En fin de course, le seul bénéficiaire d'une information sera toujours un homme ou une communauté humaine. Or l'étude de ce qu'on peut appeler la capacité d'absorption humaine en matière d'information montre que celle-ci n'est pas à la mesure de son appétit de connaissance. Les sens ne permettent pas de capter ce qui est trop petit ou trop grand, trop rapide ou trop lent; aussi les instruments doivent-ils pallier les déficiences de la faculté humaine de perception. Mais alors, la masse accrue de faits rendus perceptibles par voies indirectes se révèle trop considérable pour que le pouvoir de réflexion en tire tout le bénéfice souhaitable dans le bref laps d'existence qui est imparti à l'homme. Là encore les auxiliaires mécaniques sont indispensables pour exécuter les opérations simples de collecte, de présentation, de diffusion et quelques opérations d'analyse et de synthèse bien définies, afin de pouvoir reporter les efforts de réflexion et d'inspection de voies nouvelles dans les domaines où les machines n'apportent pas encore le moindre secours.

La technologie de l'information concerne toutes les formes d'aide méthodologique ou matérielle permettant à l'homme de concentrer son attention sur les aspects les plus importants des messages. En ce sens, la technologie de l'information est apparue avec le langage parlé, puis avec l'écriture et le calcul.

Les grands problèmes relatifs à l'information sont liés à sa collecte, son traitement, sa diffusion, sa conservation. L'état de développement des solutions varie grandement d'un secteur à l'autre. Une même question se pose à propos de la collecte et de la diffusion: où trouver l'information utile, à qui convient-il de faire parvenir telle information connue? Les réponses à ce genre de

préoccupations sont malheureusement encore largement empiriques ainsi qu'en témoignent la difficulté d'établir des statistiques, les sommes dépensées en enquêtes, les obstacles à la jonction harmonieuse de l'offre et de la demande dans le marché de l'information en pleine expansion.

L'arsenal technologique de la diffusion d'information est vaste: procédés d'impression, de duplication, de transmission à distance, il semble inutile d'en faire la revue détaillée, d'autant plus que ces techniques constituent seulement le maillon terminal des circuits de l'information traitée par voies mécaniques. Par contre, il convient d'analyser plus précisément les techniques de transformation et de mémorisation de l'information qui paraissent souvent mystérieuses simplement parce qu'elles échappent à l'observation directe.

Il est aujourd'hui inconcevable qu'un ordinateur soit capable d'exécuter uniquement les fonctions câblées dans sa structure matérielle, et chaque nouvelle machine doit être conçue avec une bibliothèque de programmes permettant d'en étendre l'usage à des applications très diverses. On ne saurait citer tous les domaines où l'emploi d'ordinateurs a pratiquement supplanté les hommes dans leur travail, ni les secteurs où ils prennent la relève de personnels voués à des tâches ingrates de surveillance: l'aspect quantitatif et social de compression d'un personnel peu qualifié a été très souvent évoqué avec beaucoup d'exagération à propos de l'implantation de systèmes informatiques, certains voyant dans cette évolution une révolution analogue à celle qui provoqua les profonds remous du début du machinisme. La comparaison n'est guère justifiée car, si les machines mues par une énergie de combustion se sont révélées très avantageuses pour libérer l'homme de peines et de fatigues musculaires, les ordinateurs électroniques ne peuvent aucunement se substituer à l'homme pour exercer à sa place la domination du monde par l'intelligence. Bien au contraire, les reconversions qu'implique l'avènement de l'informatique et des moyens de communication de l'information nous conduisent à mieux réfléchir sur nos méthodes de travail et souvent à les modifier au lieu de nous conformer à des habitudes profondément ancrées.

Nous sommes entourés par une masse d'informations, sans cesse grandissante, l'analyse des systèmes d'information et de communication et plus particulièrement dans les domaines des systèmes d'aide à la décision (SIAD) et des systèmes de gestion électronique de documents (GED) me paraît intéressante pour comprendre la gestion de toutes ces informations qui sont à la disposition de l'utilisateur.

L'évolution de l'informatique est de plus en plus forte, tant sur le plan physique que conceptuel. Cette évolution a une influence sur les deux domaines considérés précédemment qui sera présentée dans ce mémoire.

Pourquoi avoir choisi les systèmes d'aide à la décision et les systèmes de gestion électronique de documents ?

Les SIAD sont des systèmes d'informations et de communication spécialisés dans l'aide à la décision. Ils doivent pour cela travailler interactivement avec les décideurs et l'intégrer dans le processus de prise de décision. Le décideur a donc besoin pour prendre une décision d'avoir le plus grand nombre d'information à sa disposition concernant le problème posé et de pouvoir en communiquer un maximum.

Nous pensons que la GED peut apporter une partie de ces informations. De plus, depuis deux ans, je collabore au développement d'une application GED "MultiGest" ©<sup>1</sup>. Enfin, ce choix me paraît très intéressant car il présente deux domaines qui sont actuellement sous les feux des projecteurs et qui concernent de plus en plus les chercheurs car ils intègrent les nouvelles technologies et les nouveaux besoins des utilisateurs répondant de plus en plus à leurs exigences dans la collecte, le traitement et la diffusion ou communication de l'information.

Le but de ce mémoire est de présenter dans une première partie les systèmes d'aide à la décision en indiquant l'orientation des recherches actuelles. Dans une deuxième partie, nous nous intéresserons aux systèmes de Gestion Electronique de Documents en proposant une présentation sommaire de ce domaine. Ensuite, dans une troisième partie sera présentée une liste non exhaustive des nouvelles

---

<sup>1</sup> Logiciel de Gestion Electronique de Documents développés par la société ERIC (49, Avenue de la République 69200 VENISSIEUX).

technologies permettant de mettre à jour de nouvelles perspectives dans les deux domaines présentés précédemment (les SIAD et la GED). Enfin, dans une quatrième partie, nous mettrons en avant ces nouvelles perspectives dans le but de proposer des pistes de travail et de recherche.

## **Les systèmes dits d'aide à la décision (SIAD)**

Le système d'information d'une société était basé il y a encore peu de temps sur des matériels locaux à chaque site, sur le savoir faire des individus, et le flux d'informations circulait sur des supports tangibles et à majorité manuscrits. Pour des raisons de productivité, de dynamisme, de management et de sauvegardes du savoir faire, les sociétés, aidées par l'intégration de l'électronique ont dès les années 70 informatisé leur Système d'Information (SI).

Par mode ou par nécessité économique, le "boom" informatique s'est produit et l'ordinateur n'a cessé d'accroître son phénomène de dépendance dans l'organisation des entités. Le concept est très puissant, mais irréversible ce qui accroît l'impact stratégique de l'informatique qui supporte et véhicule les données du SI. Le développement croissant des SGBD (Systèmes de Gestion de Bases de Données) autorise à penser que l'informatique a d'ores et déjà atteint l'entreprise jusqu'au profond de ses centres vitaux. Désormais, la trésorerie, la comptabilité, la gestion du personnel, la gestion de production, la fabrication sont totalement informatisée et font appel de plus en plus à ce même concept: les SGBD.

Si initialement l'outil informatique a été introduit dans les systèmes d'information et de communication pour gérer tous les flux financiers et matériels représentant une grande partie de l'activité des organisations. Au début des années 1970, l'informatique était plutôt destiné aux traitements structurés ou dits "opérationnels". Depuis les années 1980, de nouvelles approches sont apparues s'orientant vers l'intégration de traitements dits "non-structurés" difficiles à structurer dans un programme. Cette nouvelle approche dite orientée vers "l'Intelligence Artificielle" s'est soldée par de nouvelles méthodes de conception, de développement touchant plutôt les connaissances que les informations. Cette orientation se caractérise par la naissances des Systèmes à Base de Connaissances. L'intérêt que porte le monde actuel à l'outil informatique est due à la nécessité accrue de disposer de l'information le plus rapidement possible, et la nécessité de communiquer et de transmettre des informations. Les moyens offerts par les nouvelles technologies conditionnent les communications, et le transfert du savoir et du savoir-faire.

Selon la définition qui en a été donnée pour la première fois par M.S. SCOTT MORTON (1971) "les SIAD sont des systèmes informatiques de type interactif qui ont pour rôle d'aider les décideurs à utiliser des données et des modèles pour résoudre des problèmes non structurés".

### ***Contexte d'aide à la décision***

Avant de définir plus amplement le concept de SIAD, il paraît nécessaire d'introduire la notion de processus de décision dans les organisations. SIMON (1977) distingue quatre phases dans le processus de décision:

- une phase de diagnostic d'un problème et d'exploration-reconnaissances des conditions dans lesquelles il se pose: c'est la phase d'intelligence;
- une phase de conception et de formulation de voies possibles offertes à la résolution du problème: c'est la phase de conception;
- une phase de choix d'un mode d'action particulier parmi les actions possibles: c'est la phase de sélection;
- une phase d'évaluation de la solution provisoirement retenue comme satisfaisante. Cette phase de bilan peut déboucher sur la réactivation de l'une des trois phases précédentes ou, au contraire, sur la validation de la solution, reconnue comme finalement satisfaisante.

L'aide à la décision repose sur une exploration, à divers niveaux, dans divers espaces de résolutions<sup>2</sup>, suivant des méthodes heuristiques. L'aspect "aide à la décision" est bien mis en évidence, c'est à dire que l'utilisateur prend lui-même la décision, et l'on comprend comment cette interactivité est nécessairement associée à l'absence de structuration et de normalisation du problème dans son ensemble, ce qui n'exclut pas la modélisation de certaines phases du processus de réflexion stockée dans une base de modèles.

### ***Origines***

C'est en observant comment les hommes résolvent certains problèmes que NEWELL et SIMON (1972) vont nous apporter les concepts de base sans lesquels on ne peut comprendre les SIAD.

---

<sup>2</sup> Cet espace de résolutions, défini par NEWELL et SIMON, est constitué de l'espace des états, des opérateurs et des informations disponibles à chaque état.

Suivant le modèle GPS (General Problem Solver) de NEWELL, SHAW et SIMON, la résolution d'un problème donné s'effectue en progressant, de situation en situation, d'une situation initiale à la résolution. Cette progression vers la résolution du problème met en évidence le concept de retour en arrière (backtracking strategy). L'ensemble de ces situations ou états représentant l'espace des états. La résolution de problèmes part d'une représentation formalisable en un espace d'états et d'une aptitude à explorer intelligemment cet espace d'états. Pour passer d'un état à un autre on utilise un opérateur. Les états, les opérateurs permettant de passer d'un état à un autre et les informations dont on peut disposer à chaque état, forment une représentation implicite du problème. C'est cet ensemble que NEWELL et SIMON appellent l'espace de résolution. Cette progression est aussi appelée HEURISTIQUE. Cette approche heuristique est le fondement des systèmes dits d'aide à la décision.

### ***Définition***

Une des définitions les plus citées est celle de KEEN et MORTON (1978):

*"Les SIAD impliquent l'utilisation d'ordinateurs pour:*

- *assister les décideurs dans leur processus de décision dans des tâches semi-structurées;*
- *aider plutôt que remplacer le jugement des décideurs;*
- *améliorer la qualité de la prise de décision plutôt que l'efficacité."*

### ***Architecture***

Comme le précisent POMEROL et LEVINE (1989), les fonctionnalités d'un SIAD sont assurées par trois sous-systèmes ou modules:

- une interface de dialogue,
- une base de données et un Système de Gestion de Base de Données,
- une base de modèles et un Système de Gestion de Base de Modèles.

L'exploitation et le contrôle de ces trois sous-systèmes sont assurés par un module dit "Module de Contrôle de Fonctions". Ils sont complétés par l'intervention des décideurs essentiellement pour l'élaboration de la décision et sa communication.

L'ensemble des données traitées par un SIAD est représenté par la base de données sur laquelle s'applique un processus structuré, basé sur l'exploitation de la

base de modèles. Le décideur n'intervient dans le processus que par l'introduction de paramètres, son domaine d'intervention est limité par ce que lui permettent les bases de données et de modèles du SIAD. Dans ces conditions, l'ensemble des connaissances du (des) décideur(s) est partiellement ignoré. Pour pallier à cette faiblesse, il sera proposée une nouvelle architecture de SIAD dans la quatrième partie de ce mémoire.

Les travaux actuels de recherche<sup>3</sup> visent à renforcer l'architecture des SIAD en intégrant deux types de modèles: des modèles de coopération homme/machine à base d'objets multimédia et des modèles de connaissances multi-agents.

### **Modèles de coopération homme-machine à base d'objets multimédia**

La base de dialogue d'un SIAD a pour rôle d'accroître la capacité de compréhension et de déduction des décideurs en multipliant les formes de représentation des connaissances (son, texte, image, graphique,...). Les modèles de coopération homme/machine à base d'objets multimédia ont pour objectif d'aider les décideurs à naviguer dans un espace de recherche formé de représentations audio-visuelles de la connaissance. Les recherches qui ont été conduites depuis une vingtaine d'années dans le domaine de l'ergonomie cognitive (McKIM 1972), et plus particulièrement dans le contexte des SIAD, ont montré que la qualité des décisions est basée sur un mode d'accès associatif aux supports de la connaissance. En d'autres termes, les messages que les décideurs reçoivent du SIAD déclenchent en eux des mécanismes de raisonnement par association (pattern matching): la multiplication des formes de représentation de ces messages les aident à imaginer (visual thinking) des stratégies de résolution aux problèmes qu'ils rencontrent.

### **Modèles de connaissances multi-agents**

Tout processus de décision dans l'entreprise place des décideurs, appartenant ou non à un même niveau hiérarchique, en situation d'interdépendance. Les SIAD de groupe (Group Decision Support Systems) et les SIAD de direction (Executive Information and Decision Systems) ont fait apparaître la nécessité d'un partage du

---

<sup>3</sup> "Les systèmes interactifs d'aide à la décision: facteurs de succès de l'IA en gestion de l'entreprise", ERNST Christian, 1993, 1<sup>er</sup> Congresso Latino Americano de engenharia Industrial, Floréanopolis BRESIL.

savoir-faire des décideurs dans l'exercice de leurs responsabilités (Gray 1986). Les systèmes multi-agents (SOUBIE et GLEIZES 1986) ont pour but de répondre à la problématique de ce type de SIAD plaçant des décideurs en situation fortement interactive (réunion de direction, négociation, communication, ...). Ces systèmes sont basés sur une architecture distribuée de la base de connaissances, respectant l'organisation physique des centres de décision: chaque unité (ou monde) de la base de connaissances fournit un modèle de comportement d'un agent expert dans un domaine de connaissance spécifique. L'apport de l'I.A. Distribuée (IAD) dans les SIAD est particulièrement efficace lorsque le comportement des décideurs fait appel à des mécanismes de coopération basés sur deux types de modèles (CHAIB-DRAA 1990):

- des modèles de négociation orientés vers la résolution de conflits;
- des modèles de planification dynamique orientés vers la coordination de plans d'action associés à chaque agent.

Tous ces modèles coopératifs à base de connaissances multi-experts sont soumis à des contraintes importantes de temps de réponse nécessitant des modèles de coopération SIAD/décideurs fortement interactifs.

Dans les recherche actuelles, les mécanismes de coopération entre décideurs sont implémentés à l'aide d'une architecture Blackboard (ENGELMORE et MORGAN 1988, NII 1986). Ce modèle multi-agents permet d'organiser le dialogue entre le SIAD et le décideur en plaçant celui-ci en situation de négociation avec les autres décideurs, représentés dans le système par des agents: la négociation peut porter sur la distribution des problèmes à résoudre, l'affectation des ressources, le choix des actions sur les contraintes. Les connaissances stratégiques qui sont mises en oeuvre dans un SIAD multi-agents sont de deux types:

- des connaissance liées à la fonction de coordination du superviseur: contrôle de cohérence des tâches effectuées par chaque agent, apportant à chacun d'eux une visibilité sur l'état courant du problème traité (contexte) et sur l'état courant de la solution (degré de résolution de chaque sous-problème);
- des connaissances liées au savoir-faire de chaque agent: choix des moyens. à mettre en oeuvre pour atteindre les objectifs fixés par le superviseur (en

termes de sous-problèmes à résoudre), actions sur les contraintes spécifiques à chaque agent (contraintes d'exploitation, contraintes financières, contraintes commerciales, ...).

Un SIAD multi-agents basé sur une architecture Blackboard permet d'accroître l'efficacité globale des décisions en simulant les interactions entre décideurs lors la résolution d'un problème par l'un d'entre eux.

### ***Conclusion***

En fine de compte, la machine reste un automate, et les agents intelligents ne sont que de simples assistants. La décision finale est de la compétence du décideur. La machine n'apporte pas la solution, mais joue le rôle d'amplificateur cognitif, c'est-à-dire qu'elle met l'accent sur une ou plusieurs phases du processus de décision.

Cet aspect purement décisionnel, nous ouvre la perspective de pouvoir stocker l'ensemble des heuristiques réalisées par le décideur dans une base de décisions. Cette base de décisions permet une meilleure capitalisation du savoir-faire du décideur et de rentrer dans la résolution de nouveaux problèmes.

L'évolution des besoins du décideur par l'intermédiaire de l'avènement des nouvelles technologies, de la mondialisation des échanges de l'information, de la démocratisation de l'outil informatique, et de la diminution des coûts pourrait permettre une interconnexion de ces bases de décisions au sein d'une entité économique. Cette interconnexion permettra une prise de décision par un ensemble de décideurs sur une base de décisions commune sans pour autant être dans un même lieu.

Pour constituer cette base de décisions le décideur doit avoir à sa disposition le plus grand nombre d'informations.

De plus en plus, les entreprises se sont lancées dans l'informatisation généralisée de l'ensemble des informations stockées par celle-ci sous les formats les plus divers. C'est l'apparition du concept de la Gestion Electronique des Documents (GED). Nombre d'entreprises et d'administrations ont opté pour ces techniques informatiques qui permettent de rationaliser la gestion de l'information et des documents. Elles y trouvent matière à de substantielles économies et améliorent de façon sensible les relations qu'elles entretiennent avec leurs clients ou leurs administrés. Les avantages que procurent le GED sont bien sûr manifestes dans

les grandes structures brassant beaucoup de documents ou générant des fichiers informatiques à longueur de temps. Mais la GED est utile pour les organisations de toute dimension. L'offre est aujourd'hui suffisamment diverse pour répondre à une multitude de besoins, du plus simple au plus complexe.

## **Les systèmes de Gestion Electronique de Documents (GED)**

Aujourd'hui la GED s'intègre dans l'environnement informatique et lui apporte les outils logiciels nécessaires pour gérer électroniquement les documents et l'ensemble des informations qui circulent dans les entreprises.

On rencontre plusieurs définitions et divers acronymes quand on commence à s'intéresser à la gestion électronique de documents et d'informations.

Le terme GED pour Gestion Electronique de Documents est apparu dans le milieu des années 80 et s'est substitué au terme "archivage électronique" précédemment utilisé.

En 1994, les professionnels français réunis au sein de l'association APROGED ont décidé d'utiliser l'acronyme GEIDE (qui provient de Gestion Electronique d'Informations et de Documents Existants) pour dépasser ce que l'appellation GED avait de restrictif dans son aspect document et élargir la gestion à tout type d'information, en particulier aux données de source informatique.

Dans les deux cas, ces acronymes désignent les mêmes technologies et les mêmes solutions.

74 % des 237 directeurs informatiques et Télécom interviewés en juin 1991 par Informat/SOFRES attendent d'un système de GED un meilleur accès à l'information. Puis de manière dégressive, une meilleure productivité, une meilleure maîtrise de l'archivage ainsi qu'une meilleure gestion de l'information.

Une seconde enquête fait ressortir quelques points divergents de l'étude précédente. Ce sont en effet, le gain de place, la réduction des coûts de stockage et la rapidité d'accès à l'information qui sont les premiers avantages cités. Le fait que 40% des utilisateurs confondent GED et archivage explique en partie ces résultats.

Une autre étude souligne trois motivations principales: la suppression des pertes de temps engendrée par une recherche documentaire classique; une réponse plus rapide aux demandes d'information et la réduction des pertes financières générées par une mauvaise utilisation de l'information.

Une dernière enquête liée plus précisément aux milieux documentaires met en exergue trois attentes principales: accroître l'accessibilité de l'information, obtenir des gains de temps substantiels et des gains d'espace.

### ***La GED en complément des applications informatiques***

La GED est, au fil des ans, devenue une composante des systèmes d'information des entreprises dans lesquels elle introduit des fonctions de gestion, de partage ou de circulation et d'archivage, fonctions qui viennent seconder des applicatifs et des programmes bureautiques, de gestion ou de traitement de documents. Historiquement, les solutions GED ont pris la suite des systèmes dits "d'archivage électronique" apparus dans le milieu des années 80.

A cette époque, les concepteurs proposaient des solutions autonomes ou parfois connectables à des réseaux locaux, dont la vocation principale était de remplacer les salles d'archives en substituant l'image numérisée du document à l'original. Cette vision n'est plus d'actualité. La GED vient désormais se greffer à une structure informatique aux côtés et en complément d'autres applicatifs et de gestionnaires de bases de données.

Concrètement, la GED est la réponse aux problèmes de gestion, de stockage, de recherche, de consultation, de traitement et de communication de fichiers ou de documents réunis sous forme de dossiers ou non. C'est l'ensemble des techniques qui permettent d'accéder rapidement et le plus économiquement possible aux masses d'informations et de documents générés ou reçus par un organisme, qu'il soit une entreprise ou une administration.

La GED est à vocation universelle. Elle trouve son utilité partout, dans tous les contextes et à tous les niveaux.

Elle le démontre là où les documents abondent comme dans la documentation technique, les dossiers d'assurances, de prêts ou du personnel, les processus industriels, etc. C'est elle qui permet d'accéder sélectivement à des documents en quelques secondes à partir d'un simple micro-ordinateur connecté à un serveur via un réseau local alors que la gestion manuelle des dossiers aurait nécessité plusieurs minutes, si ce n'est plusieurs heures, pour localiser puis regrouper le dossier. Comme nous le verrons par la suite, la GED comporte dans ses aboutissants des notions de sécurité logique et de pérennité de l'archivage si celui-ci est réalisé sur un support adéquat.

Dans l'évolution de l'informatique, la GED est une étape: celle où il devient possible d'associer des documents de toute nature à des applications bureautiques,

administratives, industrielles ou de production. Les solutions GED sont avant tout des ensembles fédérateurs d'outils qui viennent compléter des programmes bureautiques ou autres de fonctions de traitement de documents électroniques. Par document, il faut entendre non seulement la feuille que l'on numérise dans un scanner mais aussi le fichier issu d'ordinateurs centraux, le fichier de son numérique, de vidéo numérique, etc., fichiers que l'on peut relier entre eux au sein d'un dossier en créant des liens logiques.

La GED, c'est aussi le partage entre groupes de travail de documents sous forme d'images électroniques afin de restreindre la circulation des originaux, en somme un moyen de communication. Ce partage va au delà de la simple diffusion. Si on rajoute un programme de workflow<sup>4</sup> à celui de GED, le partage peut être sélectif et permettre la répartition de tâches administratives entre plusieurs agents suivant un circuit préétabli.

L'objectif de la GED, c'est de faciliter, au besoin de sécuriser, l'accès et la consultation des documents et des fichiers fédérés par une base de données. Quelques secondes suffisent pour accéder à l'ensemble des éléments constituant un dossier, les afficher sur l'écran d'un micro-ordinateur, les imprimer ou les diffuser. Les originaux ne sont plus mis en circulation; ils sont physiquement archivés une fois pour toutes, ce qui évite des pertes ou des mauvais reclassements. Les conditions de travail sont en général fortement améliorées par l'installation d'un système de GED comme le sont les relations avec les clients ou les administrés. Les solutions de gestion électronique de documents et d'informations constituent d'excellents outils pour réduire le coût des tâches administratives et font apparaître un retour sur investissement très rapides dans la plupart des cas.

Les fonctions entrant sous le vocable de GED sont nombreuses et ne sont pas forcément les mêmes d'une application à l'autre, même si l'on retrouve toujours un noyau de fonctions de base. Puisque ce sont des image numérisées et non plus des originaux sur papier que véhicule la GED, on peut faire subir à ces images des

---

<sup>4</sup> Système d'ordonnancement des flux de travaux dans une organisation. A la base du WorkFlow, il y a une modélisation des fonctions et procédures de l'entreprise. La plupart des solutions de WorkFlow prennent en charge la régulation des flux de travaux en prenant en compte des notions de synchronisme, de temps d'exécution et des alertes.

traitements électroniques et les faire circuler dans l'entreprise par le ou les réseaux locaux ou vers l'extérieur. Mais en dehors de ces images numérisées, la GED sait aussi prendre en charge la gestion, l'indexation et l'archivage de tout type de fichiers numériques. Ceux, par exemple, générés par des logiciels bureautiques, mais également des fichiers "SPOOLS" issus de systèmes informatiques centraux ou encore des fichiers sonores ou vidéo.

### ***Les différentes catégories et solutions GED***

Les applications de GED se classent en cinq grandes familles qui sont les suivantes:

#### **La GED Administrative**

Ce type d'application concerne généralement ce qu'il est convenu d'appeler la Gestion de Dossiers Electroniques et correspond au classement de documents administratifs divers et variés, souvent sous une forme d'image numérisée. On trouve parmi ces documents des bons de commandes, des factures fournisseurs, des correspondances, des fax...

Généralement la GED Administrative fait partie d'une application globale de gestion et permet à l'utilisateur d'accéder rapidement aux images des documents dont il a besoin, sans avoir à se déplacer ou à encombrer son bureau de dossiers physiques. Bien souvent et particulièrement dans les secteurs d'activité du tertiaire, on intègre des outils de communication et de gestion des flux (workflow) à l'ensemble de l'application; ils permettent d'échanger des documents via le réseau, de demander un complément d'information à un collègue, de soumettre à sa réflexion les documents extraits d'un dossier, de valider un document avec une signature électronique, etc.

#### **La GED Bureautique**

La GED Bureautique est illustrée par les produits de Lotus et Microsoft. Ces produits sont en concordance avec les plates-formes bureautiques classiques, qui se veulent de plus en plus communicantes dans un concept de "travail en groupe" (groupware).

Les outils de GED Bureautique, dont le plus représentatif est Lotus Notes, permettent de manipuler des documents dans leur format bureautique d'origine

(Word, Excel...), de centraliser leur classement sur un ou plusieurs serveurs, d'échanger ces documents par messagerie électronique... Bien souvent, ils intègrent aussi des fonctions de distribution et de télécopie, d'agenda électronique...

### **La GED COLD**

Le terme de COLD, qui est l'acronyme de Computer Output on Laser Disk, est utilisé en référence aux applications de COM (Computer Output on Microfilm) car c'est une technique qui se substitue à celle de la micrographie.

Ce type d'application, appelée "Archivage Electronique" dans les années 80, est certainement la toute première application de la GED. Elle permet de stocker et d'indexer automatiquement l'ensemble des documents générés par des applications de gestion et destinés à l'impression. Le principe consiste à récupérer le fichier d'impression (SPOOL), à le découper suivant une logique définie par paramétrage et à en extraire les critères et valeurs nécessaires à l'indexation.

Les documents concernés sont les états comptables, les factures, les bulletins de salaire, les relevés de compte...

### **La GED Documentaire**

La GED Documentaire est directement issue de la recherche documentaire particulièrement utilisée dans le cadre d'applications du type bibliothèque, documentation scientifique, éventuellement revue de presse. La GED a apporté à cette catégorie d'applications l'accès à l'image du document physique ou encore l'accès à des photos (application de type photothèque).

Ce type de GED se caractérise principalement par ses méthodes d'indexation et de recherche qui font le plus souvent appel à des thesaurus, au full-text (indexation et recherche en texte intégral) avec opérateurs de proximité, à une recherche par concept et parfois à un mode d'interrogation en langage naturel. A l'inverse des applications traditionnelles de gestion, l'indexation n'y est pas structurée au sens classique du terme.

## **La GED Technique**

La GED Technique, aussi appelée dans certains cas GED Métier, concerne toutes les applications dans lesquelles le concept de Gestion Electronique de Documents est directement liée à une profession.

Très souvent, cela se caractérise par la manipulation de documents dont le format est propre à un métier. Les applications de GED Technique en sont l'exemple de plus courant et, parmi elles, les plus nombreuses sont celles pour bureaux d'études, pour la gestion de plans ou la gestion de la documentation technique sur un projet industriel particulier. On y associe souvent la gestion des nomenclatures et la manipulation de documents structurés de type SGML.

### ***Infrastructure générale d'un système GED***

On ne peut pas parler de système de GED sans dire un mot de l'infrastructure générale qui existe nécessairement et dans laquelle le système va s'établir. Cette infrastructure est constituée des éléments suivants:

- un réseau informatique et des moyens de communication. De plus en plus fréquemment, il s'agit du réseau général d'entreprise avec des outils banalisés de communication, d'une part, internes (accès aux applications informatiques et émulation de terminaux, transferts de fichiers, messagerie interne, etc.) et, d'autre part, externes (passerelles fax, télex, accès à des messageries publiques, accès à des banques de données externes, etc.);
- des dispositifs matériels et/ou logiciels appelés serveurs dont l'objet est d'offrir des services aux utilisateurs (services d'impression, de sauvegarde de données, de bureautique, d'applications informatiques, de bases de données, etc.);
- des postes client, banalisés ou spécialisés selon la fonction de l'utilisateur. De plus en plus fréquemment, les postes de travail seront ceux dont l'entreprise est équipée : PC sous MS-Windows, PC sous OS/2, stations UNIX, etc; cette infrastructure est construite et organisée selon des choix et des règles d'architecture, par exemple une architecture de type "client-serveur"<sup>5</sup>.

---

<sup>5</sup> Le client/serveur est un mode d'interaction en réseau permettant à deux processus de communiquer entre eux, appelés respectivement client et serveur. Le client est le processus qui génère les requêtes au serveur. Le serveur est le processus qui exécute les requêtes et qui renvoie la réponse au client. Ce mode de fonctionnement est généralement opposé à la méthode d'accès par partage de fichiers.

### ***L'acquisition des documents dans une solution GED***

Il est d'abord nécessaire que l'information soit sous une forme électronique; il faut donc en faire l'acquisition. Le symbole de l'application de GED est le scanner ou numériseur qui transforme le document papier en une image électronique manipulable et véhiculable sur ordinateur. Certes, c'est un des moyens d'acquisition les plus répandus mais il faut rappeler que la GED a pour vocation de fédérer et de gérer tous les types de documents, y compris ceux qui se trouvent déjà sous une forme électronique. En plus des documents papier, les types de document les plus fréquemment rencontrés sont les suivants : télécopies, documents bureautiques issus de logiciels de traitement de texte, de tableaux, graphiques et plans, fichiers informatiques et en particulier les fichiers SPOOL, sons, images animées, documents composites.

Les techniques d'acquisition doivent être adaptées aux types de documents concernés, aux conditions de l'acquisition et aux volumes en cause.

L'acquisition peut comporter plusieurs étapes : par exemple l'interprétation ou la conversion de format du document, l'adaptation de sa structure et de sa composition, sa récupération par un dispositif de stockage ou de traitement.

A chaque forme de document correspond un type d'appareil ou un dispositif logiciel : scanner, carte fax, logiciel de transfert de fichier.

Les conditions de l'acquisition sont étroitement liées aux types de documents, aux volumes, aux contraintes de l'organisation: service de numérisation associé à un service courrier, répartition des tâches sur les postes de travail des utilisateurs, importation de documents, recours à un scanning center (centre de numérisation à façon), etc.

### ***L'acquisition des index dans un système GED***

En même temps que l'on saisit les documents, il faut procéder à l'acquisition des index qui serviront au classement et à la recherche ultérieure des documents. L'indexation est une étape essentielle et indispensable au bon fonctionnement d'un système de GED. Du choix d'une méthode pertinente d'indexation jusqu'à la conception soigneusement étudiée des index, rien ne doit être laissé au hasard.

De ces index et du système d'indexation dépendent la performance du système de GED et son adéquation aux besoins de l'entreprise. Pour acquérir des index, on a

le choix entre plusieurs techniques, dont la plus simple mais aussi la plus adéquate dans de nombreux cas, est l'acquisition manuelle par remplissage d'un formulaire associé au document. On est confronté là à un des points noirs de la GED car l'acquisition des index est une étape fastidieuse qui peut nuire à la rentabilité du système et entamer la motivation des utilisateurs.

Il est en effet plus long et plus contraignant de classer un document papier dans un système de GED en le numérisant et en l'indexant manuellement que de le glisser dans une chemise ou un dossier. Le gain de productivité, au demeurant nullement contestable, se trouve ailleurs, au niveau de la recherche, de la diffusion des documents et de la suppression des tâches de photocopies ou de reclassement. Cependant, un utilisateur mal informé ou à qui on aura fait miroiter la magie du système de GED, réducteur de toutes les tâches ingrates et champion du gain de productivité, peut se décourager très vite devant cette phase de classement plus contraignante que la méthode traditionnelle.

On peut acquérir ces index par d'autres méthodes qui dépendent des documents traités et des applications mises en oeuvre. Citons l'extraction automatique de mots-clefs par programme pour les documents électroniques formalisés du genre états informatiques, l'analyse full-text (texte intégral) des documents, la reconnaissance optique de caractères pour des textes dactylographiés ou manuscrits, l'utilisation de codes à barres ou toute combinaison de ces différentes techniques.

Il est important de noter concernant la reconnaissance optique de caractères qu'il n'existe pas une mais plusieurs solutions techniques. Chaque problématique est différente et demande de faire appel à des techniques diverses couplant la reconnaissance à proprement parler et des moyens de contrôle, soit visuels et manuels, soit automatisés par programme (contrôle de cohérence des valeurs, champs redondants...). Toujours concernant la reconnaissance optique de caractères, il faut signaler qu'elle est parfois utilisée en complément de l'acquisition des documents (par numérisation) pour transformer l'image numérisée en un fichier texte susceptible d'être retraité par des programmes adaptés (en particulier d'indexation full-text ou texte intégral).

## **OCR ou reconnaissance optique des caractères**

Avant de choisir l'OCR<sup>6</sup> (ou ROC en français) comme technique d'acquisition des index, il convient d'étudier minutieusement la typologie des documents, leur qualité, le contenu des zones que l'on veut reconnaître, les techniques de contrôle et de correction que l'on peut lui associer et enfin la volumétrie. Ce n'est qu'à l'aide de ces éléments que l'on pourra décider de l'opportunité de l'OCR et du produit à retenir. La reconnaissance optique de caractères (appliquée à la gestion électronique de documents est à considérer avec beaucoup de prudence. En effet, les moteurs d'OCR qui existent sont plus ou moins performants en fonction des cas à traiter.

L'OCR est utilisée dans la GED pour deux types d'application. La première, qui consiste à numériser des pages entières, cherche à les transformer en texte et à indexer l'ensemble en "full-text" avec des dictionnaires de correction orthographique. La seconde application fait appel à l'OCR pour reconnaître le contenu de certaines zones afin d'utiliser ce contenu comme index.

Cela fonctionne très bien quand les lots de documents numérisés sont homogènes dans leur format, que les documents sont bien imprimés, offrent un bon contraste et ne sont pas raturés (annotation manuscrite ajoutée, trait d'un cadre se superposant à une zone...). Dans la pratique, les lots de documents sont de formats hétérogènes, pas toujours bien imprimés, etc.

Les techniques à mettre en oeuvre pour remédier à ces inconvénients (algorithmes de redressement d'image, de suppression de fond de page, de filtre des couleurs de fond...) existent mais sont coûteuses et donnent rarement un résultat parfait. Le choix des moteurs d'OCR dépend aussi de l'aspect de documents. Par exemple, on ne choisira pas le même moteur pour une impression laser que pour une impression matricielle. Les moteurs logiciels sont à privilégier par rapport aux machines propriétaires et dédiées. En fait, il n'y a pas de solution "prêt à porter" pour l'indexation par OCR et il faut à chaque fois étudier avec précision les documents, faire des tests sur des lots de documents réels avec des volumes significatifs, mettre en place des programmes de vérification, de correction, de

---

<sup>6</sup> Logiciel de reconnaissance de caractères permettant la traduction de groupes de points d'une image numérisée en caractères (avec l'enrichissement typographique associé) exploitables par des programmes informatiques.

rejet et éventuellement étudier la conception du document (si cela st possible, par exemple un bordereau d'expédition).

L'OCR fonctionne aussi pour le manuscrit avec des contraintes encore plus grandes. Cependant, avec des documents bien conçus, c'est-à-dire en prévoyant des champs pré-casés, des zones redondantes pour les contrôles, des zones calculées par exemple en colonne et surtout avec un programme de correction des anomalies bien étudié du point de vue ergonomique, on peut arriver à de bons résultats. Néanmoins, dans tous les cas d'indexation automatique par OCR, il n'est guère raisonnable de s'attendre à un taux de reconnaissance supérieur à 80% (la réalité est même plus proche de 60%).

### **Le code à barres pour l'indexation automatique de documents numérisés**

Par rapport à l'OCR, l'utilisation du code à barres offre de nombreux avantages. Premièrement, la qualité d'impression des documents peut être médiocre sans que cela altère l'acuité du système. Les taux de reconnaissance de codes à barres sont proches de 100%. De plus, la position du code à barres dans le document n'a pas d'importance. De même, il peut être légèrement incliné sans que cela ait de conséquence. Par contre, son utilisation est plus limitée. En effet, la technique du code à barres est généralement applicable à des documents produits dans l'entreprise et qui y reviennent après un circuit chez des clients ou des fournisseurs mais plus difficilement à des documents produits à l'extérieur, dont on n'a pas le contrôle.

Avec le code à barres, l'indexation du document est partielle et il convient d'enrichir celle-ci a posteriori par programme, la valeur du code à barres servant de lien entre le document et des informations détenues par l'entreprise. Néanmoins, il n'est pas utopique de penser automatiser pratiquement à 100% une chaîne de numérisation/indexation/stockage de documents en utilisant ce principe.

### ***Les techniques et méthodes d'indexation***

On a décrit ci avant les techniques d'acquisition des documents d'une part et des index d'autre part. L'indexation des documents est l'opération qui consiste à décrire les documents en vue de leur exploitation ultérieure. La description d'un document va porter sur deux plans distincts mais complémentaires:

- une description externe contenant des informations sur le type de document, son origine, la date de sa prise en charge ou de sa création, pour les activités administratives ou techniques, le rattachement aux objets de base de l'entreprise (client, fournisseur, produit, etc., ...),
- une description du contenu; les enjeux de l'indexation et les difficultés se situent à ce niveau.

La nature numérique du document et son mode de représentation dans le système de GED font perdre à son contenu toute possibilité d'être compris et exploité directement: une image numérisée constituée de points n'offre aucune signification de contenu. L'enjeu de l'indexation est donc, compte tenu de l'exploitation qui sera faite ultérieurement des documents, d'en exprimer le contenu sous la forme de descripteurs.

Le descripteur d'un document sera attaché au document par un lien de localisation physique; Le descripteur contiendra les deux niveaux de description cités ci dessus. Pour le second niveau, la technique de description de contenu est l'indexation; elle consiste à décrire le document avec des mots ou des expressions convenus préalablement.

Plusieurs techniques d'indexation des documents numérisés

On met en oeuvre différentes techniques et outils pour effectuer l'indexation. L'indexation peut être manuelle et c'est alors l'utilisateur qui saisit les index au vu du contenu du document; à l'opposé, on peut mettre en oeuvre des procédés automatiques effectuant le traitement direct du contenu du document (par exemple la technique d'indexation <<full-text>>). Les principaux outils et techniques sont les suivants:

- listes de mots clés ou dictionnaires pour contrôler et sélectionner les index,
- indexation statistique qui consiste à supprimer les mots vides de sens et à ne retenir comme mots clés que ceux qui dépassent un seuil de fréquence d'apparition,
- indexation "full-text" ou texte intégral associée à des dictionnaires ou des thesaurus,
- analyse linguistique.

La finalité des techniques d'indexation ou de représentation de contenu est de constituer une base de données des différents éléments de description des documents permettant d'établir les liens entre descripteurs et documents. C'est à partir de cette base de données que se feront les recherches et les sélections de documents.

Selon la finalité du système de GED, la nature et le contenu des documents traités, on mettra en oeuvre des schémas différents d'organisation de données et de bases de données associées:

- des fichiers séquentiels indexés,
- des bases de données hiérarchiques ou relationnelles,
- des fichiers inversés.

### ***Stockage et conservation des documents numérisés dans un système GED***

La conservation est un autre élément essentiel d'un système de GED.

Elle aussi doit être étudiée avec minutie avant que ne soit défini le type de périphérique de stockage (disque magnétique, disque optique réinscriptible ou non réinscriptible, CD-WORM, juke-box) et que le stockage soit hiérarchisé en fonction des types de documents, de leur fréquence de consultation, de l'intérêt à les avoir en ligne, de la durée de la conservation et de l'importance stratégique de l'information.

Chaque type de stockage possède des caractéristiques précises qui le rend plus ou plus apte à satisfaire tel besoin ou tel budget. Il faut souligner au passage que, bien que continuant à évoluer rapidement, la technologie est aujourd'hui parfaitement maîtrisée et les moyens de stockage des systèmes de GED sont désormais parfaitement fiables et performants.

Les périphériques de stockage utilisés pour la GED sont pour une partie communs à d'autres applications informatiques.

Ainsi, il est fréquent d'utiliser des disques magnétiques en premier stockage (RAID ou non) ou de faire des sauvegardes sur bandes streamers ou DAT.

En revanche, certains périphériques, sans être exclusivement réservés à la GED, lui sont plus souvent associés qu'à d'autres applications. Ce sont les périphériques à base de disques optiques numériques ou DON.

### ***Recherche, restitution, diffusion des informations et des documents dans un système GED***

Une des finalités essentielles d'un système de GED est de rechercher des documents, soit pour les restituer à l'utilisateur qui les consultera, soit pour les communiquer à un autre utilisateur ou les diffuser à une ou plusieurs personnes. La qualité et la rapidité de la recherche dépendent de l'indexation choisie et de la qualité de l'acquisition.

La recherche de documents se fait par l'intermédiaire d'un langage qui peut être réduit à sa plus simple expression, par exemple une liste de mots clés combinés ou non, ou elle peut être très complexe et faire appel, par exemple, à des techniques d'interrogation en langage naturel. Le premier objectif de ces langages, à la bonne fin duquel contribue la méthode d'indexation, est la pertinence des recherches qui doit conduire à limiter le "bruit", c'est-à-dire la sélection de documents ne répondant pas exactement à la requête, et le "silence", c'est-à-dire le défaut de sélection de documents disponibles qui répondent à la requête.

Plusieurs techniques existent et sont mises en oeuvre dans les systèmes de GED, soit indépendamment et exclusivement les unes des autres, soit pour certaines d'entre elles de façon combinée. Les principales méthodes de recherche sont:

- la recherche booléenne qui consiste à présenter la requête sous la forme d'une expression booléenne de mots clés appartenant au dictionnaire ou au thesaurus, mots séparés par des opérateurs booléens (et, ou, sauf ...),
- la recherche plein texte ou texte intégral, qui évite l'indexation préalable des documents; ils sont décrits par eux mêmes, c'est-à-dire par la chaîne des caractères qui les constitue. Les recherches booléennes s'appliquent également à ce cas. Cependant, on associe souvent à cette technique des méthodes d'interrogation dites en langage naturel, c'est-à-dire reposant sur une analyse linguistique de la requête de l'utilisateur et conduisant à l'élaboration d'une requête "interne" à partir de laquelle la recherche est réalisée,
- recherche par navigation ou technique hypertextuelle; cette technique suppose une structuration préalable des documents avec la mise en oeuvre de liens hypertextuels.

Une fois sélectionnés, les documents peuvent être affichés sur des écrans de consultation, imprimés localement ou à distance sur des imprimantes désignées,

envoyés par réseau vers d'autres postes de travail (fréquemment sous le contrôle d'un outil de workflow ou de groupware), transmis pour traitement à des systèmes informatiques, communiqués par messagerie (par exemple en pièce jointe) ou encore envoyés par télécopie.

### ***Outils d'échange, workflow intégrés aux solutions GED***

Dans de nombreuses situations, la transmission et l'échange de documents entre plusieurs utilisateurs seront facilités ou organisés avec des outils tels qu'un logiciel de workflow ou un outil de groupware (groupe de travail).

Ces situations sont celles où plusieurs personnes travaillent en groupes organisés sur la base de règles et de procédures pour effectuer des travaux d'ordre administratif ou technique: gestion de courrier, traitement des commandes, traitement des sinistres en assurance, instruction des demandes de prêts dans les banques, etc.

Les finalités principales d'un logiciel de workflow sont l'ordonnancement et le suivi des travaux au sein d'unités de travail selon des procédures et des règles préétablies. Ce type de logiciel prend en charge de nombreuses tâches; en particulier, il régule les enchaînements d'opérations, assure la circulation des dossiers à traiter sur les différents postes de travail, surveille les priorités de travaux, gère les délais, assure les synchronisations et déclenche les alertes. Pour toutes ces actions, on établit des liens directs entre les logiciels de workflow et les documents électroniques : ces derniers sont pris en charge par le logiciel de workflow qui, en fonction de règles définies, en assure la distribution sur les postes de travail dans des corbeilles individuelles ou collectives, et prend en charge l'ensemble des déplacements tout au long de la chaîne de travail.

La mise en oeuvre de solution de workflow conduit naturellement à une forte intégration du système de GED au système d'information de l'entreprise. En effet, pour faciliter l'exécution des travaux sur les postes de travail à l'arrivée des documents, on établit des enchaînements automatiques permettant l'accès direct aux applications informatiques, aux outils de production bureautique; en retour, on facilite la recherche de dossiers électroniques à partir des bases de données de gestion. On est ainsi conduit à définir et à réaliser des postes de travail sur lesquels les différents outils sont intégrés en fonction du métier de l'utilisateur;

l'objectif recherché est l'efficacité de l'utilisateur et la banalisation de l'emploi des différentes techniques, GED en particulier.

Les liens entre documents et outils de groupware sont de même nature, avec comme idée directrice la mise en commun d'un ensemble de dossiers et de documents.

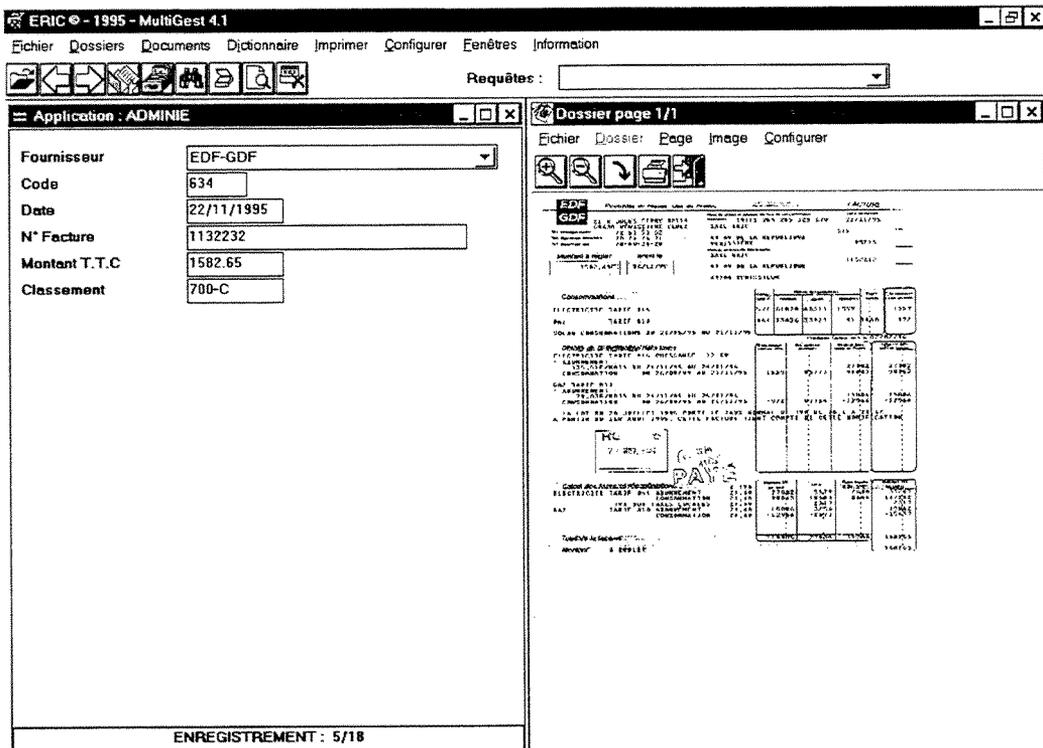
Une mention particulière doit être faite des outils de communication tels que les messageries : elles constituent fréquemment le cadre dans lequel les utilisateurs effectuent entre eux des communications non structurées de documents.

***L'application de Gestion Electronique de Documents: "MultiGest" ©***

Le but n'est pas de présenter une application de Gestion Electronique de Documents, mais à partir d'un exemple de déterminer les nouveaux besoins des utilisateurs.

Depuis deux ans, je collabore au développement de l'application de Gestion Electronique de Documents "MultiGest" ©.

Cette application est développée par la société ERIC spécialisée dans l'archivage et la GED. Depuis 1987, la société ERIC c'est engagée dans la recherche et le développement dans le domaine de la GED.



"MultiGest" © est une application qui peut s'intégrer dans les cinq grandes familles de la GED, c'est-à-dire que cette application permet de gérer l'ensemble des documents gérés par ces cinq familles.

"MultiGest" © est à la fois un système de GED structuré et non structuré c'est-à-dire que les documents sont référencés à l'aide d'une fiche d'indexation multicritères et les recherches se font directement sur le contenu du document.

"MultiGest" © est composé de huit modules qui sont les suivants:

- le module de numérisation,
- le module d'OCR,
- le module d'ICR,
- le module Codes barres,
- le module visualisation,
- le module bureautique,
- le module COLD,
- le module de recherche (requêtes, full text, hypertext).

L'intégration de ces technologies fût indispensable pour que "MultiGest" © reste compétitif face à la concurrence.

Actuellement la société ERIC se pose la problématique suivante: cette intégration rapide de tous ces outils n'a pas été faite pour l'utilisateur, mais ont été intégrés pour répondre à une innovation technologique et non pour répondre aux exigences particulières de l'utilisateur.

Comme pour les systèmes dits d'aide à la décision, l'avènement des nouvelles technologies a modifié les exigences des utilisateurs des systèmes de Gestion Electronique de Documents. L'adéquation de l'intégration de ces technologies et des exigences des utilisateurs nous ouvre une nouvelle orientation et perspective de travail.

Il est temps, maintenant, de mettre à jour la liste non exhaustive de ces technologies qui permettent de répondre aux exigences des utilisateurs.

## **Avènement des nouvelles technologies**

L'époque où les informaticiens régnaient sans partage est révolue. Les pouvoirs sont désormais aux mains des utilisateurs: dans l'entreprise, ce sont eux qui transforment leurs méthodes de travail et enrichissent leur savoir-faire.

De nombreuses entreprises ont longtemps accumulé d'énormes volumes de données sans jamais les exploiter pleinement. Les dirigeants ont enfin pris conscience de la valeur de ces mines d'informations.

Enfin, la communication devient un élément essentiel dans les systèmes d'information d'entreprise. Les réseaux locaux se ramifient à l'intérieur, pour connecter l'ensemble des acteurs de l'entreprise à ses partenaires. "Faites circuler l'information !": Messagerie, outils de "workflow", systèmes de gestion électronique de documents, etc. Fil rouge de toutes ces applications: la communication. Elle est devenue vitale pour les entreprises.

Les entreprises commencent peu à peu à employer la puissance des machines pour mettre leurs applications en image et en musique. Les innovations sont rares. Mais les savoir-faire gagnent du terrain. La traduction visuelle de certaines données complexes n'est d'ailleurs plus un luxe.

L'introduction, dans les entreprises et les organisations de méthodes et d'outils de modélisation des connaissances ouvre des perspectives nouvelles qui n'étaient pas envisageables lors de la première diffusion de la technologie des systèmes à base de connaissances.

Précisons que nous allons nous intéresser aux nouvelles technologies concernant les évolutions des systèmes d'information et de communication concernant les points suivants:

- La banalisation ou démocratisation de l'outil informatique,
- Le développement d'interface homme-machine facilitant l'accès aux informations,
- L'accélération des transferts de données,
- La diminution des coûts de communication.

### *Intranet/Internet*

L'utilisateur a toujours tendance à s'approprier les évolutions techniques majeures. Ce qui s'est produit hier avec le micro-ordinateur recommence avec l'Internet.

Internet est un ensemble de réseaux disséminés sur les cinq continents et dont certains services sont accessibles librement. Internet représente également une communauté d'utilisateurs avec lesquels il est possible de dialoguer ou d'échanger du courrier électronique.

Lorsque l'on se connecte à Internet, on accède à des centaines de milliers de gigaoctets de données en ligne. Celles-ci sont disponibles sous forme de textes, de bases de données, d'images, de vidéos, de sons. Point fort d'Internet, ces informations numérisées traitent de tous les sujets; pratiquement tout existe quelque part.

Pour les entreprises des réseaux entiers peuvent être reliés au service Internet de messagerie. En outre, le réseau peut être utilisé par une entreprise comme outil de communication de l'information.

Le développement explosif d'Internet ne doit cependant pas cacher le fait qu'il s'agit d'un système qui se cherche. L'embryon de la communication universelle est là ; il reste à lui apporter les soins et les règles nécessaires à une croissance qui, sans cela, risquerait de devenir incontrôlée.

Utiliser au sein de l'entreprise les technologies qui ont fait le succès d'Internet, c'est possible. Mieux encore, c'est déjà une réalité qui porte un nom: Intranet. Ce réseau d'un type nouveau facilite le déploiement en entreprise d'applications de communication, de partages de documents, et de travail en groupe.

Derrière l'Intranet se cachent différentes technologies disposant toutes d'un point commun: le respect des standards ouverts nés d'Internet, que soit HTTP, et HTML. Autre intérêt, et non des moindres, ces technologies peuvent être vues comme des briques que les responsables informatique d'entreprise utilisera

suivant le schéma qui lui semble le meilleur. Seul élément nécessaire: la présence dans l'entreprise d'un réseau exploitant le protocole TCP/IP<sup>7</sup>.

Le déploiement d'un Intranet commence souvent par la mise en place d'un serveur Web sur le réseau interne, dit aussi serveur HTTP (Hyper Text Transmission Protocol). Les serveurs Web deviennent au fil du temps de véritables passerelles avec le système d'information de l'entreprise. Ainsi, l'utilisation de scripts CGI (Common Gateway Interface) permet d'ouvrir ce serveur à des applicatifs ou des bases de données. A la clé, par exemple, la génération de pages HTML à la volée (pages dynamiques) obtenues à la suite d'une requête par le navigateur client sur le système de gestion de base de données.

L'architecture Intranet reprend les éléments ayant fait le succès du client-serveur, avec une partie de l'applicatif sur le poste client, l'autre sur un serveur Web, lui-même lié au reste du système d'information.

Ce qui fait le premier intérêt d'un Intranet pour l'entreprise, c'est évidemment le respect de ces standards ouverts.

Facilitant la communication interne, il aura effet de décloisonner les différents départements ou services.

Enfin, en exploitant l'infrastructure existante du réseau public Internet, cet Intranet pourra facilement être étendu et devenir ainsi accessible aux itinérants de l'entreprise, voire à ses fournisseurs ou à ses partenaires. Il permet alors de concrétiser le concept d'entreprise étendue.

La première motivation d'un Intranet est souvent d'ordre économique. Ce type d'architecture diminue considérablement les coûts de déploiement d'applications et d'administration. L'utilisation d'Internet dans le cadre d'Intranet étendus à l'extérieur limite les coûts de communication. Enfin, l'indépendance, vis à vis des fournisseurs préserve l'utilisateur des évolutions tarifaires non contrôlées.

L'intranet, avec les notions de forums de discussion et de messageries, facilite le développement de la communication interne et externe de l'entreprise.

---

<sup>7</sup> Protocole de transport d'informations sur un réseau d'ordinateurs, développé à l'origine dans le cadre du réseau Internet sur machines UNIX. Ce protocole est de plus en plus utilisé quels que soient les systèmes d'exploitations (NetWare, Windows NT, etc...). De nombreux logiciels client/serveur sont basés sur ce protocole.

Il permet la mise en place en entreprise de services nouveaux, trop chers à déployer auparavant. C'est le cas notamment des Intranet documentaires, alternatives aux solutions de gestion électronique de document (GED) pour la diffusion ou la mise à disposition de grosses masses de documents aux utilisateurs.

### ***Télécommunications - la révolution numérique***

Bien que des réseaux réservés à la transmission de données se soient développés parallèlement, on assiste à une imbrication et à une interconnexion croissante de ceux-ci avec le réseau téléphonique. Ce qui offre des possibilités beaucoup plus larges de communication entre des systèmes informatiques de taille moyenne ou petite.

Jusqu'en 1985, le réseau de télécommunication répondait pour l'essentiel aux besoins d'échanges téléphoniques et avait été constitué en conséquence. Fondé sur un principe différent, un R.N.I.S. tend à constituer un réseau d'infrastructure tirant le meilleur parti des équipements de transmission existants (lignes terminales notamment) et de commutation (électronique numérique).

Sous réserve de disposer des équipements terminaux adéquats et de se satisfaire de débits numériques limités à 64000 bits par seconde, un R.N.I.S. se prête à une infinité d'utilisations possibles pour un coût d'utilisation proche de celui du téléphone. Ce réseau permet donc à lui seul, moyennant l'existence de terminaux adaptés, des modes de communication diversifiés. Il est ainsi apte à constituer des liaisons informatiques entre ordinateurs, liaisons qui sont commutables.

### **Les différents modes de communication et les usages**

La possibilité d'une communication numérique à haut débit de bout en bout permet d'autres types d'échanges qui nécessitaient auparavant des moyens beaucoup plus coûteux (lignes spécialisées numériques):

- acheminement instantané de gros volumes de données informatiques (quelques millions de caractères alphanumériques);
- télécopie à haute définition, de grande qualité et de grande rapidité;
- transmission d'images photographiques préalablement codées numériquement;
- échanges d'images animées (vidéo) de qualité sommaire;

- échanges brefs, en mode transactionnel, de données informatiques;
- communications téléphoniques d'une qualité proche de la haute fidélité si l'on dispose d'appareils terminaux spécialement conçus;
- échanges "multimédias" qui font intervenir simultanément le son, les données, les textes, la télécopie, les images fixes ou animées (une des possibilités étant la communication visiophonique).

Le R.N.I.S. s'oriente donc plutôt vers les nouveaux modes de communication. Les plus utilisés sont les échanges d'images, la consultation à distance de documentations à partir de base de données télématiques, les échanges transactionnels de monétique, les réunions à distance, la télécopie, l'interconnexion d'ordinateurs et les échanges de données informatiques. Dans ce type d'utilisation, les terminaux sont reliés directement aux terminaisons numériques de réseau.

### ***Groupware***

Les outils informatiques du travail collectif, groupware et workflow, donnent un nouvel essor au management par objectif et par groupe de projet. Associé au workflow et à la gestion électronique de document (GED), le groupware définit une nouvelle race d'outils informatiques qui intègre l'EDI (échange de données informatisé) et la messagerie électronique.

Le groupware (le travail de groupe via les réseaux informatiques et les messageries) devient une habitude de travail qui n'est pas réservée aux ingénieurs des bureaux d'études ou aux chercheurs. En se démocratisant, les outils de groupware deviennent de plus en plus complets et faciles d'utilisation.

Les messageries d'entreprise sont maintenant capables de transmettre et stocker des documents attachés aux messages, de les filtrer ou de les émettre via des agents programmés.

Le groupware dans toutes ses nouvelles formes rétablit la continuité entre la bureautique et l'informatique. Il les renvoie à leur vocation principale : faciliter autant le travail individuel que collectif. Paradoxalement, tous ces développements technologiques réutilisent les notions traditionnelles de l'entreprise. Le vocabulaire en témoigne: des notions longtemps à la mode d'algorithme, d'entité et de relation, on revient aux concepts plus classiques de

document, de dossier ou de corbeille. Les objectifs des nouveaux outils suivent le même mouvement: de l'efficacité par la centralisation des moyens, on revient à la souplesse et à la visibilité. Les références en matière d'organisation, après avoir été la spécialisation fonctionnelle et la hiérarchie, retournent aux notions primordiales que sont le client et la réactivité d'entreprise.

Ces solutions qui utilisent la circulation informatisée des messages et des documents modifient l'architecture des applications informatiques. Elles conduisent en particulier à ouvrir de nouveaux horizons dans l'organisation de partenariats entre entreprise.

Les produits de GED se différencient en fonction des médias et des formats de documents supportés (CD-ROM, SGML, Internet ...) ou des usages (édition, documentation technique et autres affichages de dossier administratifs). C'est pourquoi les capacités d'intégration du workflow sont de plus en plus utiles. La structure en workflow permet une modernisation des moyens de production sans remettre en cause l'existant. Permettre la fédération de différents systèmes applicatifs, qu'ils soient de gestion, techniques ou bureautiques est une capacité essentielle des produits qui se juge par l'environnement de développement offert et par le moteur d'exécution des procédures. Les possibilités de description du contexte (références aux documents gérés par des systèmes de GED, SGDT, URL, etc.) sont importantes. Mais ce sont les capacités d'activation des ressources externes (DDE, OLE, ActiveX, JAVA, etc.) qui, à ce niveau, devraient jouer le rôle primordial.

### ***Data Warehouse***

L'informatique décisionnelle a bien changé de look. Néanmoins, le data warehouse, inventé par l'américain Bill INMON, n'a réellement réussi sa percée médiatique qu'il y a deux ans à peine. Sorti de l'époque pionnière, il est devenu une ingénierie nouvelle vague des systèmes décisionnels. Il a bénéficié pour cela d'une triple évolution: l'explosion de l'offre, l'émergence de nombreux acteurs et la publication de retours d'expérience plus que positifs.

Qu'offre en effet le data warehouse, si ce n'est la promesse de fournir rapidement et de façon fiable des informations utiles à la prise de décision, et ceci grâce au traitement de données jusqu'alors sous-exploités ? Reste que si le jeu en vaut la

chandelle, les difficultés ne sont pas négligeables. Car, tant en conception qu'en exploitation, les façons de procéder s'écarteront sensiblement de celles acquises au contact de l'informatique de gestion.

Quel que soit son secteur d'application, le data warehouse doit se concevoir comme une pyramide à quatre étages. En soubassement se trouvent les sources des données brutes, de production ou archivées, à partir desquelles seront extraites, filtrées et agrégées les informations qui constitueront la base du data warehouse. Celle-ci pourra être exploitée depuis un outil analytique ou encore statistique, lui même interfaçable avec des outils de visualisation graphique et de reporting. De nombreuses variantes sont possibles, mais le problème fondamental se situe moins au niveau du choix des différentes catégories d'outils que de la correcte conception de l'application.

Il faut constituer des data warehouse très ciblés, faire un constant aller et retour entre les besoins et le modèle choisi et ne penser à étendre le champ d'application qu'une fois prouvée l'efficacité du premier prototype.

Les problèmes les plus épineux: choisir les jeux de données les plus pertinents, établir les règles d'agrégation des données (méta-modèle), choisir les bons outils d'extraction mais aussi penser à optimiser la future application. Ce sont autant d'équilibres à respecter au niveau du choix des composants matériels et logiciels. Ce sont la prise en compte de méthodes d'indexation dernier cri proposées, l'adoption d'architecture 64 bits ou encore l'établissement des modèles de bases de données dénormalisées.

Avec de telles caractéristiques, l'exploitation et l'administration des applications décisionnelles s'écarteront sensiblement des traditionnelles missions imparties au "DBA" (Date Base Administrator). Il sera primordial pour l'équipe d'exploitation de faire vivre le data warehouse et, par voie de conséquence, d'établir à quelle fréquence les données seront renouvelées, nettoyées, homogénéisées .... Il sera indispensable également d'administrer les processus automatisant l'enchaînement de toutes ces tâches.

### ***Multimédia***

Dans son acception la plus générale, le multimédia permet d'accéder à n'importe quel type d'information (écrite, sonore, visuelle), à n'importe quel moment et sur

n'importe quel support.

Dans les faits, sa définition est plus restreinte. Il s'agit de la réunion sur un même support de fichiers contenant du texte, du son, des images fixes ou animées et organisés au moyen d'une programmation informatique. Une autre caractéristique fondamentale est l'interactivité, qui qualifie les matériels, les programmes ou les conditions d'exploitation permettant des actions réciproques en mode dialogué avec des utilisateurs ou en temps réel avec des appareils.

Le support privilégié du multimédia est le CD-ROM lu sur un micro-ordinateur adapté, dit multimédia; mais l'information peut également être véhiculée sur un réseau informatique numérisé à très hautes capacités et vitesse de transmission.

Association des données, du son et de l'image, le multimédia a pour vocation d'entrer dans un nombre de plus en plus grand d'applications. Des usages simples et économiques comme la visioconférence lui donnent une vraie opportunité de développement.

La démocratisation du multimédia est la conséquence de deux évolutions concomitantes. La première est la facilité d'utilisation de l'informatique, associée à une puissance accrue des postes de travail qui autorise la manipulation de données complexes. Elle explique le déploiement du multimédia dans l'entreprise vers des applications autres que la formation, l'information ou le commercial sur lesquelles il se cantonnait depuis d'années. La deuxième évolution est l'industrialisation des technologies multimédia elles mêmes. L'arrivée massive de lecteurs de CD-ROM, de contrôleurs graphiques performants et de cartes son permet leurs utilisations faciles sur les micro-ordinateurs de bureau ou les portables.

Ces PC disposent de plus en plus souvent en standard des fonctions multimédias de base. Et ce n'est qu'un début: dans peu de temps, n'importe quel poste bureautique standard sera muni d'extensions vidéo et sonores. Pourtant, l'arrivée d'outils de reconnaissance de la parole, dont IBM avec sa dictée personnelle à longterm été le fer de lance, devrait encore faire évoluer l'ensemble des applications vers une communication de plus en plus naturelle en associant aux données l'image et le son. Une simplicité à la portée de tous sera l'étape qui permettra à la reconnaissance de la parole de franchir les portes de bon nombre d'entreprises.

Le multimédia n'est pas un marché, ni une fin en soi. Il doit plutôt être vu comme un ensemble de techniques qui offre de nouvelles opportunités au système d'information. La raison d'être de ces outils se révèle dès que les entreprises ont besoin d'utiliser toutes les sortes de communication. Mais pour répondre aux besoins réels des utilisateurs, on ne trouve pas le niveau de technologie suffisant pour réaliser tout ce que l'on attend. Il faut donc apprendre à vivre avec les contraintes technologiques. Le véritable challenge est d'arriver à mettre en adéquation les technologies du moment et le niveau d'acceptation de l'utilisateur dans sa fonction de lecteur/acteur.

### ***EDI: Ingénierie collaborative***

L'ingénierie collaborative est l'ensemble des outils et méthodes à destination des ingénieurs, qui permettent de travailler ensemble et/ou de manière simultanée. Ce sont donc des outils mais aussi des méthodes qui permettent d'augmenter et de favoriser le travail en équipe en facilitant la communication entre les individus.

Avec le développement de l'informatique et surtout des moyens de communication internes et externes à l'entreprise, ces notions de travail collaboratif et concourant reviennent en force sur le devant de la scène.

En effet avec le développement des réseaux informatiques internes mais aussi avec le développement d'Internet et d'Intranet, le travail collaboratif devient possible quelque soit la localisation géographique des individus.

Les possibilités sont immenses. L'EDI qui consiste en l'échange de données informatique est un premier exemple d'ingénierie collaborative.

Le temps est devenu une donnée stratégique pour les entreprises. Le délai est une composante vitale de la qualité et notamment le délai de conception de nouveaux produits pour répondre aux besoins rapide du marché.

### **Contexte de l'ingénierie collaborative**

Le papier reste un vecteur de communication important dans les entreprises. L'informatique, et le réseau de communication d'entreprise ou inter-entreprise, sont en train de changer, voire révolutionner certaines méthodes de travail, d'organisation.

Le niveau le plus simple de ce type d'échange informatique est l'échange de fichier. Se pose dès lors l'un des premiers besoins des personnes qui échangent: le format du fichier.

**Le premier besoin: des échanges simples de données.**

Vient ensuite le problème de l'échange physique des données. C'est donc la nécessité d'utiliser le réseau pour l'échange des données. Les ordinateurs sont généralement interconnectés, par contre l'échange avec l'extérieur peut poser problème. Comment transférer rapidement et à bon prix des données tout en assurant leur sécurité et de la qualité du transfert?

Des systèmes d'EDI sont déjà utilisés par beaucoup d'entreprises. Il s'agit généralement de systèmes assez figés et assez coûteux en terme d'infrastructure. La transmission se fait généralement de manière synchrone, à l'heure fixe et est généralement reçue de manière synchrone par le correspondant. Ils ne répondent donc pas au besoin de joindre n'importe qui, n'importe quand.

La messagerie électronique permet l'envoi de fichier de manière asynchrone à l'intérieur et l'extérieur de l'entreprise.

**Le deuxième besoin: une communication souple et rapide de manière asynchrone.**

Puis, lorsque l'infrastructure réseau, et les problèmes de compatibilité de fichier sont résolus, chaque document a un cycle de vie et il est important de garder la trace du document, savoir si les personnes qui devaient le voir l'ont vu, si les modifications sont bien approuvées et par qui, si on a gardé trace de toutes les évolutions, etc. Il faut donc un véritable système de suivi et de gestion de l'information qui assure traçabilité et sécurité.

**Le troisième besoin: le stockage et la hiérarchisation de l'information et de contrôle de son flux.**

Enfin, avec le développement de l'entreprise étendue, se pose le problème de la communication informelle mais ô combien importante, qui est la communication orale et visuelle.

Comment être sûr que l'interlocuteur de l'autre côté de la terre comprends bien le message qu'il a reçu? Il faut pouvoir pour cela communiquer oralement.

**Le quatrième besoin: la téléconférence et la visioconférence.**

### ***Conclusion***

Cette liste va nous permettre maintenant de mettre en avant les différentes orientations et perspectives de recherche sur les domaines que nous avons choisi: les systèmes dits d'aide à la décision et les systèmes de Gestion Electronique de Documents. Ces orientations et perspectives concernant plus particulièrement les exigences et les besoins des utilisateurs.

## **Impact des nouvelles technologies sur les Systèmes dits d'aide à la décision et sur le systèmes de gestion électronique de documents**

L'explosion de la première phase du développement de l'informatique fait place, à présent, à une phase d'optimisation des différents potentiels de cet outil, par la conception d'architectures cohérentes. Les entreprises ressentent ainsi, de plus en plus, la nécessité de structurer leur développement informatique. La cohérence de celui-ci dépend tout particulièrement des éléments suivants:

- le traitement et la circulation de l'information,
- la productivité de l'outil informatique,
- les méthodes et les principes d'organisation de cet outil,
- l'intégration de cet outil dans l'organisation de l'entreprise.

La simplification, la rationalisation des traitements centralisés et décentralisés, permettant une gestion et une distribution efficaces de l'information, font du système d'information et de communication un outil majeur dans l'entreprise.

Le document véhicule l'emblématique de l'entreprise, il est support de représentation de ses activités. Un facteur important de la pérennité des sociétés modernes est la reconnaissance du document comme constitutif de la mémoire de l'entreprise, contenant les traces de l'histoire et du projet des hommes et des femmes qui la font vivre.

Améliorer la qualité du document, la justesse de l'information qu'il contient, sa circulation entre les différentes fonctions de l'entreprise et son accessibilité, c'est renforcer cette cohésion des acteurs de l'entreprise vers une meilleure productivité.

Un système d'information et de communication allie la notion de document comme support de l'information à celle d'application informatique au service de l'entreprise, application informatique dont les axes se définissent de la manière suivante:

- création, structuration et stockage de l'information en vue de la constitution d'un patrimoine documentaire,
- mise à disposition de ce patrimoine pour les besoins propres de l'entreprise.

Le système d'information et de communication devient donc un outil stratégique au service de l'organisation, de la pérennité et de la compétitivité de l'entreprise.

Actuellement, un système dit d'aide à la décision est composé d'un ensemble de systèmes d'information hétérogènes et essaie de les connecter. Une nouvelle approche a été proposée mettant en avant un système décrivant un domaine d'application et construit dès le départ en tant que tel. Plusieurs travaux s'inscrivent dans ce contexte, comme l'architecture d'un Système Interactif d'Aide à la Décision Avancé proposé par Laïd BOUZIDI<sup>8</sup>.

En ce qui concerne les systèmes de Gestion Electronique de Documents, comment l'intégration de ces nouvelles technologies peuvent répondre aux exigences des utilisateurs?

### ***Système Interactif d'Aide à la Décision "avancé"***

Le processus décisionnel est rarement appréhendé comme une difficulté. Seule la décision peut paraître délicate, faute d'informations disponibles. L'ensemble des nouvelles technologies présentées dans le chapitre précédent permettent cette disponibilité d'une masse importante d'information concernant la prise de décision.

Cette architecture présentée ultérieurement se compose de trois sous-systèmes permettant d'intégrer les connaissances et l'expertise du décideur dans la prise de décision. Il sera aussi présenté la nécessité de prendre en compte les connaissances de l'utilisateur dans la mise en place d'un SIAD de type SIAD avancé.

### **Présentation**

L'architecture d'un SIAD avancé intègre les connaissances et l'expertise du décideur dans la prise de décision: l'approche SIAD "avancé".

Une prise de décision est fondée sur une logique de situation (LEVINE et POMEROL 1989). Celle-ci peut être fondée sur l'exploitation de deux types de données ou informations:

- les informations statiques caractérisant de façon descriptive l'activité d'une organisation et de son environnement; on y applique différents modèles. Elles sont représentées analytiquement par la base de données et la base de modèles.

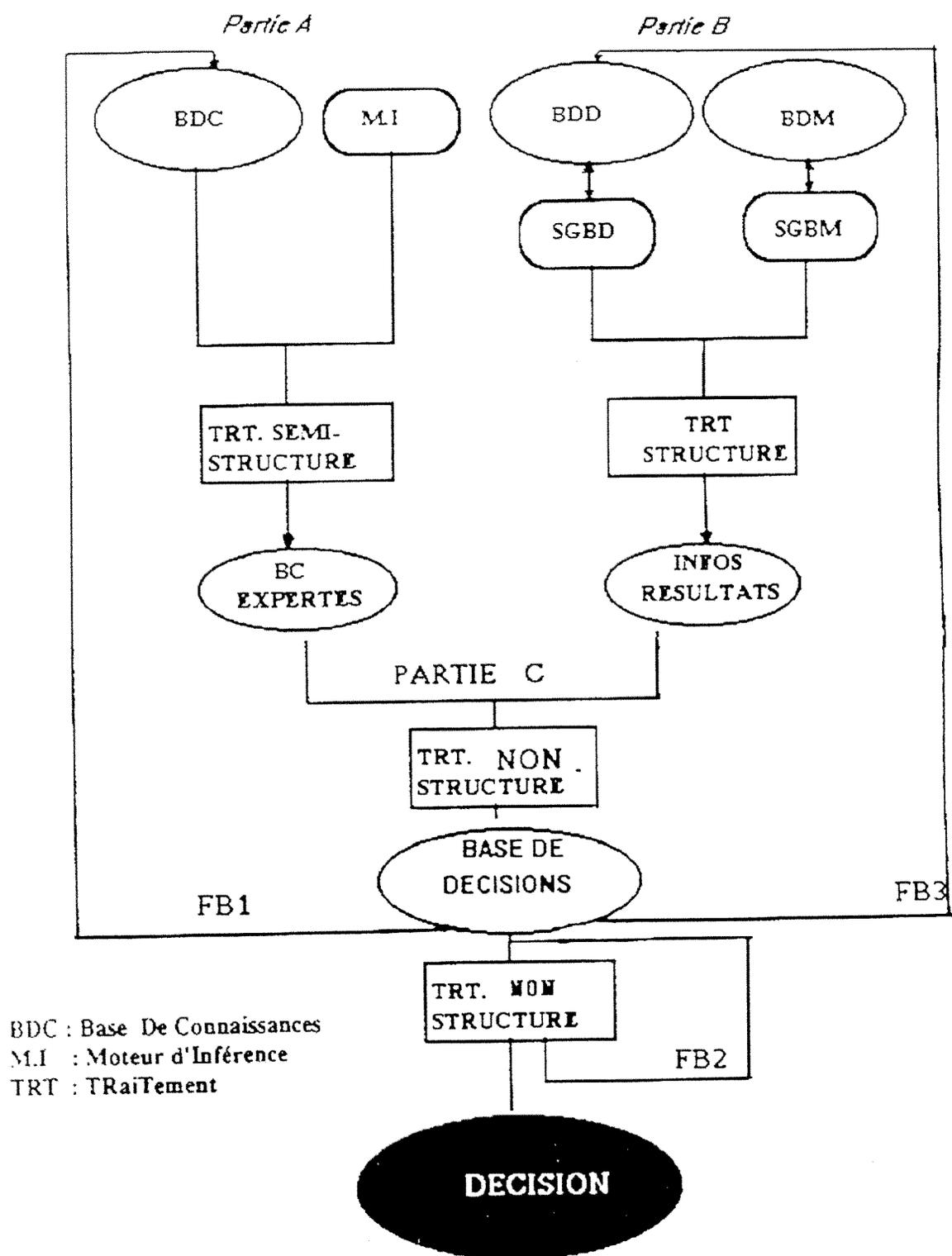
---

<sup>8</sup> Laïd BOUZIDI et Christine SYBORD, 1995, 2<sup>ème</sup> \*Congrès Biennal de l'AFCEC à Toulouse.

C'est ce qui est représenté par la base de données et la base de modèles. Ces informations sont évolutives.

- les informations dynamiques représentent les connaissances liées à l'expérience et au savoir-faire des décideurs. Ce sont des informations qualitatives.

L'architecture du SIAD "avancé" est représentée par le schéma ci-dessous. Elle est composée de trois sous-systèmes, interconnectés par une interface de dialogue et un contrôleur de fonctions.



Le sous-système A représente un système expert, permettant l'élaboration d'une base de connaissances dite "Base Expertise". Les traitements associés sont peu structurés.

Le **sous-système B** représente un système classique de traitement informatique. C'est l'exploitation de la base de données et de la base de modèles par les SGBD et SGBM. Les traitements associés sont totalement structurés.

Le **sous-système C** représente l'aspect purement décisionnel. C'est l'essentiel du niveau heuristique dans un SIAD. Les traitements associés ne sont pas structurés. Ces trois parties contiennent, dans des proportions différents, des connaissances représentant le savoir-faire du décideur.

### **Sous-système A: La modélisation des connaissances de l'utilisateur**

Une bonne connaissance de l'utilisateur est nécessaire pour construire des explications pertinentes ou pour conduire un dialogue explicatif coopératif.

Les connaissances contenues dans un modèle utilisateur sont caractéristiques d'un utilisateur individuel ou d'un groupe d'utilisateurs dont l'utilisateur individuel est un représentant. Certaines de ces connaissances sont relatives au domaine d'expertise: elles portent en particulier sur la tâche experte.

Les connaissances constituant le modèle utilisateur sont initialement acquises comme le seraient les connaissances d'expertise; il faut noter qu'elles sont de plus en plus modélisées à partir d'études empiriques d'utilisateurs réels ou potentiels. Ces connaissances forment alors un modèle utilisateur générique ou stéréotypique. Par la suite, les connaissances relatives aux utilisateurs individuels sont acquises au fur et à mesure de l'interaction avec le système. Cette dernière acquisition se fait soit de manière explicite, soit de manière implicite, par inférence.

Le modèle utilisateur sert à comprendre ce que dit et ce que veut l'utilisateur et à construire des explications ou contrôler l'interaction explicative. Autrement dit, il permet l'adaptation du système à l'utilisateur: adaptation du contenu et de la forme des explications ou du dialogue explicatif au niveau d'expertise, aux buts, aux préférences, etc., de l'utilisateur.

Cette notion de modèle utilisateur concerne le sous-système A de l'architecture SIAD avancé et permet une meilleure adéquation entre la base de connaissances et l'utilisateur d'une part, et d'autre part entre la base de connaissances et la base de décisions.

### **Sous-système B: Base de données**

Un des grands rôles d'un SIAD est la conservation et la mise à disposition des données. Le module de base de données ne se contente pas de stocker passivement les informations. En effet, grâce au système de gestion de base de données, les données participent activement au fonctionnement des SIAD.

Le choix du SGBD et la structuration des informations résultent d'un travail d'analyse. En fonction des modèles qui vont venir se greffer sur la base de données et de l'utilisation qui sera faite des informations, le concepteur décidera du type de SGBD puis l'organisation des enregistrements.

La base de données va alors servir à enregistrer des résultats intermédiaires qui ont une signification dans le cours du traitement ou qui peuvent servir de base à d'autres manipulations.

La base de données du SIAD enregistre la trace des solutions retenues. Les décisions qui, après étude, sont adoptées ou retenues comme possibles, voire les solutions de rechange, sont autant d'informations qu'il faut conserver à des fins d'examen rétrospectif. La base de données est le gisement d'information sur lequel repose le SIAD, ensuite le SGBD intervient en cours de traitement et jusqu'à l'enregistrement des résultats finaux. Les aspects informationnels et décisionnels sont complémentaires.

Après l'ère de l'imprimerie typographique, puis celle de la photocomposition et de l'offset, l'apparition de l'édition électronique marque une étape nouvelle dans l'histoire de l'écrit (et de l'image), en remettant en question la nature du support sur lequel celui-ci est lisible, et en proposant d'autres modes d'approche de l'information, à travers la constitution de banques de données.

D'ores et déjà, on peut dire qu'avec l'utilisation du texte électronique c'est une autre conception du savoir et de sa transmission qui se trouve définie. Par rapport à la masse d'information contenue dans une base de données, l'hypertexte et l'hypermédia dégagent des possibilités de lecture non hiérarchisée totalement inédites: l'hypertexte parce qu'il permet, à partir d'un mot ou d'un nom propre, d'établir une liaison continue entre l'ensemble des textes; l'hypermédia parce qu'il favorise, lui, l'établissement de liaisons suivies entre textes, sons, et images.

Cette notion d'édition électronique couplée au sous-système B permet à

l'utilisateur d'accéder à une formidable masse d'informations facilitant la prise de décision.

Le problème à étudier est la mise en place d'une interface permettant la gestion de ces différentes bases de données comme une seule, sachant qu'elles proviennent de sources différentes.

### **Sous-système C: Base de décisions**

La mise en place de cette base de décisions permet à l'utilisateur de stocker son savoir-faire dans la résolution de problèmes.

La mondialisation des entités économiques montrent que l'ensemble des décideurs d'une entreprise ne sont pas forcément situés dans un même lieu.

L'économie actuelle force les décideurs à prendre très rapidement des décisions vitales pour l'entreprise.

Cette base de décisions permet d'accélérer la prise de décisions. Cette rapidité peut être augmentée par les nouveaux moyens de communication. Ces nouveaux moyens de communication permettent de mettre en commun les différentes bases de décisions, permet de connaître les différents processus de décision de ces collaborateurs. En ajoutant à la notion de mise en commun des bases de décisions, la visioconférence permet à ces décideurs de résoudre un problème en commun et le plus rapidement possible.

Ces bases de décisions peuvent aussi rentrer dans le concept de connaissances actives sur CD-ROM, c'est-à-dire qu'il pourrait être possible de stocker ce savoir-faire d'un domaine particulier sur ce support de stockage, et permettre à décideur d'alimenter sa base de décisions.

### ***Système de Gestion Electronique de Documents***

L'évolution, l'impact des nouvelles technologies sur la GED a suivi la forte progression de l'outil informatique.

Si l'on analyse l'évolution des applications GED, on distingue cinq phases:

- **Phase 1:** le stockage de l'information, le but principal est l'archivage des informations sur un support magnétique.
- **Phase 2:** le traitement de l'information, c'est la recherche d'une information.
- **Phase 3:** la collecte automatique de l'information, c'est l'indexation

automatique des informations.

- **Phase 4:** la communication et la diffusion de l'information, c'est l'émergence des réseaux.
- **Phase 5:** la collecte à distance de l'information, c'est l'explosion d'Internet et d'Intranet.

Toutes ces évolutions correspondent aux exigences des utilisateurs. La démocratisation de l'outil informatique fait que l'utilisateur est de plus en plus exigeant de par sa connaissance accrue dans l'utilisation de cet outil.

Cet intérêt que porte l'ensemble des utilisateurs à cet outil est due à la nécessité de communiquer et de transmettre des informations.

L'utilisateur exige de plus en plus des informations fiables, pertinentes, dynamiques, compréhensives, rapides d'accès, et à faible coût.

Les systèmes de Gestion Electronique de Documents mettent à la disposition de l'utilisateur un ensemble d'informations sous des formats divers. La démocratisation de l'outil informatique que ce soit sur le plan matériel ou sur le plan conceptuel. Cette démocratisation est due au faible coût de l'outil informatique et à l'évolution des capacités de l'utilisateur. L'émergence des Réseaux et plus particulièrement d'Inter-Intranet a orienté les besoins des utilisateurs vers la consultation des informations à distance et l'indexation de ces documents en local pour mettre à jour leur Système d'Information et de communication. Le concept d'indexation et plus particulièrement celui de l'indexation automatique est depuis les années 1980 une piste de recherche largement exploitée. Etymologiquement, *indexer* signifie montrer du doigt quelque chose qu'on veut identifier à telle ou telle fin. A l'époque moderne, on désigne par ce mot l'action d'identifier tel ou tel aspect significatif de *document* quelle qu'en soit la nature de façon que cet aspect ou ces aspects servent de clés quand on aura besoin, plus tard, de le rechercher au sein d'une mémoire.

Notre société a vu se développer l'usage de beaucoup d'autres types de documents, tels que le film sous ses différentes formes, les bandes audio et vidéo, les bandes et disques utilisés comme mémoire dans un ordinateur. Tous ces documents contiennent de l'information, qu'il convient d'étiqueter clairement afin que ceux qui désirent la consulter puissent y accéder aisément.

Assez tôt dans l'histoire de l'automatisation, on s'est rendu compte qu'on pourrait se servir de machines pour indexer. A noter les travaux réalisés par l'ancienne équipe de recherche SYDO sous la direction de Monsieur BOUCH et Monsieur LE GUERN qui c'est attaquée à la faiblesse principale de l'indexation automatique: la sélection de mots isolés. La reconnaissance des syntagmes nominaux et de leur fonction syntaxique permet une recherche beaucoup plus précise.

Le concept de la recherche fait aussi partie des exigences des utilisateurs, c'est-à-dire qu'ils veulent accéder à une information le plus rapidement possible et surtout pouvoir être capable de retrouver n'importe quel document dans l'ensembles des informations de leur Système d'Information et de communication.

Les systèmes de Gestion Electronique de Documents font actuellement de plus en plus appel à la technologie du CD-ROM. Ces systèmes stockent une partie de leurs informations sur ce support amovible qui permet une diffusion plus facile de ces données. L'inconvénient actuel de ce stockage, c'est que ces données sont uniquement une copie conforme des informations. Ces informations peuvent être considérées comme des connaissances passives.

Une question se pose, serait-il possible de rendre ces informations, stockées sur CD-ROM, comme connaissances actives, c'est-à-dire, la possibilité de générer de nouvelles informations propre à l'utilisateur à partir des informations gravées sur le CD-ROM. Cette notion de connaissances actives devrait permettre d'alimenter soit une base de connaissances, soit un SGBD constituant par exemple un système dit d'aide à la décision.

## **Conclusion**

Jusqu'à présent l'informaticien monopolisait les modes d'emploi des autres, les façonnait et les mémorisait à sa façon, d'une manière implicite, par le biais d'algorithmes qu'il lui fallait concevoir à leur place. Le choix d'un modèle comme mode d'approche, tout en renforçant les déterminismes, inhibait la créativité, engendrait des frustrations chez l'utilisateur qui ne se retrouvait que très rarement dans le système qui lui était finalement livré.

Depuis les années 80, l'utilisateur participe de plus en plus dans la création et la mise en place de nouvelles applications qui correspondent à leurs réelles besoins.

Les connaissances se retrouvent entassées dans les armoires, stockées dans les têtes, soigneusement rangées dans les archives où elles peuvent être ignorées très longtemps par les acteurs, jusqu'à l'oubli de leur existence. Ceux-ci préfèrent bien souvent renouveler les études plutôt que de plonger dans la recherche d'une information, accessible certes, mais dont le contexte n'est plus connu en l'absence de ceux qui l'ont élaborée dans leur vécu quotidien.

Les techniques de l'Intelligence Artificielle, parce qu'elles permettent de travailler la connaissance, peuvent faire émerger les stratégies d'acteurs autrefois ignorées.

Les SIAD s'attaquent aux tâches non structurées qui n'ont, pour la plupart, fait jusqu'ici l'objet d'aucun protocole et sont par conséquent non normalisées et a fortiori non informatisées. C'est donc un large domaine qui, grâce aux SIAD, est ouvert à l'informaticien.

Le paradigme de la prise de décision se superpose ainsi au paradigme du traitement de l'information pour définir la cadre d'action des SIAD. Ce traitement de l'information, on le retrouve dans les systèmes de Gestion Electronique de Documents.

La plupart des nouvelles technologies sont déjà intégrées dans les applications qui sont actuellement sur le marché. L'étude de l'intégration de ces technologies dans ces applications de GED est uniquement tournée vers l'innovation.

Il faut, maintenant, tourner toutes ces avancées technologiques vers les besoins des utilisateurs de plus en plus exigeants.

Ce mémoire n'a pas l'ambition de présenter les Systèmes d'Information et de Communication et les nouvelles technologies en général, mais de se focaliser sur

deux domaines particuliers qui sont les Systèmes d'Aide à la Décision et les systèmes de Gestion Electronique de Documents et sur les nouvelles technologies les concernant.

Dans ces domaines, l'exigence des utilisateurs du fait de l'émergence et de la démocratisation de ces nouvelles technologies déterminent de nouvelles orientations et perspectives de recherche.

L'approche des Systèmes Interactifs d'Aide à la Décision Avancés constitue une première piste de recherche au niveau de l'intégration et de l'apport des nouvelles technologies dans les trois modules qui les constituent: la modélisation des connaissances de l'utilisateur, la base de données, et la base de décisions. Ce modèle théorique reste à être validé via un prototypage sur un domaine touchant les Systèmes d'Information et de Communication.

La deuxième piste de recherche est de savoir comment intégrer un système de Gestion Electronique de Documents tel qu'il est actuellement dans un Système d'Information et de Communication en général en tenant compte des besoins des utilisateurs (intégrité, confidentialité et indexation de l'information) et de l'apport des nouvelles technologies (nouveaux supports de stockage, nouveaux moyens de communication)?

A ce propos des perspectives d'indexation automatique sur des images sont développées actuellement, ainsi que le nouveau concept du "TATOO" sur l'image afin d'éviter la copie de l'information disponibles les réseaux et en particulier du réseau Internet.

De plus, la notion de connaissances actives stockée sur un CD-ROM permettant d'alimenter un Système d'Information et de Communication, peut ouvrir des perspectives intéressantes dans la diffusion et la communication d'un savoir-faire concernant un domaine d'application donnée.

Enfin, en amont du développement de ces orientations, il serait peut être opportun d'analyser d'une façon très approfondie de l'orientation des besoins des utilisateurs de tels systèmes dans un domaine d'application et selon le type d'utilisateur.

## **Glossaire**

### **Agrégation**

Première opération de préparation des données d'analyse, elle consiste à agréger les données élémentaires suivant les critères d'analyse retenus lors de la conception du data warehouse. Les niveaux d'agrégation sont multiples et peuvent être mis en œuvre dans une base de données multidimensionnelle.

### **Annotation**

Le plus souvent de nature textuelle, "collée" sur une partie de l'image afin d'y apporter un commentaire spécifique. Les annotations peuvent être de type public (accessibles à tous) ou privé (accessibles uniquement à celui qui les a écrites). Des annotations orales (par enregistrement) peuvent également être associées.

### **Base de données multidimensionnelle**

Elle offre un mode de stockage destiné à optimiser l'accès aux données grâce à la préparation des agrégats. A l'inverse du modèle relationnel, le multidimensionnel prépare les informations calculées à partir des données élémentaires, et cela afin de faciliter les analyses. Son objectif principal: le décisionnel. Inconvénient de cette structure: le nombre de dimensions, théoriquement illimité, augmente le temps de mise à jour de la base et doit être déterminé dès la conception du data warehouse.

### **Base de production**

Elles gèrent les données élémentaires de l'entreprise et sont à l'origine de tous les data warehouse. Elles sont fréquemment réparties par domaine ou par secteur géographique.

### **Business Intelligence**

Cette dénomination désigne l'ensemble des outils décisionnels destinés à l'utilisateur final. De type bureautique, ces outils servent à extraire et à présenter les résultats de requêtes effectuées dans le data warehouse.

### **Datamart**

Sous-ensemble du data warehouse, il rassemble les données élémentaires et agrégées correspondant à un métier particulier. Il peut, par ailleurs, contenir des informations issues d'un autre domaine de l'entreprise afin d'effectuer les analyses. Le datamart constitue souvent le point de départ d'une évolution progressive vers

un data warehouse. La complexité de ce dernier est alors mieux maîtrisée, et les volumes de données gérées sont réduits. La petite structure devient alors projet pilote avant la mise en œuvre effective du data warehouse.

### **Datamining**

Egalement appelé extraction de connaissance, le datamining s'appuie sur le constat qu'il existe des informations cachées dans les gisements de données de l'entreprise. Cet ensemble de techniques permet de mettre au grand jour ces précieuses informations.

### **Bitmap**

Terme informatique utilisé pour désigner la représentation interne sous forme de points d'une image.

### **CCITT (Comité Consultatif International des Téléphones et Télécommunications)**

Organisation de normalisation.

Le CCITT a émis les avis T4 et T6 définissant respectivement les standards de compression d'images monochromes Groupe 3 et Groupe 4. Ces normes sont utilisées par les télécopieurs.

### **CD-R: Compact Disc Recordable**

Voir CD-WORM.

### **CD-ROM: Compact Disc - Read Only Memory**

Très souvent le public non averti fait la confusion entre CD-ROM et DON bien qu'il y ait de grandes différences entre les deux supports. Le diamètre par exemple de 12 cm ou 4,72 pouces pour le CD-ROM et de 13 cm ou 5,25 pouces pour le DON. Le format physique du CD-ROM est en tous points semblable au CD-AUDIO. Il est simple face, et peut contenir 650 Mo en vertu d'un format normalisé (ISO 9660). C'est un support préenregistré en usine qui ne permet que la lecture. On ne peut pas y enregistrer d'informations comme on le fait avec un DON ; c'est la raison pour laquelle sa première vocation a été la diffusion (de logiciels, de catalogues, de bases de données,...). Le support lui-même est d'un coût très faible. Dupliqué en volume important (à partir de 500 copies ), il revient à environ 10 F pièce (hors frais de matrice).

### **CD-WORM: Compact Disc - Write Once Read Many**

Le CD-WORM ou CD-R pour les Anglo-saxons (CD-R pour CD-Recordable) est la révélation de ces deux dernières années en matière de stockage optique. L'unité de CD-WORM est en fait un enregistreur de CD-ROM que l'on peut tout simplement connecter sur un micro-ordinateur ou une station de travail. Son utilisation n'est pas aussi simple et transparente que celle d'un DON traditionnel car on ne peut pas enregistrer de données au fil de l'eau. On est obligé de passer par une phase qui consiste à reconstituer l'image du CD-ROM que l'on veut générer, au format logique ISO-9660, puis à copier cette image en une seule fois (monosession) ou sous formes de sessions ; cette dernière technique est appelée MULTI-SESSIONS. L'automatisation de l'exploitation d'un tel support, bien que réalisable, est moins facile que celle d'un WORM. En revanche, le CD -WORM jouit de qualités indiscutables qui font de lui aujourd'hui, parmi les supports optiques, celui qui semble promis au plus bel avenir.

Le CD-WORM ainsi généré peut être lu sur n'importe quel lecteur de CD-ROM (double vitesse et multisessions au minimum). Le CD-ROM est le seul support optique normalisé sur le plan physique et logique avec la norme ISO 9660 assurant une exploitation des données aussi bien à partir d'un compatible PC, d'un Macintosh, que d'une machine UNIX. Par ailleurs, l'enregistrement sur CD-WORM est totalement irréversible.

### **Client/Serveur**

Le client/serveur est un mode d'interaction en réseau permettant à deux processus de communiquer entre eux, appelés respectivement client et serveur. Le client est le processus qui génère les requêtes au serveur. Le serveur est le processus qui exécute les requêtes et qui renvoie la réponse au client. Ce mode de fonctionnement est généralement opposé à la méthode d'accès par partage de fichiers.

### **Code à barres**

Technique de codage de caractères basée sur des barres d'épaisseur ou d'intervalles variables. Permet la lecture directe des informations codées. Très utilisée dans la grande distribution.

### **DON : Disque Optique Numérique**

Support d'information réalisé par couches et permettant une écriture/lecture par rayon laser. Les normes des supports les plus courants actuellement sont l'ECMA 184 (1,3 Go) pour le 5 pouces 1/4 et ECMA 201 (230 Mo) pour le 3 pouces 1/2.

#### **DON WMRA: (Write Many, Read Always) ou REINSCRIPTIBLE**

Ce type de DON fait appel à des technologies permettant d'écrire, de lire, d'effacer et de réécrire les informations ou les fichiers. Deux technologies sont actuellement utilisées : le magnéto-optique (également appelé thermo-magnéto-optique) et le changement de phase. Les deux catégories de produits existent dans les formats 5,25 ou 3,5 pouces (130 et 90 mm). Ce dernier format est largement utilisé dans la bureautique mais très peu dans la GED.

On notera qu'il existe aussi sur le marché des disques effaçables de grand format, du 12 pouces (30 cm) ; ils sont utilisés soit pour le stockage de masse des ordinateurs centraux, soit dans des applications liées à l'enregistrement de vidéo numérique compressée ou non. Les caractéristiques générales de ce type de DON sont similaires au DON WORM, hormis le fait que, comme son nom l'indique, le support peut être réécrit. A ce titre, ce support se comporte comme un disque magnétique classique. De ce fait, il est un peu moins utilisé pour les applications d'archivage électronique et plus pour des applications de gestion de documents qui ont une courte durée de vie.

Les durées de vie actuellement annoncées par les constructeurs sont de l'ordre de 10 à 15 ans, voire davantage si le substrat des disques est en verre et non en polycarbonate.

Il existe sur le marché des enregistreurs/lecteurs capables d'utiliser à la fois des DON WORM ou pseudo-WORM et des médias effaçables en fonction du support qui est inséré dans l'appareil. Pour l'instant, ces unités dites MULTI-FONCTIONS existent pour le format de DON 5,25 pouces. Ce type de lecteur est apprécié justement pour sa dualité, qui permet à la fois d'utiliser des supports WORM pour l'archivage à long terme et des REINSCRIPTIBLES pour du backup quotidien par exemple. On le trouve souvent dans les juke-boxes.

#### **DON WORM (Write Once, Read Many)**

Le disque optique numérique WORM existe en deux principaux formats qui sont le 5,25 pouces et le 12 pouces. Auquel s'ajoute un disque de 14 pouces (355 mm)

de grande capacité de stockage. Le plus répandu des disques optiques numériques est le 5,25 pouces ou 13 cm de diamètre. La particularité du DON WORM est d'assurer une pérennité aux informations enregistrées sur ce support par le caractère irréversible de ce média ne permettant pas d'effacer les informations par réécriture. Il existe deux catégories de DON WORM. La première est basée sur l'utilisation de médias sur lesquels les informations sont enregistrées par une modification physique de la couche d'enregistrement (ablation, agrégat, thermodégradation, etc.). La seconde catégorie appelée pseudo WORM ou CCW utilise des disques de type magnéto-optique protégés contre la réécriture par le soft du microcode des lecteurs/enregistreurs.

On a coutume de dire que les informations enregistrées sur disque WORM ne peuvent pas être modifiées. C'est vrai pour le commun des utilisateurs, mais pas pour un bon spécialiste, notamment dans le changement de base. Les données réécrites sur d'autres secteurs sans que l'utilisateur s'en aperçoive. Dans le cas du vrai WORM, la supercherie peut être décelée par un expert.

Néanmoins, le DON WORM est un support très sûr dans un cadre normal et est très prisé comme support d'archivage.

Il est d'ailleurs considéré comme support de stockage ayant valeur légale dans certains pays et dans certains états des Etats-Unis. Les durées de vie actuellement annoncées par les constructeurs vont de 15 à 30 ans, voire plus...

L'enregistrement de données sur de tels supports se fait à l'aide d'un lecteur/enregistreur connecté à l'ordinateur ou à un serveur via une interface SCSI.

### **Ethernet**

Topologie de réseau local d'entreprise standardisée au niveau physique et logique, autorisant des débits à 10 Mb/s dans sa version actuelle. Une version à 100 Mb/s est de défini.

[Standard IEEE 802.3 pour le 10 Mb/s et IEEE 802.11 pour le 100 Mb/s (en cours de définition). Topologie logique à base de bus. S'appuie sur le protocole CSMA-CD.]

Voir aussi : TCP/IP, Ethernet

### **Formulaire électronique**

Par analogie au formulaire papier, le formulaire électronique permet de représenter un document dans un format donné, avec un certain nombre de zones à remplir. Les données pour remplir ces zones peuvent être soit saisies directement par l'utilisateur, soit extraites à partir d'une base de données.

### **Full Text**

Synonyme de texte intégral.

### **GED: Gestion Electronique de Documents**

Terme américain synonyme de EDM, DIP, GEIDE.

### **GEIDE ®: Gestion Electronique d'information et de Documents Existants**

Terme déposé par APROGED.

### **GroupWare**

Mode et outils de travail privilégiant le partage de l'information entre les membres d'un groupe. On peut recenser comme exemples d'applications le partage de fichiers, l'utilisation de messageries, les agendas électroniques, les outils d'audioconférence ou de vidéoconférence, la gestion de forums, la mise à disposition de bases de données partagées, etc.

### **HTML: Hyper Text markup language**

Langage de description de pages hypertexte permettant de définir les pages de données d'un serveur WEB accessibles à travers le réseau Internet. Ces pages peuvent contenir du texte, des objets graphiques, du son, et des liens vers d'autres pages HTML ou d'autres serveurs WEB.

Voir aussi: World Wide Web.

### **Image numérisée**

Copie électronique d'un document. L'image numérisée d'un document est généralement obtenue à l'aide d'un scanner ou numériseur.

### **Internet**

Réseau mondial d'ordinateurs, utilisant le protocole TCP/IP, permettant d'échanger du courrier électronique, de partager et de diffuser de l'information. Développé à l'origine dans le cadre des laboratoires de recherche américains, Internet devient au fil des années le réseau d'ordinateurs possédant le plus d'utilisateurs (30 millions à l'heure actuelle).

Depuis deux ans, et l'apparition des serveurs WEB, Internet est devenu très facile d'accès, d'où une explosion du nombre d'ordinateurs connectés. Internet préfigure les autoroutes de l'information.

Voir aussi : WEB, TCP/IP, HTML

### **Juke box**

Les juke-boxes, appelés aussi bibliothèques optiques, sont en fait des armoires de rangement pour les supports optiques ; armoires qui sont munies d'un ou plusieurs robots et qui sont équipées d'un ou plusieurs lecteurs. Le robot est piloté par une application, au travers d'un logiciel de gestion ou "driver", afin de charger le bon support optique dans le bon lecteur au bon moment. Les juke-boxes se connectent en général par le biais d'une interface SCSI à un micro-ordinateur ou une station de travail. Il existe également des juke-boxes intégrant un contrôleur permettant la connexion directe à un réseau local.

Les capacités des juke-boxes vont de quelques dizaines de giga-octets (Go) à plus d'un téra-octet (To). D'après les sources de l'Optical Club, les capacités des juke-boxes vendus en France évoluent entre 30 et 100 Go. Il existe des juke-boxes pour des DON WORM ou effaçables de 12 pouces, pour des médias de 5,25 pouces WORM et WMRA, pour les CD-ROM/CD-WORM et pour les DAT.

Il convient de faire une distinction entre un juke-box et un autochangeur. Un autochangeur est une unité d'enregistrement/lecture munie d'un dispositif de chargement/déchargement de disque permettant d'accéder rapidement aux informations de cinq ou six médias. Il existe aussi des tours d'enregistreurs/lecteurs. Il ne s'agit pas là de juke-boxes car ils ne présentent pas les mêmes possibilités d'extension en capacité.

### **OCR ou ROC en français (Optical Recognition Characters)**

Logiciel de reconnaissance de caractères permettant la traduction de groupes de points d'une image numérisée en caractères (avec l'enrichissement typographique associé) exploitables par des programmes informatiques.

### **OCRA et OCRB**

Polices de caractères normalisées pour faciliter leur interprétation par lecture optique. Les logiciels OCR ont levé la contrainte d'utilisation de ces polices pour les traitements automatiques.

### **ODBC: Open Database Connectivity**

Technique de communication visant à standardiser l'échange d'informations entre applications et base de données.

### **RAID**

Dispositif de stockage périphérique sécurisé fonctionnant sur le principe de la répartition des informations sur plusieurs supports. Plusieurs niveaux (1 à 5) différencient ces dispositifs selon le niveau de sécurité recherché.

### **Scanner**

Périphérique d'acquisition de documents en vue de leur transformation sous forme d'images numérisées.

### **SCSI: Small Computer System Interface**

Interface normalisée de contrôle matériel de divers types de périphériques (disques durs, scanners, lecteurs de disques optiques, etc.). Les normes en cours sont SCSI2, SCSI Fast 2 Wide.

### **SGML: Standard Generalized Markup Language**

Norme internationale définissant une représentation de marquage de texte appelée balise, en vue de faciliter l'échange de documents textuels structurés.

### **TCP/IP**

Protocole de transport d'informations sur un réseau d'ordinateurs, développé à l'origine dans le cadre du réseau Internet sur machines UNIX. Ce protocole est de plus en plus utilisé quels que soient les systèmes d'exploitations (NetWare, Windows NT, etc...). De nombreux logiciels client/serveur sont basés sur ce protocole.

[TCP Couche 4 niveau extrémité (transport), IP couche 3 réseau (routage)]

Voir aussi: Client/Serveur, Ethernet, Internet

### **Texte intégral**

Technique d'indexation de documents de nature textuelle (fichiers bureautiques) ou de champs textes de base de données, où chaque mot composant le texte est en index pour servir de clé de recherche lors d'une interrogation. On associe généralement un dictionnaire de mots vides, ceux-ci étant des mots qui ne seront jamais indexés car apparaissant de manière trop systématique dans les textes, et

donc ne présentant pas de pertinence lors de l'interrogation (les prépositions, les pronoms, etc...).

Lors d'une interrogation, les mots utilisés comme clés de recherche peuvent en général être combinés entre eux par des opérateurs logique de recherche tels que le "ET", le "OU" et le "SAUF", des opérateurs de proximité entre mots (permettant par exemple de rechercher deux mots appartenant à la même phrase), et des opérateurs de troncatures (permettant de rechercher des parties de mots).

### **Thésaurus: dictionnaire organisé**

Liste de termes adaptée à un domaine documentaire donné, organisée par affinité sémantique et complétée de liens de relations.

### **TIFF: Tagged Image File Format**

Standard de fait, Tiff définit une structure de fichier pour le stockage d'images numérisées (recommandé par APROGED).

### **WorkFlow**

Système d'ordonnancement des flux de travaux dans une organisation. A la base du WorkFlow, il y a une modélisation des fonctions et procédures de l'entreprise. La plupart des solutions de WorkFlow prennent en charge la régulation des flux de travaux en prenant en compte des notions de synchronisme, de temps d'exécution et des alertes.

### **World Wide Web (WWW)**

Serveur de données multimédia accessible à travers le réseau Internet. Un serveur WEB est constitué d'un certain nombre de pages d'informations (appelées pages HTML), pouvant contenir du texte, des graphiques, du son, etc... et dans lesquelles des liens vers d'autres pages, appartenant au même serveur ou à d'autres serveurs, peuvent être actionnés. Cette dernière caractéristique permet, à partir d'un serveur WEB, de naviguer à travers le réseau Internet d'un lieu à l'autre en fonction des liens hypertextes existants.

## **Bibliographie**

**AFCET93**, 1993, 1<sup>er</sup> Congrès biennal de l'association française des sciences et technologies de l'information et des systèmes, Système et Cognition, Versailles 1993:

**SYBORD C.**, Approche systémique pour le transfert d'expertise et l'activité cognitive, p.343.

**AFIA**, Association Française pour l'Intelligence Artificielle, Janvier 1995, N°20, Dossier Explication et systèmes à base de connaissances, p.21.

**AFIA**, Association Française pour l'Intelligence Artificielle, Avril 1995, N°21, I.A. et Internet, p.31.

**BARROT M., GANTIER P.**, 1996, Passeport pour l'Internet, Dunod.

**BLANCHE R.**, 1973, Le raisonnement, PUF.

**BONNET A.**, 1984, L'intelligence artificielle. Promesses et réalités, Inter Editions.

**BOUZIDI L., SYBORD C.**, 1995, Système d'Aide à la Décision: Proposition d'une architecture, Actes du 2<sup>ème</sup> congrès biennal de l'AFCET 1995, Toulouse.

**BRUNEAU J.M., PUJOS J.F.**, 1992, Le management des connaissances dans l'entreprise: ressources humaines et systèmes d'information, Les éditions d'organisation.

**COURBON J.-C., STABELL C.B.**, 1986, Artificial intelligence and the design of decision support systems, tutorial of the conference on economics and artificial intelligence, Aix-en-Provence, Septembre 1986.

**COURBON J.C.**, 1983, Les SIAD: outils, concepts et mode d'action, AFCET interfaces 9.

**COUTROT F.**, 1985, Typologie des SIAD: offre et marchés, in SIAD-85.

**DEMAILLY A., LE MOIGNE J.L.**, 1986, Sciences de l'Intelligence Sciences de l'Artificiel, Presse Universitaires de Lyon.

**ERNST C.**, 1993, Les systèmes interactifs d'aide à la décision: Facteurs de succès de l'intelligence artificielle en gestion de l'entreprise, 1<sup>er</sup> Congrès Latino-Américain, Brésil.

**GALACSI**, 1984, Les systèmes d'information, Dunod, Paris.

- GRUNDSTEIN M., DE BONNIERES P., PARA S.**, 1988, Les systèmes à base de connaissances, Systèmes experts pour l'entreprise, AFNOR GESTION.
- HEURGON E.**, 1985, Enjeux des systèmes interactifs d'aide à la décision (SIAD) et stratégie d'entreprise, in SIAD-85.
- LANGLEY P., SIMON H.A.**, 1981, The central rôle of learning in cognition. In : J.R. Anderson (ed), Cognitive skills and their acquisition, 1981, Hillsdale, Erlbaum.
- LAURIERE J.-L.**, 1985, Raisonnement humain et systèmes expert, in Cognitiva 85, CESTA Ed., Paris.
- LE MOIGNE J.-L.**, 1974, Les systèmes de décision dans les organisations, PUF, Paris.
- LE MOIGNE J.-L.**, 1984, La théorie du système général. Théorie de la modélisation, PUF.
- LEVINE P., POMEROL J.-C.**, 1989, Systèmes d'aide à la décision et systèmes experts, Traité des nouvelles Technologies, série Décision assistée par ordinateur, HERMES.
- NEWELL A., SIMON H.A.**, 1972, Human Problem Solving, Prentice Hall, Englewood Cliffs.
- PITRAT J.**, 1987, Naissance de l'IA, Recherche en IA, Edition Seuil, Paris.
- QUILLIAN M.R.**, Word concepts. A theory and simulation of some basic semantic capabilities, in Readings in knowledge representation, op. cit.
- SANDOVAL V.**, 1996, Intranet, le réseau d'entreprise, Hermès.
- SIMON H.A.**, 1977, The new science of management decision, Prentice-Hall, New-Jersey.
- SMITH J.M.**, 1977, Database abstraction: aggregation and generalization, ACM TODS, vol.2, N°2, PP.105-133.