

INTRODUCTION

PROBLEMES LIES A L'INFORMATION EN SANTE

Les domaines de la médecine, de la biologie et de la santé publique couvrent une part de plus en plus importante de la connaissance et de l'activité, justifiant le recours aux méthodes de traitement de l'information, l'informatique. En médecine notamment, le nombre et la variété des éléments constitutifs de cette connaissance ne font que croître. Plusieurs raisons y concourent :

- le développement de la connaissance clinique et physio-pathologique ;
- la technicité croissante des examens dits complémentaires ;
- la multiplication des structures de prise en charge ;
- l'allongement de la durée de vie des individus.

L'augmentation du nombre de paramètres nécessaires à la prise en charge des patients pose le problème de la maîtrise de l'information.

En santé publique, l'information pertinente concerne des groupes d'individus et comprend des données démographiques, épidémiologiques, sanitaires, sociales, économiques voire politiques. Le problème est plutôt ici celui de connaissances parcellaires ou isolées.

Enfin, les différents acteurs du système de santé prennent conscience de contraintes économiques et éthiques.

DYNAMIQUE DE L'INFORMATION

L'information utilisée en médecine ou en santé publique ne pose pas seulement des problèmes statiques de sur ou de sous-information. Elle intervient de façon dynamique dans des processus de décision et s'organise en systèmes et sous-systèmes.

Prenons à titre d'exemple les activités d'un hôpital universitaire où interviennent plusieurs sous-systèmes interdépendants :

- sous-système d'information de l'action médicale : l'équipe médicale et soignante recueille sur chaque malade des données, celles-ci sont confrontées à l'expérience et à la connaissance médicale pour être interprétées au cours d'un processus de décision qui aboutit à une action. Cette action interagit sur le malade. Un dossier permet d'enregistrer les données relatives au patient, il sert de support et de mémoire au processus de décision. Enfin, le malade reçoit en retour certaines informations sur son état ou son traitement.

- sous-système de la recherche : si les dossiers ont été constitués de façon correcte et homogène, ils peuvent être exploités dans la perspective de l'évaluation de la qualité des soins, de la recherche clinique ou de l'épidémiologie et contribuer ainsi à l'évolution de la connaissance médicale, en parallèle avec la recherche fondamentale.
- sous-système de la logistique hospitalière : chaque action est source d'informations et de flux d'informations vis-à-vis des autres structures hospitalières : autres services, plateaux techniques, laboratoires, pharmacie, archives.
- sous-systèmes de gestion et d'administration : ils sont concernés par la facturation des actes, la gestion du personnel, la gestion des stocks.
- sous-système de planification : il répond à la politique du développement hospitalier.

A un niveau supérieur, on pourrait considérer le système régional ou national de la planification sanitaire et sociale.

Ainsi dans de nombreuses activités, il existe un cycle à trois phases : observation/raisonnement/action qui se décline en médecine clinique en : observation/diagnostic/thérapeutique et en informatique en : acquisition/traitement/sortie. La phase de raisonnement transforme les données brutes en information, utilisable pour l'action. Le travail théorique sur la nature de l'information et les divers processus où elle intervient permet de dégager les possibilités et les limites des méthodes de traitement de l'information.

NATURE DE L'INFORMATION

Il est difficile de définir ce qu'est l'information mais on peut formaliser les situations qui la mettent en jeu : un émetteur E transmet un message ou signal S à un récepteur R via un canal (medium) ; en général un bruit ou une distorsion D se surajoute au signal qui est perçu sous forme d'un mélange $M=S+D$. Différentes configurations existent selon que la transmission est uni ou bi-directionnelle, selon le support ou selon que l'intérêt porte sur le message ou l'état du médium.

L'information, qui n'existe que grâce à un support, peut être décrite sous trois aspects :

- syntaxique : il concerne les règles de construction des messages par combinaison des symboles utilisables ;
- sémantique : il traite de la signification du message, du passage des données brutes à l'information, cette interprétation nécessitant souvent la connaissance du contexte ;
- pragmatique : il s'intéresse à l'action induite par l'interprétation, parfois réduite à une diminution d'incertitude.

Shannon en 1948 a donné une définition mathématique de l'information I, égale à l'opposé du logarithme de la probabilité d'occurrence du message ($I = -\log_2 p$ avec $0 \leq p \leq 1$). Ce contenu d'information, exprimé en bit (plus petite unité d'information, pouvant exprimer une valeur binaire comme vrai/faux) établit que la survenue d'un signal rare apporte plus d'information que celle d'un message fréquent. Le contenu d'information du message reçu ne

peut être supérieur à celui du message émis, dans un système clos ; il peut être inférieur en raison du bruit surajouté. La valeur négative de I peut être rapprochée de l'entropie qui exprime le désordre d'un système thermodynamique. Afin de réduire cette dégradation, plusieurs principes peuvent être appliqués : utiliser un capteur optimal, avoir le canal de transmission le plus court possible et avec le plus faible bruit ajouté, s'appuyer sur la redondance de l'information, disposer de connaissance a priori pour l'interprétation.

Les données peuvent se présenter sous différentes formes : nombres entiers, nombres réels, codes ou texte libre (langage naturel). Dans tous les cas, plusieurs qualités doivent être recherchées : la complétude (absence de données manquantes), l'exactitude (absence de biais, conformité à une définition) et la précision. L'utilisation de codes renforce la standardisation de l'information mais suppose une interprétation et impose stricte définition des codes. L'usage de texte libre offre le plus de souplesse mais expose aux ambiguïtés et aux données manquantes.

APPORT DES SCIENCES DE L'INFORMATION

L'informatique est la science du traitement automatique de l'information, héritière de la théorie de l'information et bénéficiant de l'apport des mathématiques, de la statistique, de la linguistique, des sciences cognitives et de la philosophie. L'informatique est aussi une technique, portée par le formidable développement technique et industriel de l'électronique.

En médecine et en santé publique, l'informatique est d'abord une méthode imposant formalisation de l'information et rigueur de raisonnement. De l'application de ces principes, plusieurs bénéfices peuvent être espérés :

- augmenter la fiabilité des données (saisie, enregistrement, transmission) ;
- aider à la mémorisation et à la réutilisation de données complexes ;
- comprendre les mécanismes d'interprétation et de raisonnement médical ;
- sélectionner les données les plus pertinentes parmi la masse des informations disponibles ;
- évaluer les résultats des actions entreprises ;
- rationaliser les choix au niveau individuel ou collectif (application de protocoles) ;
- partager l'information et fédérer les systèmes d'information ;
- faciliter l'accès à la connaissance.

Le critère primordial, souvent en raison du pronostic vital, sera celui de la validité des données.

OBJECTIFS

- Citer les problèmes de l'information en médecine et en santé publique.
- Distinguer les concepts d'information et d'informatique.

INFORMATIQUE

DEFINITIONS

Informatique = science et technique du traitement automatique de l'information.

Ordinateur = machine électronique de traitement automatique de l'information, fonctionnellement stupide, mais disciplinée, rapide et dotée d'une grande mémoire. On lui fournit des données, sur lesquelles elle effectue des opérations selon un schéma préalablement établi par l'homme : le programme, et elle fournit des résultats.

HISTORIQUE

L'ordinateur ne devient une réalité qu'après la 2^{ème} guerre mondiale avec le développement de l'électronique. Mais le traitement automatique de l'information est plus ancien et a d'abord été effectué par des connexions mécaniques. On obtient des **machines à programme figé**, spécialisées dans un certain type de travail. Dès le XVII^{ème} siècle, la machine à calculer de Pascal (1642) est construite pour effectuer des opérations arithmétiques grâce à un système d'engrenages. Le programme est codé de manière irréversible dans la structure de la machine.

Pour avoir plus de souplesse, il faut utiliser des **machines à programme enregistré** : comme le métier à tisser de Jacquard (1805) qui est un automate à programme externe, matérialisé par des cartons perforés. La suite des instructions à exécuter, c'est-à-dire le programme, est mémorisée par l'enchaînement des cartons.

La première conception théorique de l'ordinateur date de 1840 : Babbage réunit les idées de Pascal (calcul) et Jacquard (programme) et imagine une **machine analytique** qui doit exécuter une suite d'opérations grâce à des instructions successives enregistrées sur des cartes perforées. Cette machine universelle permet de changer le programme sans modifier la structure et de traiter différents problèmes. Mais les capacités techniques de l'époque ne permettaient pas de construire cette machine.

A la même époque (1854), le mathématicien **Boole** publie un ouvrage intitulé "Les lois de la pensée" où il fait de la logique formelle une discipline mathématique et non plus philosophique. Il développe une algèbre travaillant sur des mots binaires formés de 0 et de 1 : à une proposition vraie il associe le chiffre 1, à une proposition fausse le chiffre 0 et il définit des opérations à l'aide des opérateurs logiques ET, OU, NON. Cette algèbre binaire,

ultérieurement perfectionnée par De Morgan, est adaptée à une représentation physique à l'aide de dispositifs à deux états : marche/arrêt, perforé/intact ; elle est encore utilisée dans la conception des circuits électroniques.

D'un point de vue technique, les premières **machines à calculer à cartes perforées** apparaissent vers 1890 quand H. Hollerith invente une machine pour dépouiller les données du recensement américain. Les données sont représentées sur des cartes perforées, sous forme binaire (perforé ou non). Au XXe siècle apparaissent les premières entreprises d'équipements électro-mécaniques : Bull du norvégien F. Bull qui a introduit l'emploi de l'électricité dans les calculateurs et IBM, l'entreprise de Hollerith, qui, devenue la première du secteur, impose souvent ses choix technologiques.

Les besoins en calcul de la deuxième guerre mondiale conduisent à la construction de **calculateurs électroniques** où des tubes électroniques, remplaçant les relais électromécaniques, génèrent et contrôlent le déplacement des charges électriques permettant le calcul. Ainsi, ENIAC (Electronic Numerical Integrator and Calculator, 1946, USA) est une machine entièrement électronique fonctionnant avec 18000 tubes à vides qui utilise des programmes écrits sur des cartes perforées dont la lecture s'effectue en même temps que l'exécution de chacune des instructions élémentaires codées sur ces cartes, car la machine n'a pas de mémoire autre que celle conservant les résultats. ENIAC fonctionne encore selon un principe proche du boulier et le chargement du programme, manuel, pouvait prendre plusieurs jours pour des programmes complexes.

En 1947, en concevant l'EDVAC (Electronic discrete variable computer), le mathématicien **Von Neuman** réalise la synthèse entre les machines à programme figé interne et les machines à programme enregistré externe : les programmes sont manipulés comme des données et sont enregistrés dans la vaste mémoire interne de l'ordinateur avant l'exécution et non plus au fur et à mesure ; les décisions logiques sont prises par le système lui-même qui dispose d'une unité de contrôle et non plus par l'opérateur qui modifie les circuits devant un tableau de câblage. Cette conception est à la base des machines modernes à **programme enregistré** ou ordinateurs, qui sont des machines effectuant des opérations logiques et non plus uniquement des calculs. Le premier ordinateur commercial conçu sur le principe de Von Neumann est produit en 1950.

Vers 1960 apparaissent les machines à **transistors** : les transistors (inventés en 1947 dans les laboratoires de Bell, multinationale du téléphone) remplacent les tubes électroniques car ils sont plus rapides, plus petits, plus fiables, consomment moins d'énergie et coûtent moins cher. Le transistor est un circuit très petit, réalisé par la diffusion contrôlée d'impuretés dans un matériau semi-conducteur, le silicium. Son fonctionnement est celui d'un interrupteur à commande électrique, à deux états stables : allumé ou éteint, adapté donc à la représentation de l'algèbre binaire.

L'essor actuel de l'informatique est lié aux progrès de l'électronique, en particulier à l'**intégration** toujours plus poussée des composants (transistors, résistances, capacités) sur une même pastille de silicium. En 1950, le transistor est gros comme un petit pois ; en 1970,

les circuits intégrés LSI (Large Scale Integration) contiennent plus de 1000 transistors et le **microprocesseur** intègre les éléments essentiels d'un ordinateur classique sur un seul circuit (puce ou chip) de silicium. Depuis la fin des années 80, on place plus d'un million de transistors par cm^2 de silicium.

Miniaturisation et accroissement des performances vont de pair : un micro-ordinateur récent (construit autour d'un microprocesseur 32 bits exécutant 100 millions d'instructions élémentaires par seconde) est plus rapide et plus puissant que l'Eniac de 1946 qui avait une surface de 150 m^2 , consommait 200 KW et n'effectuait que 5000 additions à la seconde. La décroissance des coûts est de 30% par an, l'intégration et la puissance sont multipliées par 2 tous les 18 mois. L'électronique est maintenant le premier secteur industriel mondial et il y a plus de 100 millions de micro-ordinateurs en fonctionnement dans le monde.

MATERIEL

L'informatique fait appel à du matériel et à du logiciel :

- le matériel, c'est ce qui se touche, en anglais Hardware (quincaillerie), abrégé en Hard ;
- le logiciel est virtuel, c'est le programme, en anglais Software (par opposition à Hardware) ou Soft.

L'ordinateur, matériel composé de circuits électroniques, est en fait un support physique figé : si des programmes n'ont pas été introduits dans la machine, celle-ci reste inerte. Les machines actuelles sont digitales binaires, c'est-à-dire qu'elles ne peuvent manipuler que des informations discrètes sous forme binaire (représentées par 0/1 ou éteint/allumé)

Les **Organes** de l'ordinateur comportent :

- une **unité centrale** constituée d'une **mémoire centrale** (pré-enregistre les données et méthodes sous forme binaire avant leur utilisation, selon le principe de Von Neumann) et d'un **processeur** (effectue les opérations sur ces données)
- des **périphériques** de communication et de stockage.

L'être humain qui calcule " $12 \times 3 = 36$ " effectue le même type d'opérations qu'un ordinateur (même si l'organisation physique est radicalement différente) : il existe des données qu'il faut lire, mémoriser et consulter, une méthode de traitement, qui doit être décomposée si elle n'est pas connue par construction, un résultat qu'il faut écrire. Il met ainsi en jeu des organes d'entrée et de sortie, une mémoire centrale (mémoire consciente, à court terme), un processeur et une mémoire auxiliaire (mémoire à long terme ou livre) qui contient de façon permanente la description de la méthode.

Unité centrale

Le **processeur** (en anglais central processing unit, CPU) est un circuit intégré très complexe, capable par construction d'effectuer un nombre limité d'opérations élémentaires appelées instructions, déterminées par le constructeur. Ce circuit contrôle et exécute, il est donc fonctionnellement composé de deux parties :

- un organe de traitement (**unité arithmétique et logique** ou UAL) capable d'effectuer des opérations arithmétiques, relationnelles et logiques sur des mots binaires de longueur fixe ;
- un organe de commande (**unité de contrôle** ou UC) qui gère l'enchaînement des différentes instructions du programme chargé en mémoire centrale.

Il fonctionne séquentiellement selon un cycle : lecture d'une instruction, analyse de cette instruction, recherche des opérandes, exécution de l'instruction par l'UAL et lecture de l'instruction suivante. Ces opérations sont rythmées par une horloge interne à quartz dont la fréquence est mesurée en MegaHertz (MHz). La puissance d'exécution du processeur dépend de la vitesse de l'horloge et de la taille des mots (données ou instructions) que traite le processeur. La vitesse s'exprime en million d'instructions par seconde (MIPS) et est de l'ordre de plusieurs centaines de MIPS sur les micro-ordinateurs récents. Les constructeurs de processeurs (Intel, Motorola) développent des familles de processeurs, de plus en plus puissants, mais qui partagent le même jeu d'instructions ce qui permet aux mêmes programmes de fonctionner sur toute la gamme sans réécriture. En effet, il existe une forte complémentarité entre le matériel et le logiciel puisqu'un programme est la liste des instructions élémentaires à exécuter séquentiellement par le processeur quand il n'y a pas de circuit dédié à l'opération désirée, (par exemple, le programme de la multiplication consiste en une répétition d'additions s'il n'y a pas de circuit multiplicateur).

La **mémoire centrale** (MC) est un dispositif électronique, composé par des circuits à très grande intégration, capable de contenir sans altération d'importantes quantités d'informations : les données à traiter, les résultats, et aussi (selon le principe de Von Neuman) le programme. La mémoire centrale est organisée comme un tableau de petites cellules rangées en lignes et en colonnes. Chaque cellule contient une unité élémentaire d'information qui ne peut prendre que deux états différents et correspond à un élément binaire prenant les valeurs 0 ou 1, en anglais "binary digit" ou **bit**. Physiquement, ceci est réalisé par une tension électrique qui vaut soit 0V soit 5V, ou encore par une capacité soit chargée soit déchargée.

Les valeurs 0 et 1 constituent tout l'alphabet de l'ordinateur, on peut les regrouper en mots binaires, caractérisés par leur longueur : mot de 4 bits, mot de 8 bits ou octet ("byte" en anglais). Chaque ligne de la mémoire centrale constitue un mot qui est la quantité insécable que la mémoire échange avec le reste de l'unité centrale. Un nombre important de bits peut s'exprimer dans une autre unité : le Kilobit = 1024 (2^{10}) bits, le Megabit = 1 048 576 (2^{20}) bits. En fait on parle plus souvent en Ko ou Kilo-octet (1024 octets), Mo ou Méga-octet et même Go ou Giga-octet (2^{30} octets).

La mémoire centrale est physiquement et fonctionnellement composée de deux parties : la mémoire vive ou **RAM** (Random Access Memory) et la mémoire morte ou **ROM** (Read Only Memory). L'unité centrale peut à tout moment lire dans la RAM et dans la ROM mais seule la RAM est modifiable par écriture. La mémoire centrale est essentiellement constituée de RAM qui est **volatile**, c'est la mémoire de travail mais son contenu (programme en cours d'exécution, données, résultats) est effacé à l'extinction de la machine. Actuellement on utilise couramment des puces de RAM d'une capacité d'un Mo, alors qu'on développe des

puces de plusieurs Mo. Ainsi, dans une simple composant, on peut faire tenir plus de 500 pages (1 page = 40 lignes de 50 caractères = 2000 caractères soit 2000 octets).

Périphériques

Les **périphériques** assurent l'échange d'informations entre l'unité centrale et extérieur : l'utilisateur ou une autre machine. L'interface d'entrée/sortie (E/S ou en anglais Input/Output ou I/O) assure le raccordement de l'unité centrale aux périphériques qui fonctionnent à des vitesses et avec des signaux différents.

Il existe des **périphériques d'entrée** comme le clavier alphanumérique ou la souris (permet le déplacement d'un curseur à l'écran) et des **périphériques de sortie** comme l'écran graphique ou l'imprimante. Les périphériques d'entrée codent les données à traiter sous forme binaire utilisable par les circuits électroniques et les périphériques de sortie décodent ces valeurs binaires en les transformant en symboles compréhensibles.

Lors de la transmission de données entre ordinateurs via une ligne téléphonique, il faut intercaler un périphérique de conversion. En effet, les signaux traités par l'émetteur et le récepteur sont digitaux binaires (constitués des chiffres 0 et 1, émis avec un faible voltage), tandis que les signaux véhiculés sur les lignes de communication sont analogiques (sinusoïdaux). Un ou plusieurs paramètres (amplitude, fréquence ou phase) d'un signal sinusoïdal - la porteuse - sont modulés en fonction de l'information à transmettre. Une double transformation est donc nécessaire : modulation à la sortie de l'émetteur, démodulation à la sortie du récepteur. Cette opération est effectuée par le **modem** (MODulateur-DEModulateur) qui doit exister en deux exemplaires, un fonctionnant en modulation avant la ligne de communication et l'autre en démodulation après la ligne.

La mémoire centrale est une mémoire électronique, très rapide, mais de relativement faible capacité et qui ne fonctionne qu'alimentée en courant. Il existe donc des périphériques de stockage permanent (équivalent des livres et de la bibliothèque). Ces **mémoires secondaires** ou mémoires de masse, permettent de conserver les informations qui disparaissent de la mémoire centrale quand on éteint l'ordinateur et peuvent fonctionner en lecture et/ou en écriture. Ces informations sont regroupées sous forme de fichiers.

Il s'agit généralement de mémoires magnétiques qui existent sous différentes formes :

- disquette souple, de faible capacité (1,4 Mo), lente mais amovible ;
- disque dur, de grande capacité (de l'ordre du Go), rapide mais généralement fixe ;
- bande magnétique, d'accès long car séquentiel mais peu coûteuse.

Cependant l'usage des mémoires optiques (CD-ROM, d'une capacité de 600 Mo, adapté à la diffusion de données en lecture seule car reproduit par pressage) ou magnéto-optiques (réinscriptibles) se développe car elles offrent des capacités de stockage supérieures.

Nature des informations

Pour qu'un renseignement, une information, soit transmis et interprété (utilisé), il faut disposer d'un support et d'un code. Le support dépend du mode de transmission et le code est une convention entre l'émetteur de l'information et son destinataire. Les supports du langage écrit sont les mots quand ceux du langage parlé sont les sons. Ces deux langages sont basés sur la combinaison d'éléments plus petits, respectivement les lettres de l'alphabet et les phonèmes.

Les informations sont de natures très diverses. On peut les répartir selon leur degré de complexité :

- données **logiques** : essentiellement logique binaire à deux états, vrai ou faux s'excluant mutuellement, par exemple : le patient tousse ou ne tousse pas;
- données **numériques** : il peut s'agir de données entières qualitatives représentant une qualité ou une classe, ordonnées (testing musculaire) ou non (classification internationale des maladies); il peut s'agir également du résultat d'une mesure, exprimé par un nombre entier (nombre d'enfants) ou réel (glycémie);
- données **littérales** : texte libre descriptif d'une situation, par exemple : le nom ou le compte rendu opératoire;
- signaux **analogiques** : un signal qui varie en fonction du temps, à une dimension (ECG) ou à deux dimensions (angiographie).
- objets divers : ils doivent être représentés par un ou plusieurs nombres (opération de codification) afin d'être représentés sur un support informatique.

Support des informations en machine

Le matériel informatique est constitué de circuits électroniques. Il est facile de réaliser 2 états stables en électronique (allumé/éteint) et donc de représenter une information binaire par un circuit élémentaire. La combinaison de circuits élémentaires permet de représenter des informations plus complexes. Toutefois, l'alphabet du langage utilisé par la machine ne comportant que ces deux symboles : 0 et 1, toutes les informations (données ou instructions) susceptibles d'être traitées en informatique doivent être écrites à l'aide des deux seules lettres 0 et 1.

Cette écriture est plus ou moins facile selon la nature des informations :

- donnée logique : la transcription est immédiate, faux/vrai est équivalent à 0/1;
- donnée numérique : il faut convertir le nombre en numérotation décimale en nombre en numérotation binaire ;
- signal analogique continu : après avoir été découpé dans le temps, sa valeur à chaque instant est convertie en un nombre ;
- autre type de donnée : on doit associer à chaque objet (lettre, diagnostic, malade) un nombre le représentant, c'est la codification.

Cette codification doit, pour pouvoir être utilisée, posséder un certain nombre de qualités :

- facilité de mémorisation et de compréhension ;
- sûreté, soit facilité de détection voire de correction des erreurs de transmission ;
- économie, utilisation du minimum de positions et de symboles, mais ceci est contradictoire avec la sûreté ;
- être opératoire : facile à utiliser pour le traitement ;
- surtout biunivocité et déchiffrabilité unique.

Représentation des informations en binaire

Un mot de un bit peut coder deux positions (0 ou 1), un mot de deux bits code quatre positions (00, 01, 10, 11 soit de 0 à 3) soit 2^2 valeurs, et un mot de n bits peut coder 2^n valeurs différentes.

Le **codage des nombres** distingue les nombres entiers des nombres réels. Classiquement, les **entiers** sont représentés sur deux octets soit 16 bits, ce qui donne 2^{16} soit 65536 possibilités réparties de -32768 à 32767. Les **réels** sont représentés par convention en virgule flottante, c'est à dire écrits sous la forme d'une mantisse et d'un exposant. Selon la précision désirée, ils sont codés par des mots de quatre octets (simple précision autorisant 7 chiffres significatifs après la virgule sur une plage de 10^{-38} à 10^{+38}) ou de huit octets (double précision permettant 15 chiffres significatifs après la virgule allant de 10^{-308} à 10^{+308}).

Le **codage des caractères** consiste à établir une correspondance biunivoque entre ces caractères et les états d'un certain nombre de bits. Pratiquement, on établit une liste ordonnée de ces caractères, et on fait correspondre à chaque caractère son numéro d'ordre exprimé en base 2. Ainsi les caractères alphanumériques (l'alphabet romain + les chiffres arabes + les signes de ponctuation) sont généralement codés à l'aide du code **ASCII** (American Standard Code for Information Interchange). Ainsi, le caractère 'A' est en 65ème position et est représenté par l'octet 01000001, 'B' en 66ème position donne 01000010, 'a' en 97ème position donne 01100001, tandis que le caractère '0' en 48ème devient 00110000 et '1' en 49ème devient 00110001. Le codage s'effectue sur un octet car la réunion des 26 lettres de l'alphabet, soit 52 minuscules et majuscules, des dix chiffres et des signes de ponctuation aboutit à une centaine de symboles différents, nécessitant au minimum 7 bits ($2^6=64$ et $2^7=128$) pour coder ces différents caractères. Le 8ème bit (à gauche) est un **codage de sécurité** qui introduit une redondance : appelé bit de parité, il permet de détecter des erreurs simples lors de transmission de données.

Le **codage des instructions** consiste à numéroter toutes les instructions élémentaires du processeur et à attribuer un code à chacune. Ainsi un programme consiste en la succession des codes correspondant aux instructions à exécuter séquentiellement, complétée d'une instruction de fin de programme.

Le logiciel d'un ordinateur est composé par l'ensemble des programmes qu'il peut exécuter. Un programme, selon le principe de Von Neuman, doit être transféré en mémoire centrale avant son exécution.

On distingue au sein du logiciel :

- les logiciels d'exploitation destinés à faciliter le fonctionnement interne de la machine et son utilisation : système d'exploitation et utilitaires divers, ou servant à construire des applications : langages de programmation ;
- les logiciels d'application ou progiciels conçus pour la résolution d'un problème précis : traitement de texte, gestion de fichier ou calcul.

Langages de programmation

Le processeur ne peut exécuter que des programmes constitués de l'enchaînement d'instructions élémentaires. On peut donc écrire des programmes en langage machine interne, c'est à dire en binaire mais c'est peu commode, car le jeu d'instructions est limité et il faut donc décomposer toute opération un peu complexe en une longue suite d'instructions élémentaires, enfin ce langage est difficile à manipuler.

On a donc inventé des langages évolués, plus proches du langage humain et indépendants du processeur, comme le Basic, le Pascal ou le C, qui permettent d'écrire des programmes portables (transportables d'une machine à une autre). Avant son utilisation par le processeur, le programme écrit en langage évolué (programme source) doit être traduit en langage machine (programme objet) par un programme traducteur spécialisé. La traduction peut se faire au fur et à mesure de chaque instruction du programme source et le programme traducteur est appelé un interpréteur, ou bien une fois pour toute avant exécution, c'est alors un compilateur.

Quel que soit le langage utilisé, le but de la plupart des programmes est de résoudre un problème. Les programmes peuvent résoudre ces problèmes en manipulant des informations ou données. A cette fin, il faut pouvoir :

- fournir les données au programme ;
- les décrire et les conserver quelque part ;
- donner les instructions nécessaires à leur traitement ;
- restituer le résultat du traitement à l'utilisateur.

On peut organiser les instructions de façon que :

- certaines d'entre elles ne soient exécutées que lorsqu'une condition spécifique (ou un ensemble de conditions) est vraie ;
- d'autres soient répétées un certain nombre de fois ou tant qu'une condition est vérifiée ;
- d'autres soient décomposées en fragments qui pourront être exécutés en différents points du programme.

Voici décrites les sept notions de base de la programmation : les entrées, les types de données, les opérations, les sorties, l'exécution conditionnelle, les boucles et les sous-programmes. La plupart des langages de programmation fournissent ces fonctionnalités qui sont combinées pour traduire un **algorithme** : méthode de résolution d'un problème donné.

Système d'exploitation

Les systèmes d'exploitation (SE) sont des programmes qui permettent à l'utilisateur d'utiliser les ressources matérielles et logicielles de l'ordinateur : gestion des organes de la machine (processeur, mémoire centrale, périphériques) et des travaux qui lui sont confiés (chargement des applications, gestion des fichiers). Sans système d'exploitation, la machine est inutilisable et un programme d'application ne fonctionne que par l'intermédiaire du système d'exploitation. Le système d'exploitation sert en outre d'interface avec l'utilisateur en lui fournissant un environnement permettant d'utiliser la machine sous la forme d'un interpréteur de commandes ou d'un environnement multi-fenêtré.

Un système d'exploitation se caractérise par les processeurs qu'il peut exploiter, sa capacité d'exécuter plusieurs tâches en même temps (multi-tâche) et d'obéir à plusieurs utilisateurs indépendants simultanément (multi-utilisateur). Différents systèmes sont actuellement utilisés, dont certains sont dédiés à un type de machine et de processeur (Ms-Dos puis Windows pour les micros de type PC, Mac-Os pour les micros Macintosh, VMS pour les machines du constructeur DEC) tandis que d'autres fonctionnent sur différents types d'ordinateurs (Unix).

Progiciels

Le temps nécessaire à l'écriture et à la mise au point d'un programme est relativement élevé (plusieurs annéesxhomme). On trouve donc dans le commerce des programmes d'application tous faits, correspondants aux principales utilisations d'un ordinateur : les progiciels. Un progiciel est caractérisé bien sûr par ses fonctionnalités, mais aussi le système d'exploitation nécessaire, le type d'ordinateur utilisable, ses performances, son prix, sa facilité d'utilisation. Il existe des progiciels de traitement de texte, de gestion de fichiers, de calcul financier, de calcul scientifique et statistique, de dessin, etc ...

INTERFACE HOMME-MACHINE

On regroupe sous ce terme toutes les techniques, matérielles et logicielles, permettant à un utilisateur de dialoguer avec un ordinateur. Elles se sont beaucoup développées avec la micro-informatique, destinée en principe à des non-spécialistes pour lesquels la convivialité (facilité d'utilisation) est un critère important.

Au niveau matériel, on trouve des périphériques d'entrée comme le clavier, la souris, l'écran tactile, mais aussi la commande vocale ou la reconnaissance de l'écriture tandis que les périphériques de sortie comprennent l'écran, les imprimantes, voire la synthèse vocale.

Au niveau logiciel, il existe deux façons de dialoguer avec un programme :

- soit utiliser des mots de commandes reconnus du programme et qu'il faut donc apprendre et utiliser à bon escient ;
- soit désigner son choix (en tapant son numéro, en déplaçant un curseur à l'écran par le clavier ou la souris) dans des menus qui présentent à chaque étape les différentes options possibles. De plus en plus, les options et les objets à manipuler sont représentés sur l'écran sous forme de symboles graphiques ou d'icônes.

Ce dialogue peut intervenir selon deux modalités, non exclusives :

- soit directement au moment où on en a besoin, c'est le mode interactif ;
- soit en regroupant les ordres dans un fichier de commandes que l'on fait exécuter par le système d'exploitation, c'est le mode dit Batch (traitement par lots) qui permet une exécution automatique en différé sans présence de l'utilisateur.

Les systèmes d'exploitation gérant le multi-fenêtrage permettent d'afficher simultanément dans plusieurs zones indépendantes de l'écran (fenêtres) plusieurs parties d'un texte, plusieurs textes ou images différents ou les résultats de l'exécution de différents programmes.

SYSTEMES INFORMATIQUES

En associant un processeur, une mémoire centrale et des périphériques, on obtient un ordinateur, dont il existe diverses catégories.

Micro-ordinateurs

Le micro-ordinateur, apparu au début des années 70, est un ordinateur construit autour d'un microprocesseur. Il est à l'origine conçu pour une utilisation individuelle, par un non-informaticien, et tient sur un bureau. Il sera popularisé par une machine, l'Apple II et un logiciel, Visicalc, qui permet de faire facilement des calculs sur des tableaux de chiffres et assure sa diffusion dans les entreprises.

Actuellement, le processeur travaille sur des mots de 32 bits, la mémoire centrale a une capacité de plusieurs Mo, les mémoires secondaires sont en général des disques magnétiques de quelques Go. Les périphériques sont très développés ainsi que les logiciels, conçus pour une utilisation aisée. Le système d'exploitation est encore en général mono-utilisateur mais de plus en plus souvent multi-tâche. Les machines les plus répandues sont actuellement l'IBM-PC et l'Apple Macintosh, alors qu'est apparue une norme de fait autour du PC d'IBM depuis 1983 avec la large diffusion de machines dites "compatibles". Les micro-ordinateurs offrent le meilleur rapport performances/prix et sont les ordinateurs les plus diffusés. Ils se

dotent maintenant de capacités multi-média (traitement du son et de l'image vidéo) qui exigent de grandes capacités de calcul, de mémoire et d'affichage.

Deux tendances sont à l'œuvre en micro-informatique : l'augmentation de puissance qui rejoint celle des mini-ordinateurs et la miniaturisation qui permet de produire des machines portables voire des assistants digitaux personnels.

Systèmes classiques

Tout le reste, c'est à dire les mini-ordinateurs et les gros ordinateurs (mainframes en anglais). L'unité centrale est composée de plusieurs circuits et processeurs, la mémoire centrale compte plusieurs MO, les mémoires de masse regroupent des disques de plusieurs centaines de MO et de lecteurs de bandes magnétiques. Le système d'exploitation est complexe et multi-utilisateur. Les utilisateurs se connectent sur ce type de machine par l'intermédiaire d'une liaison spécialisée et d'un terminal, ensemble composé d'un écran et d'un clavier sans processeur ni mémoire. En effet dans ce type de système, toute la puissance de traitement est concentrée dans une seule machine centrale. Les mainframes nécessitent une équipe d'administration et de maintenance, un local aéré et protégé mais permettent à plusieurs centaines d'utilisateurs de se connecter simultanément et de partager des moyens de traitements importants. Les minis sont plus petits et un peu moins puissants. Ils correspondent aux besoins des groupes de dizaines d'utilisateurs, catégorie où ils sont en compétition avec les réseaux de micro-ordinateurs.

Réseaux informatiques

Quand on fait communiquer deux ordinateurs par un câblage, permanent ou non, on obtient un réseau. Le réseau permet le partage et l'échange d'informations, notamment l'accès aux vastes bases de données gérées sur les gros systèmes, mais aussi le partage de ressources logicielles (fichiers, applications) ou matérielles (disques ou imprimantes) et la communication entre utilisateurs (messagerie électronique).

Ces réseaux peuvent être locaux, par exemple grâce au câblage permanent de fibres électriques ou optiques à l'intérieur d'un immeuble, ou distants. La communication à distance peut emprunter le réseau téléphonique (commuté de base ou Numeris, plus rapide), on a alors un réseau télématique, non permanent. La tendance actuelle est l'interconnexion directe et permanente des réseaux informatiques locaux via l'inter-réseau mondial Internet. Pour que les machines puissent communiquer sur le même réseau, elles doivent respecter certaines normes matérielles et logicielles qui définissent un protocole. Il existe plusieurs protocoles de communication sur réseau, mais le protocole Ethernet TCP-IP, support de l'Internet, est le standard actuel.

Grâce aux réseaux, la micro-informatique, qui permet une utilisation individuelle indépendante, devient complémentaire des gros systèmes et permet aussi de répondre aux besoins des groupes d'utilisateurs.

OBJECTIFS

- Distinguer les concepts d'information et d'informatique.
- Connaître la structure générale d'un ordinateur et le rôle des principaux constituants.
- Comprendre la notion de codage et les différents types d'information.
- Définir les différentes catégories de logiciel.
- Distinguer les différents types de systèmes informatiques.
- Comprendre la notion de réseau.

GESTION DES DONNEES MEDICALES

On distingue classiquement l'information au sens usuel de renseignement, et la notion de donnée, information codée et stockée sur un support informatique en vue d'un traitement ultérieur. L'utilisation des données en informatique médicale tente de répondre à deux objectifs plus ou moins contradictoires :

- rester proche de la structure naturelle de l'information ;
- adopter la représentation informatique la plus efficace.

Ainsi, deux problèmes doivent être résolus :

- comment organiser les informations de façon à obtenir le système le plus efficace et le plus informatif : il s'agit d'un problème de **structuration** de l'information auquel répondent les **systèmes informatiques de gestion de données** ;
- comment représenter les informations afin d'en conserver le maximum de richesse sans s'interdire les possibilités de traitement automatique de l'information : il s'agit d'un problème de **standardisation du langage médical**.

Leur maîtrise permet d'envisager l'**informatisation du dossier du malade**. Par contre, en l'absence de structuration et d'organisation préalables de l'information en vue de son utilisation ultérieure, on est confronté à ce qu'on appelle un cimetière de données.

GESTION INFORMATIQUE DES DONNEES

Une structure de données correspond à une manière d'organiser et de représenter les données. Les deux types de renseignements contenus dans une structure de données sont les données proprement dites et les liens qui peuvent exister entre elles, formalisés par leur organisation. L'organisation de ces données en informatique est essentiellement celui de leur stockage et de leur accès sur une mémoire secondaire. Deux classes de systèmes peuvent être utilisées : les fichiers et les bases de données.

FICHIERS

On appelle **fichier** un ensemble de données organisées en vue d'une application déterminée. En fait, tout le monde fait des fichiers sans le savoir : ainsi, un répertoire téléphonique, un carnet d'adresses, un dictionnaire sont des fichiers.

En informatique, le fichier désigne une collection d'informations regroupées sous un

seul nom logique sur une mémoire secondaire (bande magnétique ou disque). Toutefois, un fichier informatique peut contenir un programme, du texte libre ou des données.

Les **fichiers de données** contiennent des informations de même nature (un fichier est un ensemble de fiches de même type) et surtout disposent d'une **structure interne**. Cette structure, ensemble de relations entre les différents éléments, permet l'exploitation des informations. L'ordinateur est une machine, sans connaissance du monde réel, qui ne peut s'appuyer sur un contexte pour manipuler une information. La structuration est une modélisation du monde réel afin que l'ordinateur puisse traiter correctement les éléments auxquels nous nous intéressons : noms, tailles, adresses ou autre. L'ordinateur ne sait pas que la chaîne de caractères "01.49.09.50.00" désigne un numéro de téléphone et pas un prénom, mais la structuration du fichier lui indique le format externe de cette donnée et donc la façon de la manipuler.

ORGANISATION D'UN FICHER DE DONNEES

Les entités auxquelles on s'intéresse sont décrites par un certain nombre de caractéristiques, analogues pour tous les éléments d'un fichier, les entités se distinguant par les valeurs qui sont affectées à ces caractéristiques. Par exemple, des malades seront tous représentés par leur nom, leur prénom, leur date de naissance et leur sexe, seules changeant les valeurs de ces caractéristiques pour chaque individu.

On appelle **enregistrement**, article ou fiche, l'ensemble des informations décrivant une entité. Les caractéristiques ou attributs sont appelés rubriques ou **champs** et peuvent recevoir des valeurs, appelées **occurrences** d'enregistrement ou réalisations. Afin d'optimiser la gestion informatique des rubriques, les champs sont généralement définis par leur nom, le type de donnée qu'ils vont contenir (texte, nombre, date voire image) et leur taille maximale.

Toutes ces informations, après manipulation en mémoire centrale, sont enregistrées pour mémorisation sur une mémoire secondaire et le problème est dès lors de récupérer ultérieurement cette information pour la consulter ou la modifier. Plusieurs modes d'accès à l'information sont possibles et vont influencer sur les facilités de consultation et de mise à jour.

Accès séquentiel :

Soit un fichier de malades enregistré sur une bande magnétique : les informations (fiches et rubriques) sont écrites les unes à la suite des autres :

nom1-prénom1-age1—nom2-prénom2-age2—nom3...

La recherche d'un malade par son nom ne peut se faire qu'en lisant séquentiellement tous les enregistrements le précédant, ce qui peut être très long s'il y a beaucoup d'enregistrements.

Accès direct :

Sur les disques et les disquettes, les informations sont enregistrées sur des pistes concentriques, partagées en secteurs. Chaque enregistrement a une adresse formée d'un numéro de piste et d'un numéro de secteur. On peut donc positionner directement la tête de lecture sur la piste puis lire séquentiellement le secteur sans être obligé de lire tous les enregistrements des pistes précédentes. On parle alors d'organisation directe et d'accès direct.

L'utilisateur n'a pas à connaître l'adresse physique de l'enregistrement (numéro de piste et position sur la piste) mais l'adresse logique (numéro d'ordre de l'enregistrement). Toutefois, si on veut retrouver un individu de nom donné sans connaître son numéro d'ordre, il faudra lire séquentiellement tous les enregistrements précédents.

Accès indexé :

En effet, le problème est qu'on ne connaît pas toujours le numéro d'ordre de l'individu recherché, à moins d'avoir la liste complète et à jour des enregistrements. C'est dire l'intérêt des **index** qui sont des tables de correspondance indiquant en face de la valeur du critère de recherche de chaque enregistrement (par exemple, le nom) le numéro d'ordre de cet enregistrement, de la même façon que l'index d'un livre indique à quelle page apparaît tel mot. Le critère de recherche est appelé **clé** d'index, il doit permettre d'identifier de façon unique (idéalement) un enregistrement. La gestion de l'index est normalement assurée par le logiciel de gestion de données. L'utilisateur ne voit que le fichier principal et la clé, il demande "lire enregistrement de clé "Martin"", le système récupérant le numéro d'enregistrement dans la table d'index pour accéder à cet enregistrement.

La clé d'index peut être simple ou composée de plusieurs critères (par exemple, nom et prénom) afin d'être plus discriminante. L'index peut être unique ou associé à d'autres index (on parle d'index primaire ou maître et d'index secondaires), afin de permettre un accès rapide sur d'autres clés (par exemple l'adresse ou le diagnostic). Le choix de la ou des clé(s) est du ressort du programmeur et suppose une réflexion préalable sur l'utilisation des données, comme d'ailleurs bien sûr tout le processus d'informatisation.

REALISATION D'UNE GESTION INFORMATIQUE DE FICHES

La programmation d'un tel système est longue et complexe. Il est donc plus simple d'utiliser un progiciel ad hoc dit Système de Gestion de Fichiers qui fournit les fonctionnalités nécessaires à la gestion aisée des informations. Ce gestionnaire de fichiers permet :

- de déclarer ou de redéfinir la structure des enregistrements, c'est-à-dire le nom, le type et la taille des diverses rubriques ;
- de saisir, modifier, ajouter des données ou de les supprimer ;
- de déclarer des clés d'index ou de trier le fichier ;
- de retrouver des données répondant à des critères plus ou moins complexes ;
- d'éditer ou d'imprimer le fichier, en totalité ou partie, sous une présentation variable ;
- de créer des masques facilitant la saisie à l'écran.

L'interface peut reposer sur l'emploi d'un langage de commande (en mode interactif ou programmé) ou sur un système de menus (mode assisté ou d'interrogation par l'exemple).

BASES DE DONNEES

Il est fréquent que les mêmes données soient dupliquées en totalité ou en partie dans plusieurs fichiers indépendants. Il en résulte une perte de place sur les supports physiques et

des difficultés de mise à jour : certaines fiches sont mises à jour plus souvent que d'autres et des données deviennent périmées ou incohérentes. D'autre part, l'enregistrement des données sous forme de fichiers simples ne permet pas de prendre en compte efficacement certaines relations entre les informations.

TYPES DE RELATION

Soit un fichier regroupant les données de malades ayant subi une intervention chirurgicale, chaque malade étant décrit par son état civil, son diagnostic, la date, le type et le résultat de l'intervention.

Or, il peut arriver qu'un malade subisse deux interventions. Que faire dans ce cas ? Ne pas tenir compte d'une intervention sur les deux semble un peu léger. Réviser la structure pour créer les rubriques permettant de décrire la deuxième intervention est pénalisant en terme de place sur le support car de nombreux patients n'auront subi qu'une seule opération, tandis qu'il se trouvera bien un malade qui aura subi trois opérations. Créer deux enregistrements pour ce malade fait perdre peu de place mais masque le fait qu'il s'agit de la même personne et introduit un risque d'incohérence.

Cette situation peut se représenter sous la forme d'une **relation 1:N**, c'est-à-dire pour un malade, on peut avoir N interventions. La solution consiste à rendre indépendants les malades et les interventions en éclatant le fichier unique en deux fichiers, un décrivant les malades, l'autre les interventions. Pour rétablir le lien entre le malade et ses interventions, il faut rajouter dans la fiche intervention une référence vers le malade, appelée un pointeur. Ainsi chaque malade peut avoir autant d'enregistrements d'interventions que nécessaire et il reste aisé de trouver les malades correspondant à tel type d'intervention.

Chaque pointeur est un numéro (le numéro de l'enregistrement du malade) n'occupant que peu de place, on a donc amélioré l'organisation tout en évitant une redondance trop importante, chaque information originelle n'étant enregistrée qu'une fois. En fait, on emploie généralement comme pointeur, non pas le numéro de l'enregistrement, mais une clé d'identification de l'enregistrement, comme le nom du malade. On obtient alors un pointeur logique qui, bien que plus volumineux en mémoire, permet d'éviter les problèmes liés à une réorganisation du fichier principal avec changement de l'ordre des enregistrements.

Soit une autre situation où il s'agit de gérer l'inscription d'étudiants à des cours, les étudiants étant inscrits au fur et à mesure de leur arrivée. L'enregistrement "étudiant" a comme rubriques, outre des caractéristiques d'état civil, des inscriptions à des cours : cours 1, cours 2, ... pouvant prendre les valeurs : Anatomie, Biochimie, etc..., chaque cours étant en outre décrit par son horaire, la salle, le professeur et le nombre maximum d'étudiants.

Comment rechercher tous les étudiants inscrits en Anatomie ou vérifier rapidement qu'il reste des places en Biochimie ? Comment procéder à un changement de salle pour un cours ? On peut certes balayer tout le fichier séquentiellement mais c'est long. De même, si un cours doit être modifié, il faut faire la modification pour tous les étudiants inscrits à ce cours.

La situation est ici plus complexe que dans l'exemple précédent puisqu'il s'agit d'une **relation de type N:M**, en effet un étudiant peut s'inscrire à plusieurs cours (M) mais un cours comporte plusieurs étudiants (N). La solution consiste à rendre indépendants les étudiants et les cours en éclatant le fichier précédent en trois : un fichier d'étudiants, un fichier de cours et un fichier établissant la correspondance entre un étudiant et un cours. Dans chaque enregistrement du fichier "étudiant", on retrouve l'état civil de celui-ci, dans le fichier "cours", on trouve la description du cours (numéro de salle, professeur, horaire, nombre maximum d'étudiants). Le fichier de correspondance ne comporte que deux rubriques qui sont des pointeurs : un numéro d'étudiant et un numéro de cours. Il comporte autant d'enregistrements mentionnant "Anatomie" qu'il y a d'étudiants inscrits à ce cours et autant d'enregistrements contenant "Martin" que le nombre de cours auxquels celui-ci s'est inscrit.

Bien entendu, ces questions ayant trait à l'exploitation des données doivent être posées avant de constituer le fichier, lors de la phase d'**analyse** de l'application.

OBJECTIFS D'UNE BASE DE DONNEES

La solution générale consiste à organiser les fichiers en **bases de données** qui regroupent de grands ensembles de données interdépendantes, selon les **critères** suivants :

- support informatique ;
- absence de répétition inutile ;
- partage et utilisation des données par des applications ou des utilisateurs distincts ;
- évolution indépendante des données et des applications ;
- protection et contrôle de l'accès aux données.

L'organisation et la gestion de ces bases de données, complexes, sont assurées par un ensemble de programmes rassemblés sous le terme de **SGBD** (Système de Gestion de Base de Données, Data Base Management System ou DBMS en anglais).

CONCEPTION D'UNE BASE DE DONNEES

L'objectif d'une base de données est de décrire sur un support informatique une partie du monde réel pour répondre aux besoins de certains utilisateurs. Ceci pose des problèmes de technique informatique, de modélisation de l'information et de mise à disposition de celle-ci. Ces trois aspects sont pris en compte par une conception en trois niveaux ou schémas.

Le **niveau interne** correspond à l'organisation physique et au stockage de la base de données : volumes de stockage (disques), modes d'organisation et modes d'accès des fichiers. Il concerne l'informaticien et non l'utilisateur final.

Le **niveau global ou conceptuel** correspond à la modélisation du monde réel. Il permet de décrire au mieux les différentes entités, leurs attributs et leurs relations. Il repose sur un modèle logique des données.

Le **niveau externe** concerne les applications ou les utilisateurs et regroupe un ensemble de "vues". Chacune de ces vues correspond à un ensemble structuré de données déduit du niveau conceptuel (par extraction ou transformation) et constitue une base virtuelle sans

existence physique. Elle représente ce qu'un utilisateur connaît de la base de données, cet ensemble peut varier d'un utilisateur à l'autre selon les besoins et les autorisations d'accès.

Chaque niveau répond à des objectifs différents :

- l'**optimisation** s'appuie sur le niveau interne et vise à réduire le temps d'accès en répartissant les fichiers au mieux sur les disques.
- le niveau conceptuel prend en charge la **non-redondance** et l'**intégrité des données**. L'intégrité statique concerne soit la validité intrinsèque d'une donnée (un âge est un entier entre 0 et 120), soit la cohérence des données entre elles (si sexe=M alors grossesse=non), soit des contraintes strictes (le nombre de rendez-vous de consultation ne peut excéder le nombre de créneaux libres le jour du rendez-vous). Elle comprend également le traitement des valeurs manquantes qui ne doivent pas être traitées comme une valeur "nulle". L'intégrité dynamique exprime les contraintes portant sur le passage d'un état cohérent à un autre (un âge ne peut qu'augmenter).
- la **confidentialité** est assurée au niveau externe. Une "vue-utilisateur" ne laisse voir que les éléments dont l'utilisateur a l'usage et qu'il a le droit d'utiliser en précisant le type de droit (lecture seule, lecture et écriture).

Cette organisation en niveaux interfacés assure au mieux l'indépendance entre les données et les applications qui les traitent et permet aussi l'évolution de la base quel que soit le système informatique sous-jacent, également modifiable.

SYSTEME DE GESTION DE BASE DE DONNEES

Pour répondre à ces multiples objectifs, le SGBD fournit plusieurs outils :

- un langage de description des données (**LDD**) permet de construire la base, qu'il s'agisse de l'implantation physique ou de la description logique : nom et type des données, contraintes d'intégrité ;
- un langage de manipulation des données (**LMD**) regroupe un ensemble de procédures assurant l'utilisation de la base : saisie d'informations, consultation, mise à jour, ajout ou suppression ;
- des utilitaires d'administration permettant de maintenir la base : gestion des utilisateurs, comptabilisation des accès et des opérations, sauvegarde, récupération des accidents.
- souvent, un langage dit de 4ème génération (L4G) permet au non spécialiste de développer une application. Les L4G sont dotés de générateurs de rapports ou de graphiques, d'éditeurs et de gestionnaires de fenêtres afin de réaliser des applications conviviales.

Ces langages sont de plus en plus construits autour de **SQL** (Structured Query Language, langage structuré d'interrogation) défini par IBM à la fin des années 70 et qui regroupe les fonctionnalités d'un LDD et d'un LMD.

Les SGBD, initialement développés sur les gros systèmes informatiques (par exemple, le SGBD Oracle) sont maintenant utilisables sur les micro-ordinateurs.

MODELES LOGIQUES DE BASES DE DONNEES

Les SGBD sont bâtis selon un modèle logique d'architecture des informations, modèle dont il existe différents types (hiérarchique, réseau). Le plus utilisé actuellement est le **modèle relationnel**, proposé en 1970 par Codd. Il repose sur la notion mathématique de relation, dérivée de la théorie des ensembles. Une relation est un tableau (ou table) qui a p lignes et n colonnes. Chaque ligne est une entité, qui présente n attributs, on dit que c'est un n-uplet. Chaque attribut est défini sur un domaine de valeurs. On peut établir des liens de type 1:n ou n:m entre tables grâce à une opération de jointure sur une clé ou une rubrique commune à deux ou plusieurs tables. La table résultant d'une opération (restriction, projection ou jointure) peut à son tour être manipulée comme une table originelle, il est donc possible d'enchaîner et de combiner les opérations. Ce modèle est apprécié pour la (relative) simplicité et la puissance de son langage d'interrogation, généralement SQL. Des modèles plus récents de SGBD (orientés objets, sémantiques) autorisent une plus grande richesse d'expression mais ne sont pas encore d'usage courant.

LANGAGE MEDICAL

Le langage médical est caractérisé par un vocabulaire extrêmement riche et difficile à manipuler. Il n'y a pas de consensus établi sur la définition des termes employés. Les synonymes sont nombreux (plusieurs termes désignant le même objet) tandis que le même terme peut avoir plusieurs significations selon l'auteur ou le contexte (polysémie). Les textes médicaux sont donc souvent imprécis, ambigus d'autant qu'ils font un large usage d'abréviations et d'acronymes. Pour permettre une description et une communication efficaces et dépourvues d'ambiguïté, a fortiori un traitement automatique, un minimum de standardisation du langage est nécessaire. Une première étape a consisté en un dénombrement des concepts sous forme de nomenclature alors qu'une autre phase concernait le rassemblement des concepts de sens proche sous forme de classification. Le niveau de formalisation de l'information par l'utilisateur peut être très variable, depuis de petites unités d'information sous forme de codes jusqu'au langage naturel, en passant par l'emploi de mots clés, de phrases clés voire d'un véritable langage artificiel.

NOMENCLATURES ET CLASSIFICATIONS

TERMINOLOGIE

Une **nomenclature** est une liste des éléments d'une collection de termes. Il n'y a aucun agencement particulier des termes ni de définition explicite, l'objectif étant l'exhaustivité.

Un **thesaurus** est une collection organisée des termes d'un vocabulaire, ici les termes techniques utilisés en médecine, représentés de façon normalisée par des descripteurs ou des mots clés. Chaque terme est ordonné avec une place réservée dont la référence (code) est

alphabétique ou numérique. Les références doivent être distinctes pour chaque descripteur.

Une **classification** consiste à partitionner l'ensemble des objets pour les distribuer en classes et sous-classes constituées d'éléments de plus en plus semblables, ici les termes de signification proche. Il s'agit d'un thesaurus doté d'une structure d'arbre et chaque élément ne peut appartenir qu'à une seule classe. La structure de la classification dépend de l'objectif poursuivi par son concepteur.

Un **codage** est la traduction d'un message selon un code, généralement numérique ou alpha-numérique, en vue de sa transmission ou de son traitement. Le codage doit être biunivoque pour éviter toute ambiguïté.

Thesaurus et classifications permettent de traduire un message dans un vocabulaire normalisé. Lors d'une transmission d'information, l'émetteur code le message en fonction d'un langage et du contexte, l'interprétation correcte par le récepteur suppose l'emploi du même langage et la connaissance du contexte. Or le contexte conditionne le codage, ainsi, dans le cas d'un malade hospitalisé pour chimiothérapie d'un cancer ayant développé une aplasie, le dossier serait codé selon l'étiologie cancéreuse par un épidémiologiste mais sous la rubrique aplasie si on s'intéresse à la charge en soins.

CARACTERISTIQUES DES CLASSIFICATIONS ET POUVOIR SEMANTIQUE

Les principales caractéristiques des classifications permettant d'évaluer leur capacités d'expression sont la nature du principe de classement, la prise en compte d'axes multiples et les types des relations exprimées.

Une classification correspond à une catégorisation récursive du domaine selon un critère qui s'applique à l'ensemble des éléments d'une classe, les critères se succédant de classe en sous-classe par ordre d'importance décroissante. Le **lien sémantique** qui préside à la catégorisation peut être un lien d'affectation (type est-un, par exemple, une rubéole est une maladie infectieuse) ou de partition (type fait-partie-de, par exemple, le genou fait partie du membre inférieur).

Une classification **monoaxiale** répartit en plusieurs classes disjointes l'ensemble des objets et revient à construire une hiérarchie de classes à partir d'une racine unique et commune. Les classes d'un niveau doivent couvrir l'ensemble du domaine de ce niveau (exhaustivité) sans se recouvrir (exclusivité) afin qu'un objet trouve une place et une seule.

Dans la pratique, il est très difficile de répartir les objets du domaine selon un seul critère, ce qui a entraîné le développement de classifications multiaxiales. Les classifications **multiaxiales** ou à facettes sont modulaires et combinent des termes appartenant à des systèmes différents, eux-mêmes organisés de façon hiérarchique. Ainsi, une sémantique de juxtaposition s'ajoute à la sémantique de catégorisation.

Cependant, même la juxtaposition de termes ne suffit pas à exprimer la diversité des liens sémantiques existant entre les termes médicaux (relation de causalité, d'association, d'évolution, etc...). Aussi ces liens sont-ils explicités par des **connecteurs** (équivalents des groupes verbaux : est un, a, traite, est traité par, cause, est causé par, etc...) dans certaines

classifications ou dans les langages artificiels. En outre, à côté des termes (les mots) et des connecteurs (les verbes), certains systèmes ajoutent des modificateurs, faisant fonction d'adjectif, venant préciser le degré de certitude, l'évolutivité ou l'intensité.

EXEMPLES DE CLASSIFICATION

Il existe de multiples nomenclatures et classifications, certaines d'usage universel et international tandis que d'autres sont limitées à une spécialité voire à un service ; chacune répondant (plus ou moins bien) à un besoin particulier ; nous nous limiterons donc aux plus représentatives.

CIM ET DERIVEES

La Classification Internationale des **Maladies** (CIM ou ICD en anglais) de l'Organisation Mondiale de la Santé (OMS ou WHO) a été originellement conçue pour coder les causes de décès dans une optique épidémiologique. Elle est maintenant également utilisée pour l'évaluation des soins médicaux et l'indexation des dossiers. Elle bénéficie d'une remise à niveau régulière, la version la plus récente étant la 10^{ème} révision (CIM-10, publiée en 1993). Il s'agit d'une classification **monoaxiale** avec 21 chapitres principaux dont 17 concernent des maladies et quatre concernent les signes et résultats anormaux, les causes de traumatismes, d'empoisonnement ou de morbidité, l'état de santé et les facteurs de recours aux soins. Les catégories de maladies sont définies en fonction d'un caractère commun qui peut être l'étiologie (1 = Maladies infectieuses, lettres A et B), la topographie (9 = maladies de l'appareil circulatoire, lettre I), la physiologie (15 = Grossesse et accouchement, lettre O) ou la pathologie (II = Tumeurs). La classification aboutit par subdivisions successives à un code à 3 caractères (une lettre correspondant au chapitre puis 2 chiffres) pour les maladies définies à un niveau général, décliné par l'ajout d'un quatrième chiffre (derrière un point) pour désigner les diagnostics précis et les formes cliniques ; le sous-code 9 désignant l'absence de précision (SAI = sans autre indication) et le sous-code 8 les autres formes non précédemment définies. Dans certains cas, un cinquième chiffre a été rajouté afin d'améliorer la finesse de description. La CIM-10 a introduit la notion de troubles iatrogènes. Elle compte au total 80.000 termes.

La contrainte d'avoir un seul arbre hiérarchique implique qu'une entité pathologique soit représentée une seule fois dans la classification, ce qui pose des difficultés. Ainsi les tumeurs sont extraites de leur chapitre d'appareil et regroupées dans un chapitre spécial. Parfois cependant, une même maladie peut apparaître en deux places distinctes (avec deux codes). C'est le cas lorsqu'une maladie appartient à un processus pathologique initial général (code associé à une dague), par exemple la tuberculose, et correspond à des manifestations localisées à un appareil (code associé à un astérisque), par exemple une tuberculose rachidienne. D'autre part, le principe de différenciation n'est pas constant.

La classification de l'OMS sert en France au codage des **causes de décès** ainsi qu'au regroupement des séjours hospitaliers en groupes homogènes de malades (programme de

médicalisation du système d'information ou **PMSI**) dont le but est l'analyse médico-économique de l'activité hospitalière. La CIM-9 ne permettant pas de décrire les détails cliniques d'un patient, ni les notions d'évolution ou de physiopathologie, une extension clinique a été proposée aux Etats-Unis sous le nom d'ICD-9-CM. En Europe, la classification HCIMO (Classification Internationale des Maladies et Opérations, adaptation hospitalière) est une adaptation et une extension du code ICD-9-CM.

CDAM ET NGAP

Les classifications des **actes** et des procédures sont plus variables car elles répondent à différents objectifs.

Le catalogue des actes médicaux par nature (CDAM) a été préparé par le ministère de la Santé français dans le cadre du **PMSI** afin de permettre le codage des actes effectués au cours d'un séjour hospitalier. Il est proche de l'HCIMO par les actes qui y sont définis et codés mais en a perdu la structure hiérarchisée. Il se subdivise en sept champs, correspondant aux grandes catégories d'actes, représentés par des lettres grecques, par exemple, alpha pour les actes diagnostiques et thérapeutiques ou gamma pour l'imagerie. Le code acte est formé de 4 caractères (une lettre et trois chiffres). A chaque acte est associé un indice numérique, l'**indice de coût relatif** (ICR) chiffrant la consommation de ressources (humaines, matérielles, investissement) liées à la réalisation de l'acte. De plus, certains actes sont dits classants et repérés par un "Y", car ils sont les témoins d'une prise en charge particulière et à ce titre, ils doivent figurer dans le résumé de sortie établi pour chaque séjour dans le cadre du PMSI.

La nomenclature générale des actes professionnels (NGAP) décrit des **prestations** et y associe des **tarifs**, car elle est utilisée pour le remboursement des actes par la Sécurité Sociale qui l'a établie. Une prestation est un ensemble d'actes liés dans leur réalisation ou d'objectif identique. A chaque prestation est attachée une lettre clé selon la discipline (C pour une consultation, K pour la chirurgie, B pour la biologie, Z pour la radiologie), pondérée par un coefficient correspondant à un nombre d'unités financières. Par exemple, une appendicectomie est cotée "K50" pour la chirurgie et "AK 25" pour l'anesthésie soit au total "K75". La valeur monétaire du coefficient associé à une lettre fait l'objet de négociations régulières.

Il est prévu de constituer une nomenclature commune des actes (NCA) qui remplirait les fonctions des deux nomenclatures précédentes.

SNOMED

SNOMED (Systematized Nomenclature of Medicine) combine une nomenclature de plus de 50.000 termes et une classification **multiaxiale** comportant à l'origine 7 axes : topographie, morphologie, étiologie, altération fonctionnelle, nosologie, actes médicaux. La 3ème édition compte désormais 200.000 termes et 11 axes. Chaque axe est défini par une lettre (par exemple, T pour topographie, E pour étiologie) et organisé de façon hiérarchique,

chaque élément étant associé à un code numérique à 4 ou 5 chiffres. Ainsi un diagnostic est traduit par plus d'un élément signifiant, mais chaque axe ne doit pas être obligatoirement validé. Par exemple, la juxtaposition : "T2856 (lobe supérieur du poumon gauche) / M4100 (inflammation) / F0300 (fièvre) / E2012 (pneumocoque)" correspond à la phrase "Pneumonie fébrile à pneumocoque du lobe supérieur gauche". L'ajout de connecteurs concernant notamment les liens de causalité permet de décrire un fait complexe en plusieurs phrases. SNOMED est largement utilisé car précis, cependant ce modèle pose encore des problèmes : les termes des différents axes ne sont pas complètement indépendants entre eux, l'axe Maladie fait souvent double emploi, certains concepts peuvent apparaître dans plusieurs axes.

MESH

MeSH (Medical Subject Headings) sert à indexer, cataloguer et retrouver des **références de bibliographie** dans le domaine de la Santé. Il a été conçu à la National Library of Medicine (NLM) aux Etats-Unis comme support de l'Index Medicus, répertoire des principales publications scientifiques, et est utilisé par les systèmes de recherche bibliographique Medlars et Medline. MeSH est organisé en deux parties : une liste alphabétique de termes (lexique) et une structure multiaxiale. Les 200.000 termes du lexique sont distribués selon 15 axes, allant de l'anatomie à la géographie. Les termes équivalents sont rapportés à celui des 20.000 termes principaux (descripteurs) qui exprime le mieux le concept, termes auxquels sont associés un code alphanumérique. Les descripteurs s'organisent selon une structure hiérarchique. En outre, des connecteurs permettant des références explicites entre termes expriment les relations de synonymie, de voisinage ou d'association tandis que des qualificatifs permettent de considérer les différentes facettes d'un concept (par exemple : tuberculose/traitement). Ce système qui indexe actuellement plusieurs centaines de milliers de références est mis à jour régulièrement pour suivre l'évolution des connaissances.

TRANSCODAGE

La multiplicité des classifications répond à la diversité des informations à traduire. Mais parfois, pour représenter la même réalité, les acteurs de la santé utilisent différents systèmes de classification. Ainsi, pour indexer précisément les dossiers de ces patients en fonction des actes réalisés, un chirurgien utilisera une classification ou un système de mots clés spécifiques de sa spécialité, alors que le codage des actes dans le cadre du PMSI utilise le CDAM, tandis que les organismes de Sécurité Sociale emploient la nomenclature NGAP. Cette multiplicité perturbe la vision de l'activité quand elle ne dégrade pas l'information, notamment lors de la communication entre systèmes différents.

Certaines classifications de granularité très fine permettent un transcodage dans d'autres classifications, créant ainsi des passerelles. La qualité de ces passerelles dépend du type et de la congruence des classifications ainsi reliées. Par exemple le transcodage de CIM9 en Snomed ne permet pas toujours une transcription explicite des concepts en raison de la

différence de structuration des deux systèmes. Les transcodages exposent ainsi souvent à une perte d'information mais, dans certaines situations, ils sont impossibles.

LES SYSTEMES EN DEVELOPPEMENT

META-SYSTEMES

L'absence de terminologie médicale acceptée par tous et l'existence de multiples systèmes de classification est à l'origine de plusieurs projets. UMLS (Unified Medical Language System), développé à la NLM, a comme objectif de construire un système automatisé et intelligent qui comprenne les termes biomédicaux et leurs relations, permettant ainsi aux utilisateurs de retrouver l'information de façon transparente à partir de sources multiples et hétérogènes. La structure de ce système en cours de développement comprend un méta-thesaurus, un réseau sémantique et un répertoire des sources. Le système est initialement fondé sur le MeSH, SNOMED et la CIM-9. En Europe, le projet GALEN vise à construire une représentation des concepts du domaine médical qui soit indépendante des thesaurus mais aussi des langues nationales.

LANGAGE NATUREL

Le développement des techniques informatiques dites d'intelligence artificielle pourrait faire croire que les efforts de formalisation du langage médical sont à terme vains, dès lors qu'un système informatique pourra comprendre le langage naturel.

Le problème est complexe en raison de la très grande ambiguïté des langues naturelles ou professionnelles. Ces ambiguïtés sont notamment apparues lors des essais de traduction automatique. Les systèmes d'analyse de texte doivent enchaîner des étapes d'analyse morphologique et lexicale (orthographe), syntaxique (grammaire), sémantique (définitions et sens des mots), pragmatique (connaissance du contexte). En effet, le sens d'une phrase ne dépend pas seulement de la forme de la phrase, mais même ce simple niveau syntaxique peut être source de confusions.

En fait, à court et à moyen terme, il apparaît plus réaliste d'attaquer le problème par le bas, c'est-à-dire de développer les systèmes de langage artificiel en étendant les systèmes de classifications existants. En outre, demander au médecin de structurer son langage, c'est lui demander d'explicitier son raisonnement et de mieux réfléchir à sa pratique, exercices pour le moins bénéfiques.

INFORMATISATION DU DOSSIER DU MALADE

Le dossier du malade est un élément fondamental de la prise en charge. C'est d'abord un outil de mémorisation, de communication et le support du processus de décision. Différents acteurs (médecins, infirmières, gestionnaires) sont concernés par ce dossier et en ont des vues

particulières. En fait, il existe plusieurs dossiers partiels : administratif, médical, infirmier, social, qui sont l'objet d'une utilisation individuelle ou collective. Ce dossier, constamment mis à jour, répond en effet à plusieurs besoins :

- outil de suivi du malade, particulièrement important dans le cas d'une affection chronique ;
- outil de synthèse de la démarche de soins ;
- outil médico-légal (conservation minimale légale de 20 ans dans les établissements publics, portée à 70 ans pour les dossiers de pédiatrie) ;
- outil de communication entre les différents intervenants ;
- outil de recherche clinique, d'étude épidémiologique ou d'évaluation des soins, s'il est standardisé dans le cadre d'un protocole ;
- outil de gestion hospitalière ;
- outil d'enseignement.

Cet instrument fondamental est ainsi partie intégrante de l'acte de soins, à tel point qu'en Amérique du Nord, la bonne tenue et gestion des dossiers figure depuis longtemps dans les conditions nécessaires à l'accréditation d'un établissement hospitalier.

Le dossier du malade médical (au sens large) comprend en général :

- l'identification du malade et ses coordonnées administratives ;
- le résultat de l'examen clinique initial et des examens successifs ;
- les résultats d'examens para-cliniques ;
- le ou les éventuels compte-rendus opératoires ;
- les prescriptions d'ordre thérapeutique ;
- le dossier de soins infirmiers comportant les données de surveillance et les consignes de transmission ;
- le compte-rendu d'hospitalisation ;
- les prescriptions établies à la sortie.

C'est dire qu'il contient à la fois des données "brutes" (examen clinique, résultat biologique) et des données interprétées (démarche diagnostique, compte-rendu), consignées le plus souvent dans un ordre chronologique linéaire.

Les dossiers "papier" classiques ont le mérite d'être d'usage général, de manipulation simple mais leurs défauts sont connus, notamment :

- absence de signes ou de résultats d'examens, due à une non-réalisation, à un oubli de transcription ou à une normalité de résultat, ces données manquantes venant compromettre les études comparatives dont découle pour partie la connaissance médicale ;
- hétérogénéité de présentation ;
- manque de standardisation du langage employé.

OBJECTIFS DE L'INFORMATISATION

L'informatisation du dossier par la mise sur ordinateur, la structuration et la standardisation qu'elle suppose devrait :

- améliorer le stockage, la disponibilité et la communication des informations ;

- améliorer la lisibilité des informations ;
- éviter les lacunes en systématisant le recueil ;
- permettre une saisie unique et un partage de l'information ;
- mettre en évidence l'évolutivité des informations ;
- rendre comparables les informations d'un patient à un autre ;
- intégrer des données d'origines diverses ou de nature hétérogène (signaux, images) ;
- faciliter l'emploi de systèmes d'aide à la décision ;
- aider au regroupement des données ;
- faciliter la formation ;
- améliorer la protection et la confidentialité des données.

Mais la mise en oeuvre est complexe : abondance et complexité des connaissances, modélisation des données et du langage médical, problèmes humains d'interface ou de sécurité, coût élevé des matériels et de la formation, évolution technologique incessante.

MODELISATION

La modélisation comprend la structuration du dossier et la représentation des éléments d'information le composant.

Plusieurs modes de structuration sont possibles :

- structure orientée selon la source : c'est l'organisation traditionnelle (examen-diagnostic-traitement), la plus utilisée, ce qui fait qu'elle est connue de tous, mais elle n'est pas stable dans le temps, est plus ou moins développée selon la spécialité du service et ne fait pas apparaître clairement la stratégie médicale ;
- structure orientée par problèmes : après un recueil de données de base, on liste tous les problèmes, médicaux ou autres, qui seront traités chacun selon un plan explicite comportant recherche d'informations et décisions ; cette structure favorise la mise en oeuvre de stratégies cohérentes (protocoles ou "guidelines") mais pose des problèmes de subjectivité et d'organisation des informations (indépendance ou intrication des problèmes) ;
- la recherche s'oriente vers la mise en place d'une structure informatique profonde distinguant les faits bruts (sujet - paramètre - valeur - date de l'observation) et les interprétations et pouvant être présentée à l'utilisateur selon différentes orientations.

Dans tous les cas, le dossier est généralement constitué d'un noyau de base regroupant les informations non répétitives (identité), d'un ou plusieurs dossiers satellites répétitifs correspondant à divers aspects (spécialités) ou épisodes (consultations, hospitalisations) de l'histoire du malade, pouvant eux-mêmes être décomposés en sous-dossiers, enfin des fichiers annexes correspondant aux différents thesaurus et nomenclatures employés.

La représentation des éléments d'information (symptôme, signe, résultat d'examen, compte-rendu) peut être plus ou moins facile et efficace selon leur nature :

- valeurs logiques (oui/non) ;
- valeurs numériques entières ou réelles ;
- valeurs qualitatives, correspondant à des choix dans des listes de valeurs, choix pouvant être

exclusifs ou non ;

- informations textuelles : leur saisie sous forme de texte libre n'offre comme avantages que la lisibilité et la disponibilité mais n'améliore ni leur homogénéité ni leur exhaustivité, enfin surtout leur exploitation (sélection de dossiers, statistiques) est très aléatoire ; au contraire leur codage au moyen d'une classification ou d'un langage artificiel autorise les traitements automatiques de ces données, au prix d'une relative simplification.

Afin d'améliorer la complétude de l'information, il est préférable de recourir à un dossier standardisé, qui couvre tout le sujet en énumérant a priori tous les items à enregistrer et en définissant a priori la forme des réponses, plutôt que d'avoir une seule rubrique fourre-tout de type commentaire en texte libre. En outre, le problème des données manquantes est crucial pour toute exploitation statistique et il est important de pouvoir distinguer une réponse négative d'une absence de réponse (ne sais pas, inconnu ou non-disponible).

MISE EN OEUVRE

La mise en œuvre peut faire appel à des SGBD généraux ou à des systèmes spécialisés de gestion de données médicales :

- les SGBD généraux doivent être considérés plus comme des outils de développement destinés au programmeur que comme des produits directement utilisables par le médecin ; ils sont généralement performants, ouverts sur l'extérieur mais parfois peu aptes à prendre en compte la dimension temporelle et évolutive des informations médicales ou la diversité de certaines informations (signaux, images) ; de nouveaux modèles de SGBD (modèles dits orientés objets) devraient mieux y répondre ;
- les logiciels spécialisés de gestion de données médicales offrent l'avantage d'une structure pré organisée (dossier de base et dossiers satellites), de l'intégration de fonctionnalités de bureautique médicale (éditions de comptes-rendus ou d'ordonnances, gestion de rendez-vous) ; en contrepartie, il s'agit souvent de systèmes développés par de petites équipes, avec des outils non standardisés, assez fermés et dont la pérennité est souvent préoccupante.

Dans tous les cas, la mise en place d'un système de gestion de dossiers implique un investissement important, en matériel, en formation et en maintenance (contrainte de disponibilité). L'acceptabilité du système par les utilisateurs conditionne largement son succès, celle-ci dépend directement d'une bonne **analyse préalable** des besoins, des tâches et des circuits de l'information. La qualité des informations enregistrées lui est également liée, ainsi la saisie de l'information doit se faire le plus près de la source, par ailleurs la personne qui utilise le système doit en retirer un avantage afin d'être directement intéressée à la qualité de son travail. La validité des données entrées en machine est en effet un point capital, en raison de leur possible influence sur le pronostic du malade, or trop souvent les données qui sortent d'un ordinateur apparaissent auréolées de certitude, alors que cette machine fonctionne selon le principe de l'auberge espagnole ("garbage in, garbage out").

INDICATIONS

L'expérience accumulée depuis plus de vingt ans a montré l'échec des systèmes généraux voulant tout faire (action médicale, logistique, administration) et les réussites partielles des systèmes à finalités limitées et bien définies. Parmi ceux-ci, on peut individualiser :

- dossier minimum : il s'agit soit d'un dossier minimal commun à tous les services d'un hôpital, ne mentionnant que les éléments majeurs du dossier, accessible notamment aux urgences, soit d'un résumé de sortie standardisé utilisé pour l'analyse d'activité médico-économique ;
- dossier de spécialité : la relative homogénéité d'une spécialité, tant du point de vue des malades, des examens pratiqués ou du langage employé, la limitation du domaine de connaissances à considérer ont permis le développement de dossiers informatisés dans plusieurs disciplines (diabétologie, néphrologie, réanimation), utilisés pour le suivi thérapeutique, l'épidémiologie ou l'analyse d'activité ;
- dossier général : dans ce cas, la diversité des pathologies, l'hétérogénéité des pratiques conduit généralement à ne considérer que les fonctions de bureautique médicale ou de codage minimal ;
- registre : il s'agit ici de dossiers très simplifiés, systématiques, destinés à alimenter des bases de données multicentriques voire nationales ou internationales dont l'objectif est essentiellement épidémiologique ;
- questionnaires d'essais thérapeutiques ou d'enquêtes épidémiologiques, très détaillés et précis, mais limités dans leur objectif et dans le temps, répondant à un protocole explicite.

OBJECTIFS

- Définir les notions de fichier, d'enregistrement, de rubrique, d'index.
- Comprendre les deux types de relation entre des informations.
- Distinguer les différents niveaux de conception d'une base de données.
- Connaître les objectifs et les principes d'un système de gestion de base de données.
- Définitions d'une nomenclature, d'un thesaurus, d'une classification.
- Comprendre le principe du codage et du décodage de l'information médicale.
- Connaître l'organisation et les indications de la CIM, l'intérêt des extensions.
- Connaître les caractéristiques du langage SNOMED.
- Connaître les fonctions du dossier patient.
- Comprendre les objectifs et les bénéfices attendus de l'informatisation du dossier.
- Comparer les possibilités du langage naturel et d'un langage artificiel.
- Quels sont les moyens et les difficultés de l'informatisation du dossier patient.

SYSTEMES D'INFORMATION

Par système d'information, on considère ici un système qui permet d'exécuter sur des informations, tout ou partie des actions suivantes : recueil, archivage, extraction, traitement, interprétation, réduction, évaluation, présentation, communication. Certains de ces systèmes intègrent des sous-systèmes d'informations ou communiquent avec d'autres systèmes. S'agissant de données médicales, individuelles ou collectives, la plus grande attention devra être portée à : la fiabilité de la technologie, la validité des données recueillies puis enregistrées ou transmises, la représentativité des données, la protection des données et le respect du secret médical.

TRAITEMENT DES SIGNAUX PHYSIOLOGIQUES

Les signaux physiologiques que fournissent les explorations fonctionnelles ou les enregistrements électriques : ECG, EEG, spirométrie, analyseurs de biochimie, sont constitués de suites de nombres : de plusieurs dizaines à plusieurs millions par seconde. Or le médecin a généralement besoin d'un petit nombre de paramètres caractéristiques du signal afin d'interpréter celui-ci. C'est l'objet du traitement d'effectuer les transformations rendant ces données directement utilisables pour la prise de décision.

TECHNIQUES DE TRAITEMENT

Le signal électrique analogique produit par un capteur est un signal continu, variant en fonction du temps, à 2 dimensions, sa fréquence et son intensité. Comme toute information, il doit être mis sous forme binaire pour être manipulable par un ordinateur, c'est l'opération de **conversion analogique-digitale** (ou numérisation) qui procède en trois étapes :

- le signal est d'abord découpé en segments de durées égales, c'est l'**échantillonnage** dont la fréquence doit être au moins double de la fréquence maximale du signal (théorème de Shannon) pour éviter toute perte d'information ;
- la hauteur de chaque segment est alors **quantifiée** (en prenant une valeur moyenne) ;
- cette valeur est ensuite codée sous forme **numérique** ; plus la longueur du mot binaire utilisé pour représenter la hauteur est grande, plus on peut définir de niveaux différents d'intensité du signal et donc plus la précision sera importante : 1 bit ne permet de coder que deux niveaux et correspond à un signal en tout ou rien (par exemple, froid ou chaud), 2 bits

autorisent 4 niveaux possibles alors qu'un octet correspond à 256 (2^8) niveaux différents possibles.

Au total, la **séquence de traitement** comporte quatre phases :

- **acquisition** du signal analogique par un capteur et **numérisation** par un convertisseur analogique-digital ;
- pré traitement simple visant à l'amélioration de la **qualité** du signal (extraction du signal sur le bruit, amplification, filtration) ;
- traitement **analytique** permettant l'extraction de paramètres, par exemple les complexes QRS d'un ECG, le plus souvent par des méthodes mathématiques ;
- **interprétation** des résultats.

Il n'y a pas de pré traitement universel, en effet, celui-ci dépend à la fois de la source des informations, de la nature des données initiales et de la transformation à effectuer.

APPLICATIONS

L'**intérêt de l'informatique** pour le traitement tient au fait que les techniques numériques de traitement de signaux sont plus puissantes et plus souples d'emploi que les techniques strictement électroniques (un programme est plus facile à modifier qu'un câblage). Le signal, une fois numérisé, est facile à stocker, à manipuler et à transmettre (télésurveillance). L'ordinateur permet d'enregistrer des signaux pendant des périodes éventuellement prolongées, d'en assurer le dépouillement automatique, de présenter les informations de façon variée, d'augmenter ainsi la capacité de perception du clinicien, de déclencher des alarmes ou d'établir une boucle de rétroaction, par exemple en pilotant une perfusion. Les applications de l'analyse des signaux sont nombreuses :

- l'analyse automatique de signaux ECG est une des applications les plus anciennes et les plus répandues. A partir du signal acquis sur plusieurs dérivations puis numérisé, le programme procède à un lissage par une méthode de moyenne mobile, puis à un calcul de la ligne de base et à un repérage du complexe QRS par analyse de la pente ; il fournit enfin la fréquence, la durée et l'amplitude des différentes ondes. Ces paramètres simples peuvent être fournis au médecin ou transmis à un programme d'aide à la décision qui indiquera la nature du trouble éventuel. Le signal, enregistré, peut être analysé et interprété à distance (transmission par réseau pour télésurveillance).
- l'analyse de l'EEG : les signaux EEG sont plus difficiles à interpréter ; l'analyse des potentiels évoqués a pour but d'étudier la réponse du cortex à différents stimuli et leur numérisation permet de synthétiser et de présenter les résultats sous forme de cartes d'activité cérébrale ; dans l'analyse des troubles du sommeil où les enregistrements durent plusieurs heures, les programmes permettent de repérer les phases du sommeil, de les quantifier et de sélectionner les passages pertinents pour une analyse manuelle ;
- La surveillance prolongée en ambulatoire avec des systèmes portables peut concerner l'ECG ou la mesure de la pression artérielle ;
- la surveillance en soins intensifs est caractérisée par un nombre élevé de paramètres et le

caractère parfois urgent d'une intervention correctrice ; les systèmes informatiques permettent d'intégrer des signaux d'origine multiple (ECG, pression, volume, débit, concentration), de les présenter sous une forme synthétique, de mémoriser et d'interpréter des variations sur des durées prolongées, de déclencher des alarmes avec plus de pertinence que des appareils isolés ; la multiplicité des sources a motivé le développement d'interfaces standard de communication ;

- les systèmes de rétrocontrôle fonctionnent selon le principe "perception-interprétation-rétroaction" ; pour par exemple adapter de façon automatique le débit d'une perfusion ou d'une pompe à insuline, les paramètres d'un appareil de ventilation ou de dialyse rénale, appareils auxquels il faut restituer un signal analogique par un convertisseur digital-analogique.

L'intégration des systèmes de traitement des signaux dans des systèmes d'information plus vastes, par exemple au niveau d'un hôpital, permet de les mettre en relation avec les données du dossier médical ou de communiquer avec les laboratoires de biologie afin de travailler sur des informations de meilleure qualité ou d'élaborer des décisions plus synthétiques.

SYSTEME D'INFORMATION EN IMAGERIE

Depuis les années 70, trois nouvelles techniques, basées sur le traitement informatique, ont bouleversé l'imagerie médicale : la tomodynamométrie (scanner), l'angiographie numérisée et de l'imagerie par résonance magnétique nucléaire. La diffusion croissante des systèmes informatiques a bénéficié également à la scintigraphie, à l'échographie, à l'endoscopie vidéo et à la radiologie conventionnelle qui sont devenues peu ou prou numériques par conversion d'images sources. Plus récemment est apparu le concept de système informatique dédié à l'imagerie.

NUMERISATION D'IMAGES

L'image est un signal ayant deux dimensions (2D) dans l'espace. Il y a donc nécessité d'un **codage spatial** de l'image qui est représentée par une matrice (grille) de $n \times m$ carrés élémentaires ou pixels (contraction de "picture element"). La précision augmente en fonction du nombre de pixels par surface : on passe de 128×128 pixels en scintigraphie à 1024×1024 en angiographie numérisée. Dans le cas d'images en trois dimensions (3D), on considère un tableau à trois dimensions dont chaque élément est un cube appelé voxel.

Il faut aussi traduire l'**intensité** lumineuse de chaque point par un nombre binaire lors d'une opération de quantification. La taille du mot binaire induit le nombre de graduations possibles ou résolution de contraste : un bit permet de coder noir/blanc quand un octet permet de coder 256 niveaux d'intensité lumineuse. Dans le cas d'une image monochrome, les 256 niveaux correspondent à 256 niveaux de gris. Pour une image en couleur, le même principe

s'applique à chacune des trois couleurs fondamentales (rouge, vert, bleu dans le codage RVB).

Certaines sources délivrent des images animées (angiographie) ainsi une numérisation en "temps réel" doit pouvoir générer 30 images par seconde. La **résolution temporelle** désigne le temps nécessaire pour créer une image. Une séquence d'images 2D nécessitera un tableau à 3 dimensions alors qu'une séquence d'images 3D nécessitera un tableau à 4 dimensions.

Lors d'une conversion analogique-digitale, le codage s'accompagne généralement d'une perte de l'information contenue dans l'image initiale, en raison du nombre limité de pixels et de niveaux distincts alors que le phénomène codé peut avoir une variation quasi-continue. Cette perte d'information est parfois théorique, lorsque l'information perdue est inaccessible en pratique à l'œil humain qui ne peut distinguer plus de 256 niveaux de gris sur une radiographie classique. Dans la pratique, les systèmes de numérisation utilisent des matrices carrées correspondant à des puissances de 2 (de 128x128 à 2048x2048) avec un codage d'intensité sur 8 à 24 bits.

SOURCES D'IMAGES

Elles sont multiples, mais on peut distinguer d'abord celles où les images sont obtenues directement sous forme numérique :

- la tomodensitométrie est basée sur le balayage axial du corps par un mince faisceau de rayons X dont l'atténuation est mesuré par un capteur. L'atténuation des rayons observés sur tous les points du plan de coupe permet la reconstruction d'une image par un ordinateur. Ces images sous forme de matrice de 256^2 à 512^2 pixels nécessitent de 0,5 à 2 secondes de temps de pose.
- l'imagerie par résonance magnétique nucléaire fournit des images morphologiques mais aussi fonctionnelles des organes. Les images, obtenues par reconstruction numérique, reposent sur l'excitation des atomes qui émettent un signal radio dont l'intensité et la durée dépendent des caractéristiques biologiques des tissus traversés.
- l'angiographie numérisée permet de visualiser le trajet vasculaire en soustrayant les structures environnantes. Les premiers clichés, obtenus avant injection du produit de contraste sont numérisés pour permettre le calcul d'un masque qui sera soustrait des clichés pris après injection et également numérisés.

Les sources suivantes préexistaient à la numérisation mais celle-ci, généralement appliquée après acquisition, permet d'améliorer les caractéristiques du signal et de faciliter la manipulation :

- la scintigraphie repose sur l'injection d'un isotope radioactif ayant un tropisme particulier pour un organe dont il donne une image morphologique et fonctionnelle, basée sur le comptage des radiations émises et qui reste de petite dimension (128x128 pixels).
- l'échographie est basé sur l'émission et la réception d'ondes ultrasonores. Ces signaux acoustiques sont convertis en signaux électriques puis numérisés et traités pour être

notamment affichés sur un écran vidéo.

- la radiographie conventionnelle utilise des radiations ionisantes, qui après atténuation par traversée du corps, viennent frapper un film chimique radiosensible. Ce procédé a des temps de pose très courts et une excellente définition. La numérisation peut s'effectuer secondairement sur un film développé ou directement en remplaçant le film par une plaque au phosphore.
- la vidéo-endoscopie permet d'obtenir une image de l'intérieur du corps. Le signal vidéo correspondant peut être ensuite numérisé. Il en est de même des images microscopiques ou des images photographiques argentiques. Dans ce cas, l'image papier est convertie par un numériseur, vulgairement appelé scanner, qui effectue une analyse de la densité et de la couleur de l'image, au cours d'un balayage pixel par pixel. Les appareils de prise de vues fournissant directement une image numérique sont maintenant courants mais leur résolution est encore inférieure à celle obtenue par un support argentique. Une attention particulière doit parfois être apportée au calibrage des couleurs (dermatologie).

ANALYSE D'IMAGE

L'analyse d'une image s'effectue au cours d'une succession de procédures constituant la chaîne de traitement : acquisition - pré traitement - mise en évidence des caractéristiques de l'image - extraction de paramètres quantitatifs d'intérêt clinique - interprétation.

Le pré traitement consiste d'une part à compenser les imperfections du système de génération des images (distorsion) et à atténuer le bruit dont sont affectées les images. Le bruit est responsable d'une diminution du contraste et d'un gommage du contour des objets. Sa réduction, généralement par des techniques de filtrage passe-bas, améliore l'aspect général de l'image mais peut, en contrepartie, masquer certains détails qui se traduiraient également par de hautes fréquences spatiales.

Les imperfections liées aux capteurs et au bruit rendent souvent difficile la détection de certaines anomalies, aussi plusieurs techniques permettent d'améliorer la visibilité des structures. Le calcul de l'histogramme de densité fournit la répartition des pixels en fonction de la densité de l'image. Il permet d'apprécier le contraste général de l'image que des opérateurs permettent ensuite de modifier. La méthode de fenêtrage de visualisation (ou seuillage) permet, dans le cas d'une image comportant de trop nombreux niveaux d'intensité, de choisir deux seuils de visualisation entre lesquels apparaît certaine fraction de l'image. La technique des iso-contours consiste à tracer les courbes de niveau de densité d'une image, ce qui permet au mieux de mettre en évidence certaines structures morphologiques. La soustraction d'images permet de mettre en évidence certains détails (cf angiographie numérisée). Au point de vue spatial, des opérateurs géométriques permettent le déplacement en rotation ou translation comme l'agrandissement d'une zone particulière. Enfin, la mise en relation d'images 2D correspondant à des coupes étagées du corps (tomodensitomètre ou IRM) permet des reconstructions d'images en 3 dimensions.

Le traitement des séquences d'images a pour but d'extraire des informations de nature

dynamique. Aux techniques précédentes, s'ajoutent celles liées notamment au recalage des images d'une même séquence, en raison de différences causées par des instabilités de la technique ou des mouvements parasites mais indépendantes du phénomène à étudier. Ces techniques reposent sur l'optimisation d'un critère de similitude entre les images.

La détermination de paramètres quantitatifs à partir d'une image dépend du domaine d'application ; on peut citer : le calcul de surfaces ou de volumes, la détermination d'une densité ou le comptage d'éléments. La radiothérapie s'appuie sur le calcul de doses et de position des sources maximisant l'effet thérapeutique et minimisant la radiotoxicité sur les organes voisins.

L'interprétation automatique des images, comme aide au diagnostic, est complexe et reste du domaine de la recherche. Elle fait appel à de nombreuses techniques, notamment de reconnaissance des formes et d'intelligence artificielle, et combine des informations de natures diverses : le problème consiste à identifier les paramètres et les structures significatifs puis à les comparer à des structures connues ou à les confronter à des connaissances théoriques ou expérimentales.

EXEMPLES D'APPLICATIONS

- L'évaluation de la contractilité myocardique peut s'effectuer à partir de la numérisation d'images échocardiographiques. Après sélection des images correspondant respectivement à la diastole et à la systole puis traçage des contours internes et externes du myocarde, le programme calcule les modifications d'épaisseur du muscle dans les différents secteurs, permettant de repérer ainsi les zones d'hypocontractilité (infarctus).
- L'identification de chromosomes pour la détermination du caryotype, repose sur des techniques de reconnaissance des formes. L'image de la préparation cytologique est numérisée. L'analyse des variations de densité permet de repérer les chromosomes. Ensuite leur regroupement en paires est le résultat de plusieurs phases : détermination du grand axe, calcul de la largeur le long de l'axe pour repérer le centromère et obtenir la longueur de chaque bras.
- La transmission d'images par réseau pour consultation par un expert permet des applications de télédiagnostic ou de télésurveillance, notamment en radiologie ou cytopathologie.
- La reconstruction en trois dimensions d'images des organes montre les rapports des structures entre elles (organes, tumeurs, structures vasculaires). Particulièrement employée dans le domaine de la neurologie, elle peut déboucher sur la création d'un espace en réalité virtuelle où le médecin peut se déplacer ou sur la production automatique de moules en trois dimensions, afin que le chirurgien puisse repérer les voies d'abord ou répéter l'intervention.
- La chirurgie assistée par ordinateur associe aux phases d'acquisition et d'interprétation d'images, deux étapes de raisonnement et de commande robotique. L'objectif est de faciliter la réalisation de gestes médico-chirurgicaux complexes. A partir d'images reconstruites, souvent à partir de plusieurs sources, le raisonnement constitue un modèle du patient et

permet de simuler l'intervention (geste virtuel). La dernière étape peut prendre la forme d'une aide passive (détection d'écarts au geste prévu), semi-active (système de contraintes) voire active (autonomie du robot).

SYSTEMES INFORMATIQUES

Une image de 1024x1024x8 bits occupe 1 Mo donc l'imagerie entraîne des besoins importants en **mémoire** : mémoire centrale pour les calculs et mémoires de masse pour l'archivage. C'est ici que s'emploient au mieux les systèmes d'archivage sur disque optique à grande capacité : disques optiques numériques (DON) de grande capacité (plusieurs Go) qui peuvent être réunis en juke-boxes. Les calculs manipulent des données de grande taille mais ils sont souvent répétitifs et basés sur la manipulation de vecteurs et de matrices. Ils sont facilités par l'emploi de processeurs spécialisés : processeurs d'images (disposant de fonctions de traitement d'image précablées) ou processeurs vectoriels. Enfin le transfert d'images entre ordinateurs nécessite la mise en place de réseaux à grande vitesse (soit locaux comme Ethernet avec un débit maximum de 10 à 100 Mbits/s, soit à grande distance comme Numéris à 64 Kbits/s et ATM permettant plusieurs Gbits/s). La numérisation des séquences d'images a entraîné le développement des techniques de compression d'images qui font l'objet de standards. Des facteurs de réduction de taille de 2 à 3 sont obtenus sans perte de qualité tandis que des rapports de 8 à 10 n'induisent qu'une dégradation minime de l'information.

Les systèmes informatiques d'imagerie, dénommés PACS (Picture Archiving and Communication System) ont comme objectif de collecter les images provenant de multiples modalités, de les stocker dans une banque d'images, afin d'en faciliter la communication et la consultation uniforme sur des stations de visualisation ou de manipulation, dispersées dans l'hôpital. Leur développement pose d'une part des problèmes matériels (intégration de sources disparates, capacité de stockage, vitesse de transmission) mais aussi logiques (intégration avec le système d'information de l'hôpital, modélisation de l'information, identification des patients ou des images) ayant conduit à la définition d'un standard dit ACR/NEMA-DICOM.

Dans un autre registre, le développement de réseaux publics à grande vitesse (Numeris) et l'intégration de fonctionnalités multimédia (son et image vidéo) sur les micro-ordinateurs favorise l'émergence d'applications de téléconsultation ou de télé-expertise.

SYSTEME D'INFORMATION DE LABORATOIRE

L'informatisation des laboratoires est un des domaines d'application les plus anciens et en est désormais au stade de l'exploitation industrielle. L'intérêt de ces systèmes est généralement dans le gain de productivité et de qualité du service effectué. L'origine de la demande d'informatisation est l'augmentation du volume d'informations à gérer : inflation de la consommation d'examens, augmentation du nombre de résultats fournis par des systèmes d'analyse de plus en plus automatiques.

Quelle que soit la discipline (biochimie, hématologie ou bactériologie), ces systèmes permettent les mêmes fonctionnalités :

- l'enregistrement des demandes d'analyse : il nécessite l'identification de plusieurs paramètres, le patient, le service demandeur et l'examen. L'identité du patient peut être enregistrée ou transmise par un autre système informatique ; le service demandeur est généralement identifié par un code tandis que l'examen peut être identifié par saisie manuelle au clavier ou plus souvent par lecture directe du document demandeur ou par transmission informatique.
- le tri des examens et ventilation par poste de travail : le tri des échantillons est en partie manuel mais implique l'attribution d'un numéro et l'édition d'étiquettes, aisément réalisées avec un système informatique. L'emploi de systèmes d'identification par codes à barres est courante.
- l'acquisition des résultats : il peut s'agir d'une acquisition automatique par interfaçage avec un analyseur automatique qui exige néanmoins une vérification et une double validation par le technicien de laboratoire et par le biologiste. Mais certains examens (hématologie, cytologie, anatomo-pathologie, bactériologie) sont largement manuels et la saisie l'est donc également. Dans le cas de comptes-rendus, les contraintes liées à la formalisation du langage médical s'appliquent ici et ont débouché sur des solutions de type questionnaire standardisé et langage de classification (par exemple, SNOP ou Adicap en anatomo-pathologie).
- la consultation et l'édition des résultats analysés : elle peut être accélérée par la consultation sur écran ou la transmission télématique des résultats. D'autre part, l'enregistrement informatique permet de se reporter aux résultats antérieurs, comme d'effectuer des recherches multicritères. Hors contexte d'urgence, les résultats doivent être validés par le biologiste avant transmission.
- la gestion du laboratoire : il s'agit de la comptabilité des actes effectués, de la gestion des stocks de matériel consommable, de l'édition de statistiques d'activité ventilée selon divers critères.
- l'archivage des dossiers : il est de la responsabilité du biologiste. Les facilités et les contraintes (protection, confidentialité) sont celles de tout dossier informatisé.

A côté des avantages induits par les fonctions précédentes : économie, ergonomie, meilleure organisation du travail et meilleure gestion ; d'autres retombées, de nature plus médicale, sont possibles :

- Contrôle de qualité : un système informatique permet de réaliser un certain nombre de contrôle de validité des résultats, notamment de vérifier sur des séries qu'il n'y a pas de dérive dans le temps. L'emploi d'échantillons étalons permet de calibrer le système.
- Connaissance épidémiologique : dans le cadre de protocoles d'enquête bien définis, ces systèmes permettent d'obtenir une description uni ou multidimensionnelle d'échantillons de référence de sujets normaux ou atteints de pathologies identifiées débouchant sur le calcul de valeurs de référence ou l'évaluation de la valeur décisionnelle des examens.

Cependant, l'usage de tels systèmes appelle quelques remarques :

- la tendance des industriels d'intégrer en un seul système les différents analyseurs et le système informatique pour constituer un système de gestion de laboratoire déporte le problème de l'intégration vers d'autres systèmes d'information : services cliniques, autres laboratoires, hôpital. L'échange de données sera facilité par la définition d'un standard logique d'interface.
- la transmission des résultats en urgence doit se limiter au contexte de l'urgence si les résultats n'ont pas été validés. Elle pose des problèmes de confidentialités importants liés à la multiplicité des récepteurs potentiels et à la nature des informations transmises.
- le laboratoire de bactériologie est caractérisé par une activité encore largement manuelle alors que les perspectives y sont importantes : suivi de l'évolution de la population microbienne, adaptation des thérapeutiques, dépistage des infections nosocomiales.

SYSTEME D'INFORMATION DE L'UNITE DE SOINS

L'unité de soins est l'entité dont la fonction est de produire des soins médicaux à visée diagnostique, thérapeutique ou d'évaluation. Elle s'appuie sur une équipe soignante et prend en charge des patients hospitalisés ou des consultants externes.

Les fonctions d'un système de gestion de l'unité de soins comprennent d'abord des fonctions de prise en charge et de suivi individuel des patients :

- identification médico-administrative ;
- rendez-vous et accueil des patients ;
- accès au dossier antérieur ;
- saisie des données de soins ;
- prescription de soins ;
- saisie des actes ;
- saisie et mise à jour des résultats d'examen ;
- compte-rendu, résumé de sortie et synthèse.

Il s'y greffe des fonctions intéressant l'unité dans son ensemble :

- fonctions de communication ;
- gestion du courrier et bureautique ;
- bilans d'activité ;
- gestion des ressources ;
- accès aux banques de données, enseignement et recherche.

L'informatisation des unités de soins pose des problèmes complexes de modélisation du système d'information et de stratégie de mise en place. Les demandes des différents acteurs de l'unité de soins ne sont pas toujours convergentes. La communication avec d'autres systèmes comme l'intégration dans le système d'information de l'hôpital est en partie dépendante de l'infrastructure technique parfois déjà en place.

QU'EST-CE QUE L'HOPITAL ?

L'hôpital est une institution dont l'objectif est de soigner et si possible de guérir des malades. S'il est en quelque sorte un producteur de santé, il est abusif de l'assimiler à une industrie car il ne maîtrise aucunement ses entrées. Il assure deux fonctions : une fonction d'accueil et une fonction technique qui prédomine de plus en plus. Pour accomplir ces fonctions, l'hôpital dispose de certaines ressources et emploie un personnel nombreux, réparti en catégories aux fonctions distinctes. En conséquence, l'hôpital est une énorme machine économique et financière. Les hôpitaux représentent environ 50% des dépenses de soins, lesquelles sont encore croissantes, dépassant 10% du PIB. Une meilleure connaissance de leur fonctionnement, notamment par la mise en place d'un système d'information adapté, devrait permettre une meilleure gestion et utilisation des ressources tout en contribuant à améliorer la qualité des soins.

L'HOPITAL, SYSTEME D'INFORMATION

L'hôpital est en fait une fédération de sous-systèmes fonctionnellement distincts mais non disjoints, à l'intérieur desquels et entre lesquels circulent des flux d'informations.

- le sous-système d'action médicale concerne l'activité mise en œuvre par l'équipe soignante pour répondre au problème du malade : l'information recueillie sur le patient, la constitution et la consultation du dossier du malade, les connaissances médicales, les processus de décision.
- le sous-système logistique comprend l'ensemble des flux résultant des actions médicales : prescriptions, résultats, transferts, archivages. Il met en jeu les divers services cliniques et plateaux techniques de l'établissement pour appuyer l'activité de l'équipe soignante.
- le sous-système de recherche et d'études travaille sur des regroupements de dossiers, à condition que ceux-ci aient été correctement constitués, à des fins épidémiologique ou d'évaluation de la qualité des soins, alimentant en retour la connaissance médicale ou les sous-systèmes d'administration et de planification.
- le sous-système de l'administration quotidienne de l'hôpital s'intéresse à la facturation, à la gestion du personnel, à la gestion des stocks et d'une manière générale à la comptabilité.
- le sous-système de planification hospitalière a une vision plus stratégique. Il s'appuie sur l'analyse d'activité ou les études de morbidité hospitalière pour engager des décisions d'investissement structurels, matériels et humains. Il est en rapport avec des entités extérieures : autorités de tutelle, offre de soins environnante, état de santé de la population desservie.

Tous ces sous-systèmes sont interdépendants et sont pour une large part centrés sur le dossier du malade. Ainsi même si l'information "médicale" et l'information "administrative"

ne sont pas recueillies par les mêmes personnes, ne mettent pas en œuvre les mêmes procédures ni les mêmes connaissances et ne s'intéressent pas a priori aux mêmes faits, l'action médicale ne peut s'abstraire d'informations de type administratif (identité, profession, coût des procédures) tandis que l'hôpital ne peut être convenablement géré sans considérer sa finalité de soins (qualité des soins, progrès des connaissances, adaptation aux besoins de la population).

CONCEPTION DU SYSTEME INFORMATIQUE DE L'HOPITAL

Le système d'information est une réalité intrinsèque à l'hôpital, indépendante de toute informatisation. La mise en place d'un système informatique est souvent l'occasion de son réexamen, en vue de son automatisation plus ou moins complète, aboutissant à la constitution d'un système d'information hospitalier (**SIH**). Cette automatisation recouvre généralement les fonctions de mémorisation et de communication, voire de traitement, mais elle ne saurait accomplir l'essentiel des fonctions de décision où elle apparaît comme un support ou une aide mais jamais comme un substitut. L'informatisation, c'est-à-dire la (ré)organisation, la formalisation et l'automatisation des flux d'information devrait apporter une gestion plus rationnelle de son activité, une meilleure connaissance du fonctionnement de l'hôpital, une amélioration de la qualité des soins, un meilleur support pour la recherche et l'enseignement.

Quelques principes doivent guider la mise en place du système informatique de l'hôpital :

- conception globale ;
- position centrale du malade et de son dossier ;
- priorité de l'information médicale ;
- saisie unique de l'information à la source, partage et retour de l'information ;
- souplesse : l'information prime sur l'ordinateur d'où une exigence d'interfaces ;
- mémorisation et communication ;
- protection des données ;
- disponibilité.

D'autre part, l'analyse des flux comme le respect des principes ci-dessus conduit à proposer une organisation structurale du système informatique en strates concentriques :

- identification des malades ;
- mouvements ;
- fonctions cliniques ;
- administration ;
- gestion.

Cette organisation ne préjuge pas de la configuration matérielle du système informatique, laquelle est également conditionnée par la taille et le mode de fonctionnement de l'hôpital.

Identification et mouvement des malades

L'identification est le noyau central. Un système d'identification devrait répondre aux

critères suivants, en partie contradictoires :

- unique pour chaque patient ;
- permanent ;
- universel ;
- disponible ;
- économique ;
- propre au système de santé.

En général, on essaye de doubler l'identification littérale par un numéro permanent d'identification.

La strate des mouvements a pour but principal de localiser à tout instant chaque malade actif, c'est-à-dire de répondre à la triple interrogation "qui ? quand ? où ?". Il s'y adosse les fonctionnalités suivantes :

- gestion des rendez-vous ;
- édition préalable des listes de consultants ou d'hospitalisés ;
- éditions d'étiquettes correspondant au séjour (identité du malade, date et lieu d'accueil) ;
- enregistrement des transferts du malade, dont la sortie.

Cet ensemble participe au sous-système logistique qui gère les flux au niveau de l'hôpital.

Strate des fonctions cliniques

Sur les malades identifiés et localisés sont recueillies des données diverses : données cliniques, laboratoire, imagerie, thérapeutiques. C'est ici que se constituent le dossier médical et le dossier de soins. Cette strate s'appuie sur l'informatisation des dossiers, des unités de soins, des laboratoires et des plateaux techniques. Elle est potentialisée par l'intégration de ces différents systèmes informatiques, ce qui met l'accent sur la définition d'interfaces de communication, matérielles et logicielles. En effet, s'il existe un cloisonnement assez net des ressources matérielles et humaines de l'hôpital en services, la prise en charge d'un malade est de plus en plus rarement le fait d'un seul service, d'autant que la distinction entre service clinique et médico-technique tend à s'estomper, chaque entité pouvant devenir demandeur et prestataire.

L'ouverture du SIH vers l'extérieur, en direction des autres structures de soins, médecine libérale et santé communautaire, est encore en développement, afin de décroisonner médecine hospitalière et extra-hospitalière et d'assurer une continuité de suivi du patient (réseaux de soins).

Structures dédiées à l'information médicale

Deux structures, dédiées à l'information médicale, occupent une place importante dans le fonctionnement d'un SIH.

- le Service Central des Dossier Médicaux (SCDM) : il est en principe obligatoire dans les Centres Hospitaliers Régionaux. Il permet la mise en place d'un dossier unique par patient, en assurant la coordination de la centralisation de toutes les informations relatives au

malade. Le SCDM est chargé de la bonne conservation des archives, de leur communication dans de bonnes conditions et de l'intégrité du dossier (physique et logique) afin de faciliter la consultation et l'exploitation scientifique. Mais ceci exige une stricte unité de stockage (on distingue toutefois archive morte et archive vivante) et une forte disponibilité matérielle et humaine.

- le Département d'Information Médicale (DIM) prend en charge les problèmes liés à la gestion de l'information médicale dans un hôpital : participation à l'élaboration du schéma directeur du SIH, exploitation du SIH (gestion des dossiers, exploitation statistique, cohérence du système, respect du secret médical), développement de systèmes d'informatique médicale, conseil méthodologique, réalisation d'études, formation. Sa direction doit être médicale, afin d'assurer une interface avec le corps médical.

On pourrait y adjoindre une troisième structure, la bibliothèque médicale ou centre de documentation, assurant la diffusion des connaissances et l'accès aux banques de données.

STRATEGIES ET SOLUTIONS TECHNIQUES

En fait, si l'approche centrée sur le dossier patient est la plus conforme à la finalité de l'hôpital, on retrouve souvent, pour des raisons historiques ou humaines, une organisation plus ou moins calquée sur les structures existantes. En effet, celle-ci est bien perçue des acteurs de l'hôpital et valorise la notion d'équipe de soins. En contrepartie, elle a tendance à éclater le système d'information du patient en sous-systèmes avec leur risque de redondances ou d'incohérences. Plus rarement, la structuration reprend la répartition des personnels en différentes catégories, avec des risques d'incohérence plus grands. La mise en œuvre d'un SIH peut faire appel à diverses solutions techniques, qui sont en grande partie le reflet de l'évolution de l'informatique elle-même.

Les premiers systèmes datent des années 70 et se caractérisent par une architecture centralisée, souvent mono-constructeur, comportant un ordinateur central et des terminaux. Cette approche, dite verticale, centrée sur le dossier patient, favorise la saisie unique et le partage de l'information. Elle rend l'établissement fortement dépendant d'une solution "clés en main", qui oblige à une standardisation forte et dont l'évolution est difficile. Pour ces raisons, elle paraît plutôt indiquée pour des hôpitaux de petite taille.

L'architecture basée sur les systèmes départementaux, dite approche horizontale, est contemporaine de la mini-informatique. Elle consiste en l'équipement de chacune des structures de l'hôpital d'applications spécifiques. Cette approche, qui a intéressé tout d'abord l'administration et les plateaux techniques, permet une meilleure adaptation des produits aux besoins des utilisateurs qui disposent de plus de choix, ainsi qu'une évolution plus douce. Par contre, si elle autorise le regroupement de données dans le cadre de services de même spécialité, elle conduit souvent à une forte hétérogénéité à l'intérieur d'un même établissement, génératrice d'incompatibilités ou de surcoûts. Cette approche est plutôt celle d'institutions hospitalo-universitaires (Assistance Publique, Massachusetts General Hospital).

L'objectif d'une approche mixte, dite distribuée, est de dépasser les limites précédentes.

Les applications correspondant aux grandes fonctions d'un SIH (identité, mouvement, résultats de laboratoire) sont assurées par des systèmes centraux serveurs, accessibles par réseau informatique aux applications dédiées à l'utilisateur final (médecin, infirmière), selon un mode dit "client-serveur". Cette approche nécessite la sélection d'une architecture matérielle, logicielle et de réseau adaptée, appartenant de préférence à un standard industriel, afin de n'être pas dépendant d'un constructeur. Sa plus grande complexité tient à la nécessité d'une approche globale et à la définition d'interfaces de communication.

EXEMPLES DE SIH

Le système PCS (Patient Care System) d'IBM est un SIH centralisé, utilisé dans de nombreux hôpitaux nord-américains, et à Paris à l'hôpital Robert Debré. HELP, développé à l'Université d'Utah, associe à un SIH centralisé un système d'alarmes et d'aide à la décision.

L'hôpital cantonal de Genève est un hôpital universitaire de 2000 lits où une équipe médicale (J.R. Scherrer) s'est lancée dans les années 70 dans la réalisation d'un SIH, dénommé Diogène. Il s'agit d'un système informatique travaillant en temps réel, sur un mode interactif, structuré autour d'une base de données ordonnée autour du fichier de patients. Il est orienté sur les besoins des utilisateurs (interactivité), sa base de données est historique, il est capable de fournir des données synthétiques d'activité et de coût, les dossiers sont codés en combinant la CIM-9 et le SNOMED ce qui favorise l'exploitation des données, enfin une attention particulière a été portée à la sécurité et à la protection des données (badge d'identification). Il évolue vers le système Diogène 2, conçu selon une approche distribuée, intégrant notamment un système PACS et des stations de travail multimédia.

L'Assistance Publique-Hôpitaux de Paris (AP-HP) s'est engagée depuis 10 ans dans la modernisation de son système d'information. Auparavant essentiellement consacrée à la gestion comptable et fonctionnant selon une logique administrative de séjour, celui-ci privilégie maintenant une logique patient. Cette réorganisation a nécessité la mise en place d'un numéro permanent d'identification, de serveurs d'identité, d'un réseau informatique intra et inter-hospitalier et d'applications clientes au niveau des unités de soins ou des plateaux techniques. Cependant, la taille de l'établissement (50 hôpitaux), une centralisation certaine et la nécessité d'intégrer l'existant sont la cause de difficultés.

RESSOURCES

Les coûts de mise en place et de fonctionnement d'un SIH sont encore difficiles à estimer. Ainsi, la plupart des établissements sont dans une phase de croissance de leur informatique et non de renouvellement et la part de budget consacrée à l'informatique peut varier de 1,5 à 5%.

La prise en compte des ressources humaines conditionne largement le succès d'un SIH, qu'il s'agisse de l'intégration des compétences administratives, médicales et universitaires, ou du dépassement de la dualité informatique administrative - informatique médicale. Dans tous les cas, le SIH ne peut fonctionner sans participation active de l'ensemble des personnels

concernés. Cette participation doit être initialisée dès la phase d'analyse sous peine de rejet ultérieur et implique en outre une formation adéquate de l'ensemble des acteurs.

DOSSIER INDIVIDUEL

Les données relatives à la santé d'une personne sont disséminées dans différents systèmes d'information : du généraliste ou spécialiste au système d'information hospitalier en passant par les organismes payeurs. Aussi, afin de garantir une certaine unicité de l'information, est-il prévu de doter chaque patient d'un dossier portable, constitué d'une carte à microprocesseur.

La carte à microprocesseur (carte à puce) réunit dans un support plastique au format carte bancaire : un microprocesseur, un système d'entrées-sorties, un bloc de sécurité qui assure les contrôles sur l'accès aux mémoires, une mémoire de données réinscriptibles, d'une capacité de 1 à 8 pages de caractères alpha-numériques, permanente et dont la mise à jour nécessite un lecteur-enregistreur spécial. La carte à processeur peut gérer elle-même ses données et les droits d'accès, ce qui assure une sécurité de haut niveau et permet de l'utiliser comme une clé de sûreté.

Le dossier portable sur carte à puce a fait l'objet de plusieurs études de faisabilité. Le dossier d'une carte de santé généraliste (par exemple la carte Santal expérimentée à Saint-Nazaire) comporte généralement les zones suivantes : identification de la personne, groupe sanguin - vaccination - allergie - urgence, antécédents, histoire de la maladie, thérapeutique en cours. La carte peut être utilisée dans un contexte de spécialité, par exemple, pour aider à la prise en charge du malade dialysé (expérience Dialybre) où elle assure l'enregistrement des données de suivi du patient. Dans tous les cas, le succès de la carte dépend des structures médicales dans lesquelles elle s'insère, de l'adhésion des professionnels et d'une large diffusion parmi les patients, corollaire d'une standardisation du codage de l'information et de la disponibilité de lecteurs. D'autre part, elle pose aussi des problèmes de partage de l'information.

Une autre application de la carte à usage du patient est la carte d'ouverture de droits de l'assuré social (carte VITALE). Ici, la carte sert à identifier le porteur et les bénéficiaires de soins, à préciser ses droits et à fournir au prestataire de soins l'information nécessaire à la déclaration des coûts du traitement vis-à-vis de l'Assurance Maladie (taux de prise en charge). La version Vitale 2 dispose en outre d'une zone réservée à des informations médicales afin de remplacer le carnet de santé. L'emploi de la carte comme outil monétique apporterait une meilleure transparence de l'activité médicale, un allègement des tâches de retranscription d'informations et un meilleur contrôle des fraudes. La carte Vitale devrait être diffusée à l'ensemble des assurés sociaux d'ici la fin 1999.

Environ 70% des 150.000 médecins français exercent une activité libérale, partielle ou totale. La constitution de cabinets de groupe est de plus en plus fréquente, de même que l'organisation des médecins en réseaux regroupant généralistes et spécialistes. D'autre part, la continuité du suivi d'un patient incite au renforcement des liens entre hôpital et prise en charge extra-hospitalière, ceci dans un but de qualité des soins mais aussi en raison des contraintes économiques, lesquelles débouchent sur une évaluation de l'activité médicale et un contrôle renforcé de la part des organismes de Sécurité Sociale (RMO : références médicales opposables).

Les tâches du généraliste, médecin de première ligne, sont particulièrement étendues :

- fonctions diagnostique, thérapeutique et d'évaluation ;
- fonctions de surveillance ;
- fonctions de dépistage, de prévention et d'éducation ;
- fonctions médico-légales ;
- fonctions d'enseignement, d'évaluation et de recherche.

Il interfère ainsi avec ses collègues libéraux, le système hospitalier ou universitaire, la santé communautaire, le secteur paramédical, biologique ou pharmaceutique, les organismes de Sécurité Sociale.

L'informatisation du cabinet médical peut contribuer à améliorer les pratiques comme l'efficacité du praticien mais elle nécessite une certaine maîtrise pour en faire un outil utile et intégré à sa pratique et éviter les désillusions. Ainsi, on peut envisager les fonctions suivantes, plus ou moins aisées à mettre en œuvre :

- gestion administrative et financière du cabinet ;
- bureautique et secrétariat ;
- gestion du dossier patient : elle est compliquée notamment par la brièveté de la consultation et l'étendue du domaine à considérer. En outre, il n'existe pas en France de nomenclature réellement adaptée à la pratique extra-hospitalière (codes Read en Angleterre). Il importe donc de définir rigoureusement les finalités de ce dossier, en particulier son usage individuel ou collectif ;
- communication avec l'extérieur (transmission de résultats de laboratoire, téléconsultation) ;
- aide à la décision médicale : il s'agit surtout de l'accès aux banques de données ou de l'aide à la prescription médicamenteuse ;
- formation continue.

La mise en œuvre de ces fonctions est facilitée par la diffusion de la micro-informatique, l'amélioration de la convivialité des logiciels et l'extension des réseaux de communication, qui se concrétisera par un réseau informatique dédié : le Réseau Santé Social.

D'autres applications de l'informatique médicale intéressent le praticien :

- La carte du Professionnel de Santé (CPS) s'adresse notamment au médecin généraliste. Il

s'agit d'une carte à processeur qui sert de carte d'habilitation. C'est d'abord un outil de sécurité qui permet d'identifier et d'authentifier le porteur mais aussi de mémoriser ses droits d'accès vis-à-vis de telle ou telle information (accès au SIH) ou sa situation conventionnelle et de faire office de signature électronique. Elle devrait être généralisée à l'ensemble des professionnels de santé (médecins, pharmaciens, dentistes, infirmières, kinésithérapeutes, etc) à partir de 1998 pour mettre en œuvre la télétransmission des feuilles de soins électroniques.

- la participation à un réseau télématique de médecins sentinelles contribue à une meilleure connaissance épidémiologique de certaines pathologies ; dans le même ordre d'idée, des associations de généralistes ont mis en place des études sur les motifs de recours ou sur l'évaluation de la qualité des soins ;
- la télémédecine s'appuie sur l'existence d'un réseau informatique performant mettant en relation le généraliste isolé et le spécialiste, souvent hospitalier, afin de transmettre des images pour interprétation, d'établir une visio-conférence faisant office de staff ou de participer à des séances de formation.

En France, seulement 20% des cabinets médicaux sont informatisés, contre près de 90% en Angleterre ou en Hollande, en raison d'une définition des cahiers des charges par les organisations professionnelles, mais aussi d'incitations financières plus précoces et d'une organisation de la médecine qui réduit le nomadisme des patients. Les ordonnances d'avril 1996 réformant le système de soins prévoient l'informatisation des cabinets, notamment pour leur interfaçage avec les organismes d'Assurance maladie. La création de réseaux de soins intégrés, pour mieux coordonner et rationaliser l'offre de soins, pose le problème de la standardisation du langage et des procédures et celui de la confidentialité des informations médicales.

SYSTEME D'INFORMATION EN SANTE PUBLIQUE

Il s'agit essentiellement des systèmes d'information concernant la connaissance de l'état de santé de la population et à moindre degré, le fonctionnement du système de soins. En effet, la mise à disposition des usagers d'informations leur permettant de gérer leur propre santé est une notion encore embryonnaire en France. Quoiqu'il en soit, la situation de la Santé Publique est caractérisée par la diversité des sources d'information et surtout leur relative absence de mise en relation ou d'interconnexion. On peut citer, parmi beaucoup d'autres systèmes :

- le système d'information sur la mortalité et les causes médicales de décès : cette enquête permanente utilise les données des certificats de décès, établis par le médecin, convoyés par la mairie et la DDASS pour être transmis à l'INSERM et à l'INSEE. L'INSERM assure la codification des causes de décès (CIM) et établit les statistiques nationales de décès.
- les registres des maladies sont des enquêtes permanentes s'intéressant à l'enregistrement

continu de l'ensemble des cas d'une pathologie donnée (cancer, maladies cardio-vasculaires) apparaissant dans une population géographiquement déterminée (département ou région). Ils répondent à un objectif d'information sanitaire (surveillance du niveau de morbidité dans le temps et dans l'espace) et servent de support à des enquêtes épidémiologiques plus détaillées.

- les réseaux téléinformatiques de surveillance épidémiologique par télématique s'appuient sur un centre informatique serveur, chargé de collecter les données et d'établir des statistiques, relié par un réseau télématique à un échantillon de médecins sentinelles qui signalent les cas au centre serveur par le moyen d'un terminal (minitel ou PC). Les données transmises sont anonymes et concernent l'état de santé et l'état vaccinal. Elles permettent d'estimer en temps réel les incidences des maladies concernées. Aussi ces réseaux concernent-ils plus particulièrement les maladies transmissibles.
- les observatoires régionaux de la santé (ORS) sont des organismes qui maintiennent des systèmes d'information permanents permettant l'aide à la décision médicale et de santé publique, fondé sur le recueil, l'intégration et l'analyse de données d'origines diverses.

EVALUATION DE L'ACTIVITE MEDICALE

L'augmentation continue des dépenses de santé pose à la collectivité le problème de la répartition optimale des ressources et du choix de critères permettant de concilier l'exercice d'une médecine de qualité et le respect de contraintes économiques. La première étape d'une maîtrise des dépenses passe par une meilleure connaissance de l'activité de soins, notamment hospitalière, c'est l'objet du Programme de Médicalisation du Système d'Information (PMSI), initié en 1982.

LE PMSI s'applique pour l'instant aux activités de court et de moyen séjour des établissements, publics et désormais privés, et vise à adapter leur dotation à leur activité. Pour décrire l'activité, chaque séjour d'un patient doit faire l'objet d'un résumé systématique, dit résumé standardisé de sortie (RSS), sur lequel figurent les caractéristiques anonymes du patient, la durée du séjour, les différentes unités de soins concernées ainsi que le diagnostic principal, les diagnostics associés et les principaux actes réalisés. Le codage des diagnostics s'effectue avec la CIM-10 tandis que le CDAM permet le codage des actes. Le choix d'un diagnostic principal est en rapport avec le problème qui a mobilisé le plus de ressources, mais il est souvent difficile, surtout dans le cas d'une pathologie chronique ou d'un séjour multi-services. Tous les résumés sont collectés par le DIM qui effectue un certain nombre de contrôles puis procède à un groupage des séjours. Cette opération s'effectue à l'aide d'un logiciel groupeur, défini au niveau national, qui répartit les séjours en groupes homogènes de malades (GHM). Ces groupes homogènes de malades, au nombre de 500 environ, représentent des situations médicales relativement semblables d'un point de vue pathologie et consommation de ressources, comme par exemple, "intervention sur l'œsophage, l'estomac et

le duodénum, âge de 18 à 69 ans, sans complication ni comorbidité". L'activité annuelle de l'hôpital, ainsi résumée sous forme d'un certain nombre de GHM, est communiquée à la DRASS, qui effectue un contrôle de validité des données et détermine le niveau de la dotation de l'établissement. En effet, une étude nationale permet de calculer la valeur financière de chaque GHM en unités fictives dits points ISA (indice synthétique d'activité). Cette valeur est traduite en francs au niveau régional, elle permet de répartir le budget régional entre les établissements, en corrigeant les sur ou sous-dotations. Au niveau de l'établissement, un système de comptabilité analytique permet de déterminer sur quels secteurs (investissement, matériel, consommables, personnel) l'hôpital est économe ou trop cher, et ceci pour chaque GHM et pour chaque unité de soins.

Le PMSI ne cesse d'évoluer et s'étend ainsi depuis 1998 au moyen séjour. Il présente des défauts : il fonctionne selon une logique de séjours et non de patients, alors que la prise en charge d'un même problème peut s'effectuer en un seul séjour prolongé ou en plusieurs séjours brefs. A terme, l'aspect strictement comptable devrait être équilibré par la prise en compte de critères de qualité des soins.

Ainsi, les ordonnances de 1996 portant réforme hospitalière ont institué à l'échelon national une Agence Nationale pour l'Accréditation et l'Evaluation en Santé (ANAES) qui doit définir les normes de qualité pour les établissements de soins et procéder à leur évaluation. Au niveau régional, les Agences Régionales de l'Hospitalisation (ARH) sont chargées de réguler le fonctionnement et de déterminer la dotation des établissements qui dépendent d'elle. En effet, les établissements devront se soumettre régulièrement à une procédure d'accréditation leur donnant autorisation ou non de fonctionner sur des critères de sécurité ou de qualité des soins (respect de procédures plutôt que résultats).

OBJECTIFS

- Connaître le processus de digitalisation d'un signal analogique.
- Expliquer les avantages de la digitalisation des signaux physiologiques.
- Citer des exemples de digitalisation des signaux physiologiques.
- Citer les avantages et les contraintes informatiques de la numérisation des images.
- Décrire la chaîne de traitement des images ; connaître quelques exemples.
- Définir un système de communication et d'archivage d'images.
- Citer les possibilités offertes par l'informatisation des laboratoires.
- Identifier les sous-systèmes du système d'information d'un hôpital.
- Principales fonctions informatisables et bénéfiques attendus d'un SIH.
- Rôle d'un service central de dossiers médicaux et d'un DIM.
- Intérêt et faisabilité d'un dossier médical individuel.
- Applications d'informatique médicale pouvant intéresser un médecin libéral.
- Indiquer quelques systèmes d'information en Santé Publique

- Connaître l'objectif et le principe de fonctionnement du PMSI.

SYSTEMES D'AIDE A L'ACTION MEDICALE

La médecine est une discipline scientifique mais aussi une discipline d'action qui requiert souvent une prise de décision. Ce processus résulte de la confrontation d'un problème réel à l'expérience acquise et à un corpus de connaissances théoriques. Plusieurs catégories de systèmes informatiques peuvent participer à des phases diverses de ce processus : systèmes de banques de données, systèmes d'aide à la décision, enfin systèmes d'aide à l'enseignement.

BANQUES DE DONNEES

L'accroissement des connaissances (maladies, traitements, biologie, etc...), l'inflation des dossiers médicaux, comme la multiplication des publications scientifiques posent des problèmes de maîtrise de l'information. L'emploi de bases de données à des fins de mémorisation des connaissances permet de faciliter l'accès à l'information par l'organisation de banques de données. Les **banques de données** sont des collections d'informations sur un domaine particulier, organisées en vue de leur consultation par un grand nombre d'utilisateurs.

On distingue couramment deux catégories de systèmes :

- les banques de références **bibliographiques** ou systèmes documentaires indiquent les références où trouver l'information pertinente ;
- les banques de données **factuelles** ou banques de données proprement dites qui permettent un accès direct à l'information désirée.

SYSTEMES DOCUMENTAIRES

PRINCIPE

Le temps consacré à la documentation scientifique est forcément limité d'où l'intérêt d'un système informatisé qui permet de sélectionner les documents pertinents et de donner un aperçu du contenu des références retenues.

Au préalable, cependant, les références doivent être entrées dans la base après analyse et codage du document par un professionnel : analyste ou documentaliste, qui réalisera l'**indexation** du document, c'est-à-dire sa description résumée et standardisée. Il utilise pour cela un langage artificiel, reposant sur des mots clés, utilisés seuls ou combinés avec des connecteurs logiques, relationnels ou de liaison (cf MeSH). Plusieurs travaux de recherche

portent sur l'indexation semi-automatique par analyse directe du texte.

Lors de la recherche d'informations, le même langage doit être employé par l'utilisateur pour définir sa requête, éventuellement avec l'aide d'un documentaliste. De l'adéquation entre le codage de la requête et l'indexation des documents dépend le résultat qui comprend généralement les références bibliographiques des articles correspondants (titre, auteurs, journal, année, volume, pages, etc...). Les systèmes les plus perfectionnés permettent de commander une copie ou d'obtenir en temps réel un résumé voire la totalité de l'article (bases de données en texte intégral).

MODE D'INTERROGATION

Deux modes d'interrogation sont possibles :

- l'interrogation ponctuelle correspond à une question précise posée une seule fois au système ;
- le profil permet à un utilisateur de définir ses centres d'intérêt et d'être informé des références de façon systématique et périodique.

Mais la contrainte d'apprendre un langage artificiel est un frein pour le non-professionnel, aussi a-t-on développé des systèmes d'interrogation par menus, permettant par choix successifs de préciser la requête (par exemple, GratefulMed pour Medline). D'autre part, la recherche s'oriente vers la réalisation d'un langage artificiel unifié, UMLS, permettant d'interroger de façon homogène différentes bases n'ayant pas le même langage d'indexation.

EXEMPLES

Parmi les systèmes disponibles dans le domaine de la Santé, on peut citer :

- le système de documentation Medlars (Medical Literature Analysis and Retrieval System), avec sa version à accès direct **Medline** (Medlars on line), est un des plus grands systèmes automatisés mondiaux (9 millions de références). Créé aux Etats-Unis par la National Library of Medicine (NLM), il intègre chaque année plus de 400.000 nouveaux documents issus de près de 4000 périodiques. Ces documents sont indexés à l'aide du thesaurus MeSH par des analystes professionnels. Medline est accessible dans tous les pays. Le MeSH a été traduit en français par l'INSERM et le CNRS. Certains banques concernent des sous-ensembles spécialisés de Medline : Cancerline, Toxline, Chemline, Aidsline. L'interrogation se fait également par le langage MeSH en combinant les mots clés, les noms d'auteurs ou de périodiques, les périodes à l'aide d'opérateurs logiques booléens (et, ou, sauf). Afin de faciliter l'accès, plusieurs serveurs disposent de duplicata de la base Medline et proposent des logiciels d'interrogation simplifiés utilisant des systèmes de menus.
- Le système Cancernet est d'origine européenne, il est spécialisé dans la cancérologie. Il utilise un thesaurus multilingue et l'indexation est réalisée par des chercheurs en cancérologie.
- la base européenne Embase couvre les domaines de la pharmacologie, de la biologie et de la médecine ; elle est caractérisée par une intrégration très rapide des références.

- le système Pascal, français, est géré par le CNRS. il couvre l'ensemble du domaine scientifique (périodiques, congrès et thèses) et compte 9 millions de références.
- En santé publique, la BDSP (Banque de données en Santé Publique) regroupe plus de 100000 références documentaires sur la santé et les systèmes de soins.

PERFORMANCES D'UN SYSTEME DOCUMENTAIRE

Soit une question posée à une système documentaire contenant au total N documents, son résultat peut être complètement décrit par le tableau suivant :

	----- Document -----		
	Pertinent	Non pertinent	
Document Transmis	a	b	a+b
Document Non Transmis	c	d	c+d
	a+c	b+d	N

Plusieurs indices mesurent la qualité du système :

- le taux de rappel ("recall") est la proportion de documents transmis parmi ceux qui sont pertinents : $R = a / (a+c)$, il est analogue à la sensibilité ;
- le silence en est le complémentaire : $1-R = c / (a+c)$;
- le taux de chute ("fallout") est la proportion de documents transmis parmi les non pertinents : $C = b / (b+d)$;
- la spécificité en est le complémentaire : $Sp = d / (b+d)$.

Le système est d'autant meilleur que son taux de rappel et sa spécificité sont plus élevés et proches de 1. Mais il y a une certaine contradiction entre les deux qualités de rappel et de spécificité : en effet si on cherche à augmenter le rappel en ramenant tous les documents pertinents, on le paye souvent de l'extraction de références non pertinentes plus nombreuses.

Cependant, les paramètres précédents ne sont pas directement mesurables par l'utilisateur qui, le plus souvent, ignore le nombre total de documents comme le nombre de documents pertinents. Il ne connaît que les documents transmis (a+b) sur lesquels il peut mesurer :

- la **pertinence** ou précision : $P = a / (a+b)$;
- le bruit, son complémentaire : $B = b / (a+b)$.

Au-delà des critères quantitatifs, le choix d'un système documentaire est d'abord gouverné par le domaine concerné, mais aussi la période couverte, la nature des documents référencés, la fréquence de mise à jour, la qualité et la variété des informations transmises, enfin le mode d'interrogation.

BANQUES DE DONNEES

Il existe différentes catégories de banques de données. Toutes donnent accès à une information factuelle. Quel que soit le volume d'informations (souvent considérable) contenu dans la base, le temps d'accès est constant. Leur mise à jour se fait par ajouts et modifications des informations. Enfin, elle disposent de fonctions de recherche multicritères. Elles touchent

tous les domaines scientifiques.

BANQUES DE FAITS THEORIQUES OU EXPERIMENTAUX

- La toxicologie et la **pharmacologie** sont le sujet de nombreuses banques de données. En France, la banque d'informations sur les médicaments (BIAM) contient des informations sur près de 3000 substances et 8000 spécialités pharmaceutiques. Elle permet de consulter les propriétés d'un médicament ou de rechercher des interactions entre produits. L'accès peut se faire selon plusieurs critères : nom commercial, principe actif, classe pharmacologique, propriétés, indications, posologie, précautions d'emploi. La base Thériaque, élaborée par les pharmaciens hospitaliers du Centre national d'information sur le médicament hospitalier (CNIMH) ou la base Vidal offrent des fonctionnalités comparables.
- La biologie, plus particulièrement la **biologie moléculaire** et la génétique, en raison du récent et fort développement de cette discipline (projet "génomique humaine"), est couverte par de nombreuses bases de données : protéines et molécules, cartes génétiques, séquences de nucléotides (par exemple, la banque européenne de l'EMBL compte plus de 300000 séquences nucléotidiques). Leur accès et leur mise à jour rapide par la communauté mondiale des chercheurs au moyen du réseau Internet constitue un support important de la recherche.
- En médecine interne, la base ADM (aide au diagnostic médical, réalisée à Rennes, Prs Lenoir et LeBeux) contient la description de 2500 maladies.

DOSSIERS MEDICAUX

Les banques de données constituées à partir de dossiers **standardisés** de malades correspondent plutôt aux applications de type registre, axées sur l'épidémiologie d'une pathologie (cancer), sur un suivi de patients atteint d'une maladie chronique (hypertension ou diabète) ou sur la pratique d'une procédure thérapeutique (greffe de moelle). L'exploitation de ces bases, à condition qu'elles soient représentatives du problème étudié et que leurs données soient valides, permet d'enrichir la connaissance médicale.

BANQUES DE CONNAISSANCES

Les systèmes précédents restent au niveau des faits. Or les faits ne sont pas suffisants pour créer des connaissances et l'accumulation ne remplace pas l'organisation et la synthèse. Passer des faits aux connaissances implique de modéliser le raisonnement et de doter les bases de données de capacités de raisonnement, constituant une base de connaissances. C'est l'objet des systèmes d'aide à la décision comme ADM qui décrit les maladies, mais aussi les liens entre maladies et entre signes permettant ainsi, outre une simple description encyclopédique, une aide à la décision de type diagnostique.

INFRASTRUCTURE TECHNIQUE

Deux technologies sont employées de façon concurrente pour l'accès aux banques de

données : en ligne ("on-line") et hors ligne ("off-line").

- Dans le cas d'un accès en ligne, on utilise : un centre serveur qui stocke les informations sur un ordinateur, un réseau de communication, un poste de consultation qui peut être un terminal standard, un minitel ou un micro-ordinateur. En France, de nombreuses banques sont accessibles à moindre coût par le réseau télématique et un minitel, en contrepartie, seules des informations de type texte peuvent être diffusées en raison notamment du faible débit de la transmission. De plus en plus de banques de données sont accessibles par le réseau informatique mondial Internet dont le débit plus rapide permet la consultation et la transmission d'informations multimédia. Les serveurs de type **WWW** ("World Wide Web" ou toile d'araignée mondiale), originellement développés au CERN à Genève, contiennent des bases de données intégrant texte, images, sons ou programmes. Ces informations sont consultables sur l'ordinateur de l'utilisateur, connecté au réseau, grâce à un programme client dit navigateur, reposant sur une interface de type **hypermédia**. Un document hypermédia contient des informations textuelles mais aussi des zones sensibles, appelées boutons, dont l'activation permet l'accès à une autre information (texte du même document ou d'un autre document, image ou son) ou déclenche une action. Ces serveurs, reliés entre eux par le réseau Internet, forment un maillage, ainsi les informations recherchées peuvent être physiquement réparties sur plusieurs machines dispersées, mais le consultant n'a pas besoin de le savoir car l'accès est totalement unifié et transparent. Les résultats d'une recherche peuvent être, bien sûr, consultés sur écran mais souvent aussi imprimés ou téléchargés sur l'ordinateur du consultant. L'avantage des banques accédées en ligne réside dans leur mise à jour continue et dans l'accès à l'intégralité de la base.
- La diffusion des informations de la banque de données peut aussi se faire au moyen de supports physiques : les disquettes et maintenant essentiellement le **CD-ROM** qui contient 650 Mo de données. Il s'agit alors d'un abonnement avec une mise à jour périodique, mensuelle ou annuelle. L'avantage réside dans l'accès permanent et généralement plus aisé à l'information, par contre l'information peut être segmentée en plusieurs disques, même si, à côté des disques périodiques contenant une information intégrale, on dispose de disques de compilation regroupant les résumés de plusieurs années. L'accès simultané à plusieurs banques nécessite l'utilisation d'une chaîne de plusieurs lecteurs de CD-ROM.

REMARQUES

- Toute banque de données, même bien élaborée sur le plan conceptuel, ne vaudra que par la validité des données qu'elle contient, ceci est particulièrement vrai des informations glanées sur Internet dont la fiabilité est variable ;
- Quelle que soit la souplesse du langage et de l'organisation technique d'une banque de données, certaines réponses sont plus difficiles à obtenir que d'autres ; l'interrogation isotrope (non prévue à la conception) reste un idéal ;
- Dans le cas d'une banque de dossiers médicaux, sauf si elle est exhaustive, il se pose en général des problèmes de représentativité de l'échantillon sur lequel ont été recueillies les

données. On peut en rapprocher le problème de la domination culturelle des banques de données nord-américaines qui entraîne une sous-représentation des activités des autres pays et un biais de prise en compte de certains sujets.

- Les banques de données sont comme les banques : elles concentrent des notions dont l'accumulation et l'organisation sont précieuses et qui sont le fruit du travail de nombreuses personnes. Il s'agit donc de les protéger de toute utilisation erronée ou frauduleuse. En cas de données médicales individuelles, a fortiori nominatives, se pose en outre le problème de la confidentialité des informations privées.
- Les systèmes de consultation, qu'il s'agisse d'accès par réseau ou de CD-ROM sont pleinement opérationnels sur les ordinateurs de bureau qui disposent également de la puissance nécessaire aux applications interactives multimédia. L'accès aux connaissances sur de petits systèmes nomades (organiseurs personnels) bénéficie du développement des réseaux de télécommunication sans fil et de la miniaturisation des mémoires secondaires.
- Une nouvelle pratique de la formation et de la décision médicale, la médecine basée sur le niveau de preuve ("**evidence-based medicine**", d'origine nord-américaine), repose sur la revue exhaustive et systématique de la littérature et l'évaluation critique des publications. Elle doit permettre de dégager, par l'analyse standardisée des essais cliniques (publiés ou non), ce qui est scientifiquement prouvé à un moment donné (efficacité d'un médicament, intérêt d'un test). Des collaborations internationales (comme la collaboration Cochrane) se sont données comme objectif de produire et de diffuser largement ces informations synthétiques.

AIDE A LA DECISION

Les systèmes d'information, les bases de données, les dossiers informatisés, facilitent la prise de décision en améliorant l'accès aux données pertinentes et leur mise en perspective. Mais il ne s'agit que d'une aide indirecte présentant des faits sur lesquels le décideur doit appliquer un raisonnement. Les systèmes d'aide à la décision ont l'ambition d'assister l'homme, voire de le suppléer, en remplaçant ou en reproduisant le raisonnement humain.

POSITION DU PROBLEME

Une décision suppose la confrontation et l'application d'un modèle de connaissance à un cas du monde réel dans le but d'effectuer un choix. Trois types d'informations entrent en jeu :

- les faits observés ;
- les connaissances théoriques, le savoir ;
- l'expérience acquise au cours de l'exercice d'une activité ; c'est à dire le savoir-faire.

La difficulté de la prise de décision, en particulier en médecine et en santé publique, vient de la **situation d'incertitude**, qui tient à plusieurs raisons :

- l'incertitude sur les connaissances : certaines connaissances sont d'ordre statistique

(fréquence des maladies ou des signes) et sont associées par nature à un risque d'erreur, mais d'autres connaissances sont incomplètes, par défaut d'exploration ou par insuffisance de conceptualisation (physiopathologie) ;

- l'incertitude sur les faits : la description de l'état présent n'est jamais parfaite, soit par manque de moyens ou de temps (urgence), soit par défaut de mesure ou mauvaise interprétation d'un symptôme, d'un signe ou d'un résultat ;
- l'incertitude du langage : le flou et l'ambiguïté des notions manipulées perturbent le traitement et la transmission de l'information.

D'autre part, la décision, bien que portant sur un objet précis dans le cadre d'un domaine scientifique déterminé, ne peut s'abstraire de l'environnement (psychologique, social, culturel, économique) de l'objet d'étude ou de l'observateur.

CARACTERISTIQUES DES SYSTEMES

Plusieurs caractéristiques permettent de décrire les systèmes d'aide à la décision : le type de problème, le mode d'intervention, la nature de l'interaction, le domaine d'études ou enfin la méthode employée, qui sera développée à part.

Les systèmes d'aide à la décision peuvent s'appliquer à deux types de problèmes :

- un problème de **classement** ou de diagnostic (médical ou autre) dont le but est de séparer ce qui est de ce qui n'est pas, compte tenu de l'incertitude sur la situation réelle de l'objet d'étude (patient, organe, population) ;
- un problème d'**optimisation** dont le but est d'indiquer la démarche la plus efficace (par exemple, une stratégie thérapeutique) compte tenu de l'objectif et de contraintes (coût, risque, difficulté, environnement).

Plusieurs modes de fonctionnement de ces programmes informatiques sont possibles, chaque système ne fonctionnant généralement que dans un mode.

- le mode **passif** est le plus fréquent, il suppose l'intervention explicite de l'utilisateur pour décrire le problème (par exemple, l'état du patient) et interroger le système. On distingue deux types de comportement : un système consultant fournit en retour une conclusion ou un conseil (par exemple un diagnostic ou un traitement) tandis qu'un système critique demande que lui soit décrite la stratégie envisagée par le décideur humain, ce qui lui permet de la commenter ou de la critiquer en indiquant les failles du raisonnement.
- le mode **semi-actif** correspond à un système dont le déclenchement automatique répond à une intervention humaine. L'objectif est de jouer le rôle de garde-fou en rappelant en temps réel des informations ou des règles indiscutables : système de rappel automatique qui permet d'éviter des prescriptions inutiles, contre-indiquées ou exposant à des interactions, système d'alarme qui alerte sur un changement d'état du patient.
- le mode **actif** est celui d'un système à déclenchement automatique et autonome. Il agit sans intervention du décideur selon une boucle de rétro-contrôle pour actionner un système de traitement ou de surveillance.
- Dans le cas d'un système passif ou semi-actif, deux modes de dialogue sont possibles : un

mode interactif où s'enchaînent questions et réponses (l'utilisateur renseignant et questionnant le système qui demande des renseignements et émet des conclusions), et un mode programmé où après avoir reçu un bloc d'informations initiales, le système n'émet qu'une réponse finale. Le support du dialogue est généralement constitué de textes ou de nombres, mais de plus en plus souvent, il met en jeu des informations plus riches et plus complexes, de type signal, image ou son.

Si l'aide au diagnostic médical est le premier sujet évoqué, les systèmes informatiques d'aide à la décision ont comme objectif général de modéliser un système réel ou son comportement afin de prédire son état présent ou à venir. Ce système peut donc être un individu, malade ou non, mais aussi un organe isolé, une population d'individus, ou une organisation (hôpital).

METHODES

Différentes méthodes sont susceptibles d'être employées, de façon exclusive ou combinée.

METHODES MATHEMATIQUES

La modélisation mathématique concerne généralement la simulation du comportement de populations, de systèmes biologiques ou physiologiques. On distingue deux grandes catégories de modèles. Les modèles **déterministes** décrivent l'évolution de concentrations ou de quantités continues à l'aide de fonctions mathématiques et de systèmes d'équations différentielles. Les modèles **stochastiques** s'intéressent au comportement d'objets individualisés (personnes ou molécules) qu'on ne peut connaître avec certitude mais qui obéissent à des lois de probabilités connues ; la combinaison des comportements individuels aléatoires induit le comportement du système global et la génération répétée de séries aléatoires par l'ordinateur permet d'obtenir la distribution des variables d'intérêt. Quel que soit le formalisme mathématique sous-jacent, les résultats que fournit le modèle en sortie, en réponse aux données d'entrée, doivent être **validés**, c'est-à-dire comparés à ceux observés dans la réalité pour les mêmes valeurs des paramètres d'entrée. Sous réserve d'une bonne adéquation, le modèle peut alors être utilisé pour décrire le système réel mais surtout pour prédire son évolution future ou son comportement face à des modifications des paramètres d'entrée, ces modifications pouvant correspondre à des phénomènes naturels ou à des interventions humaines.

Les modèles mathématiques peuvent servir à la décision médicale ou de santé publique, en mode passif ou en mode actif (contrôle automatique). Leurs applications sont donc nombreuses et variées :

- modèle pharmacocinétique : il permet de représenter et de quantifier les différentes phases du métabolisme des médicaments (absorption, diffusion, transformation en métabolites actifs ou non, élimination). Le modèle peut être construit sur les données d'une population ou d'un individu. Il est utilisé pour adapter au mieux la posologie, en particulier dans le cas

de médicaments à fenêtre thérapeutique étroite.

- modèle épidémiologique : il a pour objectif la représentation de l'évolution dans une population d'une maladie, souvent contagieuse, et des facteurs influençant cette diffusion dans le temps ou dans l'espace. Lorsqu'un tel modèle a été validé, il permet d'une part d'avoir une idée des facteurs de risque ou de protection possibles et de leur importance relative, mais surtout il fournit une estimation de l'efficacité potentielle des diverses mesures sanitaires envisageables. La pandémie à VIH est l'objet de nombreux modèles de ce type.

METHODES NEUROMIMETIQUES

Les méthodes neuromimétiques ou connexionnistes sont inspirées des structures neuronales et du fonctionnement cérébral d'où le nom fréquent de réseau de neurones formels. Le réseau neuronal est un programme qui met en jeu des nœuds ou neurones formels reliés entre eux par des arcs, équivalents des axones et des dendrites. Chaque neurone réalise la sommation des stimuli des neurones afférents, chaque connexion étant affectée d'une pondération. Cette sommation obéit à différentes formes de fonction (tout ou rien, fonction en escalier ou sigmoïde) et si le résultat dépasse un certain seuil d'activation, alors le neurone envoie un stimulus aux neurones efférents. Il existe différents modèles de réseaux, notamment selon le type d'interconnexions des neurones. La mise au point du réseau s'effectue par modification automatique des pondérations des connexions au cours d'un processus d'apprentissage, en fonction de l'écart observé entre la sortie observée et la sortie attendue.

Les systèmes neuronaux ont trouvé leur premières applications dans le domaine de la reconnaissance de formes (caractères, images, sons), en effet, on peut assimiler la couche d'entrée à une rétine. Ils sont bien adaptés aux problèmes de classification diagnostique à condition que l'on dispose d'une base de cas suffisante en nombre et en variété pour le processus d'apprentissage. Dans cette situation, la couche d'entrée correspond aux signes et la couche de sortie aux diagnostics.

L'usage d'un réseau de neurones permet de ne pas spécifier un modèle mathématique théorique, mais en contrepartie le réseau est un modèle **empirique** qui ne peut explicité.

METHODES STATISTIQUES

Les méthodes statistiques concernent essentiellement les méthodes **de régression ou de classification multidimensionnelles** qui permettent d'expliquer la valeur d'une réponse ou l'appartenance à un groupe en fonction des valeurs de plusieurs variables dites explicatives. Il s'agit notamment de l'analyse discriminante, de la régression logistique dans le cas d'une réponse qualitative (groupe), de la régression multiple dans le cas d'une réponse quantitative, ou du modèle de Cox dans le cas d'une variable de réponse censurée (analyse de survie), méthodes qui sont toutes dites paramétriques car imposant des contraintes de distribution (par exemple gaussienne) ou de forme de liaison, alors que la régression ou la classification par

arbre (méthode CART : classification and regression tree) est une méthode non-paramétrique. Les variables explicatives peuvent être qualitatives et quantitatives.

Ces méthodes répondent à deux objectifs : **réduire** le nombre souvent considérable de variables potentiellement explicatives à un sous-ensemble d'usage ou d'obtention plus aisé, et obtenir une **équation prédictive**.

Dans le cas d'un modèle paramétrique, le modèle se présente sous la forme d'une équation mettant en rapport la réponse ou le groupe (diagnostique ou pronostique) et les variables explicatives affectées de coefficients. Dans le cas de la méthode CART, on obtient un arbre qui répartit l'échantillon de départ en sous-groupes de plus en plus homogènes vis-à-vis de la réponse en fonction de la valeur des variables explicatives. Dans tous les cas, un premier échantillon, dit de calcul, permet de déterminer quelles sont les variables explicatives, leur forme, leurs interactions éventuelles et la valeur de leurs coefficients (modèle paramétrique). Puis la performance du modèle obtenu est évaluée sur un deuxième échantillon, **l'échantillon test**, qui doit répondre aux mêmes critères de représentativité que le premier, afin d'avoir une estimation non optimiste de la capacité de prédiction ou de classification. On peut ensuite procéder à l'utilisation effective du modèle sur un nouveau sujet dont les valeurs des variables explicatives sont connues mais dont la réponse est inconnue.

Ces modèles sont parmi les plus fréquents mais plusieurs remarques doivent être faites : ils nécessitent pour leur mise au point des échantillons de taille souvent importante, constitués dans le cadre d'un protocole d'enquête avec les contraintes de représentativité et d'absence de biais déjà mentionnées ; les modèles paramétriques sont pénalisés dans le cas de valeurs manquantes ; ils s'appliquent à des problèmes diagnostiques, thérapeutiques ou pronostiques généralement restreints ; enfin leur forme n'est pas toujours très parlante pour le clinicien.

METHODES ALGEBRIQUES

Les méthodes algébriques font d'abord appel à l'algèbre de **Boole** ou algèbre binaire qui concerne les variables ne pouvant prendre que deux états : vrai/faux ou présent/absent. L'utilisation de l'algèbre de Boole vise à reproduire le raisonnement médical et à formaliser la connaissance au moyens d'arguments binaires, en Oui ou Non, et d'opérateurs logiques : Non, Et (Syndrome néphrotique = protéinurie et hypoprotéinémie), Ou (Eruption rougeole = exanthème ou Koplick), Implique (Signe de Babinski => syndrome pyramidal). Il est alors possible en principe de décrire les maladies par la présence ou l'absence de signes, et les règles de raisonnement clinique par des implications.

Cependant, l'utilisation du formalisme booléen pose des problèmes. Il implique une rigidification de la pensée et l'élimination des diverses formes d'incertitude, fréquentes en médecine. Enfin, la décision suggérée par un tel système est, elle aussi, booléenne, c'est-à-dire en tout ou rien alors qu'elle est toujours plus ou moins probable en réalité.

En dépit de ces simplifications extrêmes, plusieurs systèmes ont été construits sur des

principes booléens. L'usage le plus fréquent de l'algèbre de Boole a consisté à développer des petits algorithmes booléens, sous la forme d'arbres ("guidelines") décrivant la démarche diagnostique ou thérapeutique, destinés aux auxiliaires médicaux des pays en développement et leur permettant de résoudre les problèmes de prise en charge les plus fréquents.

METHODES PROBABILISTES

Les méthodes probabilistes, reposant sur l'application du **théorème de Bayes**, permettent de calculer la probabilité d'une maladie connaissant les signes du malade. L'application sur un problème de diagnostic médical du théorème de Bayes, encore appelé théorème d'inversion des probabilités, établit que la probabilité a posteriori d'un diagnostic D_i , lorsque le signe S est présent, est fonction de la probabilité a priori du diagnostic et de la probabilité conditionnelle d'observer le signe lorsque le diagnostic est présent :

$$P(D_i/S) = P(D_i) \cdot P(S/D_i) / P(S) = P(D_i) \cdot P(S/D_i) / (\sum_j P(D_j)P(S/D_j))$$

La méthode bayésienne tient compte des signes positifs et négatifs et permet de calculer le diagnostic le plus probable.

La détermination des probabilités a priori (prévalence de la maladie) et des probabilités conditionnelles (fréquence d'un signe dans une maladie) suppose l'exploitation d'une base de données médicales correctement constituée, enregistrant de façon non biaisée les signes et le diagnostic de chaque cas. Les performances sont généralement acceptables, avec une concordance observée avec les experts de plus de 70%.

Le système bayésien le plus connu est celui développé par l'équipe de De Dombal à Leeds, qui a fait l'objet d'évaluations dans différentes situations de soins. Il s'agit d'un système d'aide au diagnostic des douleurs abdominales aiguës pouvant nécessiter un geste chirurgical. Son utilisation permet au médecin de poser plus souvent le bon diagnostic que sans assistance. D'autre part, les faux négatifs (diagnostic médical conduisant à une abstention chirurgicale entraînant une aggravation) et les faux positifs (intervention à tort) sont moins fréquents. Une expérience française similaire a permis de montrer l'importance de constituer une base de cas représentatifs.

Un intérêt du théorème de Bayes est qu'il est possible de l'utiliser de façon itérative, qu'il permet d'apprécier l'efficacité diagnostique de chacun des signes et de construire ainsi la meilleure séquence de recueil. Les limites de l'approche bayésienne classique sont liées soit à la nature même des probabilités : exhaustivité des diagnostics, soit à des simplifications calculatoires : exclusivité des maladies et indépendance des signes. L'exhaustivité est liée au fait que la somme des probabilités des maladies doit être égale à 1. Il faut donc obtenir une liste complète des maladies ou des décisions à prendre, ce qui est généralement obtenu en créant une catégorie "autre" qui n'a guère de sens clinique ou thérapeutique. L'exclusivité signifie qu'on exclut la combinaison de maladies. L'indépendance suppose que la fréquence d'un signe dans une maladie n'est pas modifiée par la présence ou l'absence d'un autre signe alors que les signes sont souvent plus ou moins redondants et liés de façon différentielle aux maladies. Pour toutes ces raisons, les systèmes bayésiens s'appliquent à des domaines

relativement restreints.

METHODES DECISIONNELLES

La décision la plus probable n'est pas toujours la plus utile à prendre. Ce problème est particulièrement net lorsque les décisions placées en tête ont des probabilités très voisines alors qu'elles ont des conséquences très dissemblables. Aussi, l'analyse de la décision s'attache à rationaliser le choix d'une action en prenant en compte le caractère aléatoire des états comme le **coût attaché aux décisions**.

La démarche comporte plusieurs étapes. On dénombre tout d'abord l'ensemble des actions possibles répondant à un problème donné, les combinaisons de ces actions permettant de définir un certain nombre de stratégies possibles. On recense de même toutes les éventualités possibles à caractère aléatoire traduisant l'état de la nature (existence d'une maladie) ou l'efficacité d'une procédure (diagnostique ou thérapeutique) et on leur affecte des probabilités, en effet, alors qu'une action est effectuée avec certitude, on reste incertain quant à la présence d'une maladie. On définit et on mesure, afin de choisir les solutions du problème, des quantités appelées utilités. Elles permettent de donner des degrés de préférence ou d'indiquer le gain correspondant des différentes stratégies selon les éventualités, par exemple "donner un traitement anticoagulant si le malade souffre d'une embolie cérébrale". La représentation adoptée est généralement celle d'un arbre de décision : depuis la racine représentant le problème partent des branches aboutissant aux nœuds "actions" qui eux-mêmes pointent sur les différentes éventualités. En bout de branche figurent les utilités. On peut alors calculer pour chaque stratégie d'action l'espérance mathématique de son utilité en pondérant le gain par la probabilité de la situation selon la formule :

$$U(A_i) = \sum_j p(E_j/A_i) \cdot U(E_j)$$

où $U(A_i)$ est l'utilité de la i ème stratégie, $p(E_j/A_i)$ est la probabilité de la j ème conséquence après avoir exécuté l'action i et $U(E_j)$ est l'utilité associée à la j ème conséquence. La solution du problème consiste alors à retenir la stratégie qui a l'espérance maximale d'utilité.

L'intérêt de la théorie de la décision réside dans la formalisation et la systématisation de la réflexion, la prise en compte simultanée de la connaissance et de facteurs éthiques ou de qualité des soins. Cependant, elle souffre de plusieurs limites. La première semble être la difficulté de la détermination des utilités : si le couple survie/décès peut se traduire par 1/0, la plupart des situations sont plus nuancées et si on ne dispose pas d'utilités objectives comme des probabilités de survie, il faut faire appel soit au jugement clinique, soit au jugement du patient lui-même (qualité de vie). On peut également remplacer les utilités par des coûts mais il semble délicat d'introduire dans le même modèle des notions de qualité de vie et de coût financier, et mieux vaut alors réaliser deux analyses distinctes. Enfin, ces méthodes simplificatrices ne s'appliquent qu'à des problèmes relativement limités.

METHODES SYMBOLIQUES

Toutes les méthodes précédentes sont dites numériques car elles ne font que manipuler

des chiffres sans réel raisonnement. En ce sens, malgré leurs bonnes performances, ce sont des "boîtes noires" fournissant un résultat mais incapables d'expliquer le pourquoi et le comment de celui-ci et limitées à des problèmes réduits et stables. Ces limites conceptuelles expliquent l'espoir mis depuis les années 70-80 dans les méthodes symboliques d'**intelligence artificielle**.

L'intelligence artificielle est une discipline informatique dont l'objectif est de faire traiter par l'ordinateur des problèmes usuellement résolus par l'homme et dont la solution exige des connaissances, de la perception, du raisonnement, de l'apprentissage, car ces problèmes n'ont pas de solution algorithmique théorique ou pratique. Il s'agit de la reconnaissance de formes, de l'analyse du langage naturel et des systèmes experts.

Les **systèmes experts** sont des logiciels de résolution de problèmes. Ils permettent de représenter sous forme explicite et déclarative les connaissances et le comportement d'un expert humain, afin qu'il puisse être reproduit par un programme. Deux types de notions doivent donc être formalisées : un ensemble de connaissances théoriques ou expérimentales et le raisonnement qui permet de les utiliser. Les premières seront gérées dans une base de connaissances tandis que le second sera réalisé sous la forme d'un programme interpréteur de connaissances : le moteur d'inférences. Les autres modules du système expert sont la base de faits qui contient la description du problème à résoudre et l'interface homme-machine, en charge du dialogue avec l'utilisateur. Cette conception modulaire permet un développement de la base de connaissances sans refonte du reste du système. D'autre part, le **caractère explicite des connaissances** permet au système d'expliquer comment il a atteint telle conclusion ou pourquoi il demande telle information à l'utilisateur.

Analyse du raisonnement en médecine

L'objectif général de la démarche médicale à visée décisionnelle est de réduire au maximum l'incertitude quant à l'état actuel ou futur du patient par l'acquisition d'informations complémentaires, acquisition guidée par la mise en œuvre de connaissances multiples et complexes.

Différents modes de raisonnement sont possibles :

- le raisonnement déductif obéit aux principes de l'implication logique et permet de passer des prémisses à leur conséquence par l'application d'une règle du type "si prémisses alors conclusion" ; il permet également de passer du cas général au cas particulier ;
- le raisonnement inductif permet d'avancer une loi générale à partir de l'observation de faits particuliers ;
- le raisonnement par abduction part des observations pour en évoquer les causes possibles ;
- le raisonnement par défaut permet de supposer une valeur possible pour les données temporairement manquantes et de la remettre en cause lors de l'acquisition de nouvelles informations ;
- le raisonnement causal s'appuie sur l'analyse détaillée de la chronologie des faits, l'imputation des effets à la cause nécessitant l'antériorité de la cause, un délai d'apparition

compatible avec la connaissance, et au mieux la réversibilité de l'effet et une relation quantitative entre la cause et l'effet.

Ces raisonnements peuvent s'enchaîner au cours de plusieurs étapes, selon une démarche globale hypothético-déductive :

- acquisition de données ;
- formulation d'hypothèses ; cette phase est conditionnée par l'expérience du praticien qui lui permet de se focaliser rapidement sur les bonnes hypothèses (abduction) ;
- interprétation des données cliniques où le clinicien doit sélectionner les informations significatives parmi toutes celles disponibles ; élaboration d'une stratégie de résolution de problème comportant la recherche de données pertinentes ;
- le choix de la solution s'effectue sur les hypothèses réévaluées à la lumière des nouveaux renseignements obtenus ; en l'absence de choix unique expliquant toutes les manifestations ou d'une conclusion ayant une certitude suffisante, le processus réitère sur l'acquisition de données ; souvent, cependant, il se produit avant la fin du processus diagnostique un échappement vers la phase de décision thérapeutique.

Représentation des connaissances

De multiples connaissances sont nécessaires à la décision : anatomie, biologie, physiopathologie, sémiologie, épidémiologie, pharmacologie, etc. Elles s'organisent selon deux modèles : connaissances empiriques concernant les associations entre signes et maladies ; connaissances théoriques basées sur un modèle profond, physiologique et physiopathologique. Différents formalismes de représentation permettent de les exprimer au mieux dans le système :

- les règles de production expriment de petites quantités de connaissances de la forme "si prémisses alors conclusion" , les prémisses correspondant généralement à des signes et les conclusions à des syndromes, des hypothèses diagnostiques ou des actions, par exemple "si Koplick alors rougeole". Certains systèmes associent des coefficients de crédibilité à la règle afin d'attribuer un degré de certitude à la conclusion. Les règles sont bien adaptées à la représentation d'une connaissance dynamique : conduite à tenir.
- les objets structurés (en anglais "frames") permettent de décrire des concepts médicaux complexes en précisant leurs caractéristiques (rubriques) et les moyens de les évaluer ou de les utiliser (fonctions ou règles locales). Ces objets peuvent être regroupés en classes partageant des caractéristiques communes. Ils sont bien adaptés à la représentation d'une connaissance statique : maladies ou traitements.

Exemples

- Mycin est le premier système expert en médecine, il est encore le plus cité bien qu'il n'ait jamais été utilisé en routine. Initialement conçu pour le diagnostic et le traitement des méningites infectieuses, il a donné lieu à toute une série de développements théoriques : isolement du moteur d'inférences sous la forme d'un moteur essentiel qui a été appliqué à

d'autres bases de connaissances comme Oncocin en chimiothérapie ou développement d'un système à vocation pédagogique dénommé Guidon ;

- Internist est un des systèmes à vocation large : la médecine interne, avec 600 maladies et 4500 signes ; chaque maladie est décrite par des signes dotés d'un coefficient de sensibilité et de spécificité ; les performances de ce système ont été évaluées sur des cas cliniques du New England Journal of Medicine ; cependant il est inutilisable en pratique en raison notamment du temps de consultation ; une version simplifiée et à vocation didactique, QMR, est maintenant disponible sur micro-ordinateur ;
- Sphinx reste le système expert français le plus connu ; il utilise un formalisme mixte, objets structurés pour les concepts médicaux et règles de production pour le raisonnement ; il a été appliqué dans le domaine du diagnostic des ictères et de la thérapeutique du diabète, où il a été évalué auprès de médecins généralistes ;
- le système Help est l'exemple d'un système d'aide à la décision intégré à un SIH ; il fonctionne en mode semi-actif, la mise à jour des données du dossier du patient déclenchant les modules d'aide à la décision ce qui permet la réalisation d'alarmes intelligentes dans le cadre de la prescription thérapeutique (contre-indications, interactions), notamment d'antibiotiques (détection d'infections nosocomiales, de résistances), ou de produits sanguins.

Remarques

Les systèmes experts trouvent leur indication principale dans la résolution de problèmes individualisés dans des domaines spécialisés où la disponibilité d'experts reconnus rend possible la constitution et la maintenance des bases de connaissances. Ils permettent ainsi de mettre à disposition une expertise là où elle n'existe pas.

Les limites tiennent aux contraintes de formulation de la connaissance (connaissance imparfaite, relations multiples, causales, physiopathologiques ou temporelles, difficulté de formalisation du raisonnement, contrainte du dialogue avec l'utilisateur, nécessité de consensus). D'autre part, le système part d'une information déjà interprétée par l'utilisateur, il lui est difficile de détecter qu'il sort de son domaine d'expertise.

MISE EN PLACE DES SYSTEMES D'AIDE A LA DECISION

En pratique, le nombre de systèmes d'aides à la décision utilisés en routine reste faible. En effet, au-delà des limites liées à chaque méthodologie, plusieurs contraintes pèsent sur la diffusion de ces systèmes.

Interface homme-machine

L'interface homme-machine, qu'il s'agisse de la disponibilité du système ou de la possibilité d'employer le langage naturel pour les questions ou les explications, conditionne largement l'acceptabilité des systèmes interactifs. Ceux-ci doivent être assez réactifs et assez simples d'emploi pour s'intégrer dans les conditions de travail et ne pas nécessiter de

formation prolongée. Dans le cas d'un système d'aide à la décision intégré à un SIH, il ne doit pas y avoir ressaisie d'informations.

Constitution des bases de connaissances

La plupart des systèmes d'aide à la décision numériques nécessitent une base de dossiers médicaux valides permettant le calcul des paramètres du modèle (probabilités, coefficients, etc...). La qualité du système dépend donc de la fiabilité des informations saisies et de la représentativité des dossiers vis-à-vis du problème à résoudre. Dans le cas des systèmes experts, les informations permettant de résoudre le problème ne sont pas obtenues par quantification, mais sont l'expression du jugement d'un ou plusieurs experts humains. Or l'ensemble des connaissances d'un domaine peut être difficile à représenter avec un formalisme nécessairement simplificateur, d'autant que la base de connaissances doit être complète et non contradictoire. L'exploitation d'une base de dossiers permet parfois d'épauler la constitution de la base de connaissances. D'autre part, la connaissance qu'exprime le système doit faire l'objet d'un consensus de la part de la communauté médicale.

Evaluation et validation

L'évaluation des systèmes d'aide à la décision recouvre plusieurs aspects. La performance pure d'un système d'aide au diagnostic s'exprime souvent par des notions statistiques de pourcentage de sujets bien classés ou de concordance avec la référence (étalon-or ou "gold standard") et en cela ne diffère guère de celle d'un examen complémentaire. Parfois, les experts du domaine ne sont pas d'accord sur la solution à adopter. Enfin, le critère de bon classement ne s'applique pas aux problèmes d'optimisation (par exemple, thérapeutique) où une ou plusieurs décisions peuvent être jugées comme optimale, correcte, inefficace ou dangereuse. La sécurité des systèmes (fiabilité, protection contre les erreurs de manipulation, absence de résultats dangereux) est bien sûr une contrainte majeure dès lors que le pronostic vital peut être engagé. Enfin, le système doit être évalué sur le terrain, ce qui permet de réintroduire les critères d'ergonomie, d'impact sur la prise en charge et de pertinence vis-à-vis des objectifs.

Conclusion

Tous les systèmes d'aide à la décision à méthodologie numérique et la majorité des systèmes symboliques ne traitent que des problèmes préalablement identifiés par le clinicien, limités et structurés. Ceci est particulièrement criant pour les systèmes passifs qui, de ce fait, à moins d'englober idéalement toute la médecine, sont d'un intérêt pratique limité. Au contraire, de multiples petits systèmes actifs (alarmes intelligentes) peuvent s'intégrer dans un système d'information plus vaste.

En terme de finalité, au delà de l'amélioration de la qualité des décisions dans les centres hospitalo-universitaires où ces systèmes sont pour l'essentiels conçus, l'objectif devrait être d'apporter la connaissance de pointe là où elle n'est pas accessible facilement :

auprès du non-spécialiste, de l'auxiliaire médical, des malades chroniques, mais aussi des étudiants. Le principal bénéfice de ces systèmes vient plus de l'évaluation de la qualité de l'information et de sa pertinence, de la formalisation du raisonnement et de la méthode qu'ils supposent que de leurs résultats immédiats.

SYSTEME D'ENSEIGNEMENT

Si, comme on le dit souvent, la répétition est à la base de la pédagogie, les ordinateurs, automates dotés de mémoire, devaient logiquement occuper le terrain de l'enseignement. En fait, une formation efficace nécessite plus que des connaissances factuelles, mais, ici comme ailleurs, une méthode.

PEDAGOGIE

Les travaux sur l'enseignement assisté par ordinateur (EAO) bénéficient des recherches en sciences cognitives afin d'identifier les techniques pédagogiques utiles. On distingue classiquement deux types de connaissances : les connaissances académiques et les connaissances d'expérience. L'enseignement traditionnel est plus axé sur la transmission des théories et des faits. L'apprentissage des savoir-faire (heuristiques) s'effectue mieux à partir d'expériences pratiques, préalables à l'exercice réel. D'autre part, alors que la durée des études est nécessairement limitée, le volume de connaissances utiles ne fait qu'augmenter et la durée de vie de ces connaissances raccourcit. Il importe donc d'apprendre à apprendre. La pédagogie doit donc aider l'apprenant à définir des objectifs de formation et à les réaliser, individualiser l'enseignement, favoriser le travail individuel pratique (acquisition de savoir-faire) et de réflexion (intégration des connaissances, développement de stratégies).

INTERET DE L'EAO

Dans cette optique, les logiciels d'EAO ou didacticiels visent à compléter l'enseignement traditionnel :

- ils permettent une plus grande individualisation de l'enseignement (rythme, sujet, répétition, disponibilité) ;
- l'acquisition de connaissances se fait sur un mode interactif, éventuellement à connotation ludique ;
- la combinaison des supports (textes, images, sons) en environnement multimédia renforce le message et favorise la mémorisation ;
- l'informatique permet la mise en place d'environnements de simulation ;
- le processus de contrôle peut être intégré au processus d'apprentissage dans un but de test ou d'adaptation du programme ;
- l'exploitation statistique des séances d'utilisation des didacticiels permet de les faire évoluer.

METHODES

Les didacticiels s'adressent aux différentes phases de la transmission de connaissances : exposition, répétition, contrôle, mise en application.

Enseignement tutoriel

Il s'agit de renforcer l'enseignement traditionnel au travers de programmes relativement simples basés sur des QCM, des études de cas ou des banques d'images;

- les enseignements programmés traditionnels de type QCM décomposent les connaissances à transmettre en petits blocs enchaînés dans un ordre préétabli par le concepteur du didacticiel. La progression de l'étudiant d'un bloc au suivant dépend de la bonne réponse à la question courante ou d'un mécanisme d'échappement, l'adaptation au profil de l'apprenant est donc modeste.
- les enseignements du type "étude de cas" reposent sur la résolution par l'étudiant d'un problème fictif au fur et à mesure d'étapes prédéterminées, pouvant comporter l'utilisation de données multimedia. Ceci revient à parcourir un arbre depuis la racine (exposé du problème) jusqu'à un noeud terminal (succès), chaque étape devant correspondre à la résolution d'un sous-problème. Cette méthode place apparemment l'étudiant en position de simulation. Cependant le parcours est là encore prédéterminé par l'enseignant lors de la conception du didacticiel et le nombre d'interventions possibles de l'étudiant lors d'une étape est fortement contraint. Ces programmes peuvent facilement être couplés à une procédure de contrôle des connaissances.
- les banques d'images peuvent se limiter à offrir un accès facile à un ensemble sélectionné d'images de références, avec possibilité de sélection multicritère. De plus en plus, elles prennent un caractère interactif, grâce aux techniques de programmation hypermédia. Ainsi, le programme ADAM présente l'anatomie humaine sous la forme d'une dissection interactive, offrant un environnement propre à développer une exploration active de la part de l'étudiant.

Modélisation et simulation

La mise en pratique des connaissances théoriques est l'objectif des systèmes de simulation. Ces simulations peuvent être statiques quand le problème présenté n'évolue pas en cours de simulation alors que les simulations dynamiques où l'état du patient évolue en fonction du temps et des interventions de l'étudiant permettent une interaction plus riche. Elles peuvent concerner des phénomènes biologiques (circulation, respiration, équilibre hydro-électrolytique) complexes où les interventions de l'étudiant permettent de lui faire saisir la dynamique du système, notamment au travers de représentations graphiques. Les simulations portent également sur des cas cliniques. Ces cas, posant des problèmes diagnostiques ou thérapeutiques, peuvent être préenregistrés dans une base de cas ou générés par programme. Ils sont souvent associés à des banques d'images ou de sons. Lors de l'exercice, le programme fournit les réponses correspondant aux demandes d'examen de

l'étudiant. Par rapport à l'enseignement tutoriel, l'étudiant a une plus grande latitude de choix puisqu'il n'y a pas de chemin de résolution préenregistré. Ainsi le système "Iliad" est basé sur un système expert en médecine interne qui évalue la démarche décisionnelle en fonction des probabilités a priori des maladies, de la valeur prédictive des signes et du coût des examens.

EAO "intelligent"

De même que les systèmes experts ont pour objectif de résoudre un problème du monde réel à l'imitation de l'homme grâce à une base de connaissances du domaine, les didacticiels "intelligents" ont pour vocation d'offrir un enseignement automatique mais réellement individualisé grâce à la mise en oeuvre de connaissances pédagogiques. Il s'agit donc de systèmes experts pédagogues mais aussi experts du domaine, selon le principe que l'on ne peut enseigner que ce qu'on l'on sait. Le prototype en fut Guidon, version pédagogique du système expert Mycin, dont l'évaluation permit de montrer qu'il ne suffit pas d'être un expert du domaine pour être un enseignant du domaine : ainsi, la connaissance de l'expert est efficace mais elle est souvent trop condensée pour être transmise en l'état, elle a besoin d'être décomposée et argumentée pour être comprise. Les problèmes de modélisation de la connaissance sont évidemment plus complexes puisque s'ajoutent à la connaissance du domaine, nécessaire pour générer le cas simulé et proposer la démarche optimale, une connaissance pédagogique, pour adapter l'interaction ou la difficulté, et un modèle de l'étudiant, qui est une carte des concepts maîtrisés ou non par l'étudiant. Enfin, ces systèmes, objets de travaux de recherche, nécessitent une interface homme-machine d'une grande souplesse.

REMARQUES

Les didacticiels ne permettent pas encore d'exprimer ou de simuler dans toute leur complexité les relations enseignant-enseigné, étudiant-patient ou médecin-malade. Aussi doivent-ils s'intégrer dans un projet pédagogique à côté des formations et apprentissages classiques, chaque composante devant être évaluée. Les EAO ont l'avantage de laisser une part plus importante d'autonomie à l'étudiant et d'offrir de plus en plus souvent un environnement multimédia attractif, cependant, cela peut parfois aboutir à une navigation entre des bribes de connaissances, sans fil conducteur ni mise en perspective.

UTILISATEURS ET POPULATIONS CIBLES

L'objectif général des systèmes d'aide à l'action médicale est d'assurer des soins de meilleure qualité, c'est-à-dire d'élever le niveau de la santé individuelle et collective, de façon immédiate ou non. On peut ainsi recenser diverses catégories d'utilisateur :

- médecin généraliste : accès aux bases de données, télémedecine, formation continue.
- médecin hospitalier : accès aux bases de données, aide à la décision, SIH intelligent.

- malade : éducation vis-à-vis du traitement (diabète), surveillance automatique du traitement (pompe à insuline ou pace-maker), aide au traitement (diabète ou HTA).
- population générale : éducation (conseil aux voyageurs, pré-consultation).
- étudiants : enseignement, banques d'images.
- décideur, responsable de santé publique : aide à la décision (systèmes de simulation), banques de données.

OBJECTIFS

- Définir un système documentaire, une banque de données médicales.
- Quel est le principe d'interrogation d'un système documentaire ?
- Citer quelques systèmes documentaires utilisables en médecine, biologie ou santé publique.
- Quels sont les critères permettant d'évaluer un système documentaire ?
- Décrire les moyens techniques permettant d'accéder à une banque de données.
- Décrire les principales méthodes d'aide à la décision en médecine et santé publique.
- Décrire les principales modalités d'intervention d'un système d'aide à la décision.
- Indications, faisabilité et limites des systèmes d'aide à la décision, notamment des systèmes experts, en médecine.
- Donner les avantages et les inconvénients de l'EAO par rapport à l'enseignement traditionnel.
- Définir les différents types d'EAO.
- Citer les usages des systèmes d'aide à la décision en fonction des acteurs du système de soins.

ASPECTS DEONTOLOGIQUES

La profession médicale touche à l'intimité de la personnalité humaine tant sur le plan physique que moral. On conçoit donc qu'un certain nombre de règles doivent être observées par les médecins pour préserver la vie privée des patients. De plus, l'exercice de la médecine impose souvent des choix difficiles.

Depuis le début de l'exercice de la médecine, dans l'antiquité, cet aspect de la profession médicale a été mis en évidence. Ainsi le serment d'Hippocrate, 400 ans av. J-C. répond à ce type de problème. Les principes à respecter et les solutions pratiques édictées par Hippocrate et son école ont survécu jusqu'à ce jour, car ils étaient et sont toujours adaptés à l'exercice de la médecine. Les principales mises à jour des règles régissant l'exercice de la médecine sont relativement récentes et apparaissent liées d'une part à l'augmentation brutale de la démographie médicale, d'autre part à l'évolution des techniques d'exploration des malades. Néanmoins, les principes de base demeurent. Tous peuvent se résumer par la phrase suivante: "le médecin exerce son métier dans le seul but de soulager son patient, et dans le strict intérêt de ce dernier".

D'autre part, l'émergence de l'informatique et la création de nombreux fichiers nominatifs engendrent de nouveaux risques pour la confidentialité et la protection des données, notamment médicales. Ils ont nécessité une législation générale ainsi qu'une adaptation récente de la réglementation de l'exercice médical.

Enfin, un ensemble de mesures organisationnelles, matérielles et logicielles permet d'assurer au mieux l'intégrité des données informatiques et la sécurité. L'intégrité des données signifie que l'information est toujours disponible à ceux qui y ont droit tandis que la sécurité signifie que l'information n'est jamais disponible à ceux qui n'y ont pas droit. Cependant il faut rester conscient du fait que la protection absolue est impossible car contradictoire avec l'usage des données. Au delà de ces deux exigences et les justifiant, se situe l'intérêt du malade.

CADRE DEONTOLOGIQUE ET LEGAL

Le texte réglementaire régissant la profession médicale est le code de déontologie médicale. Il comporte les principales règles qui doivent présider à l'exercice de la médecine. La dernière version figure dans le décret 95-1000 du 6 septembre 1995.

D'autre part, le secret médical, c'est-à-dire la protection de la confidentialité des informations connues d'un médecin au sujet de son patient, est inclus dans la notion de secret professionnel qui est protégé par le code pénal (article 378).

Enfin, pour l'aspect informatique, une loi (loi n°78.17 du 6 Janvier 1978), relative à l'informatique, aux fichiers et aux libertés, précise de nombreuses notions liées à la confidentialité des données recueillies, stockées dans des fichiers et destinées à être traitées par informatique. Elle précise les notions de fichier nominatif et de devoir d'information préalable des individus

CODE DE DEONTOLOGIE

Nous ne retiendrons que les grands principes :

- La volonté du malade doit toujours être respectée, dans toute la mesure du possible.
- Le secret couvre tout ce qui est venu à la connaissance du médecin dans l'exercice de sa profession, ce qu'on lui a confié mais aussi ce qu'il a vu, entendu ou compris.
- Le secret professionnel est institué dans l'intérêt du malade. Il le protège des indiscretions. Il ne peut pas lui être opposé (sauf art. 35). Il s'impose même entre médecins, sauf si plusieurs médecins concourent aux soins du même patient, c'est la notion de secret partagé. Dès lors, tous, y compris ceux qui assistent le médecin, se conforment à l'obligation du secret médical.
- Le médecin doit veiller à la protection contre toute indiscretion des documents concernant les malades. Lors d'une communication scientifique, il doit faire en sorte que l'identification du malade soit impossible.
- Le médecin doit juger de ce qu'il peut dire au patient. C'est l'article 35 : "Toutefois, dans l'intérêt du malade et pour des raisons légitimes que le praticien apprécie en conscience, un malade peut être tenu dans l'ignorance d'un diagnostic ou d'un pronostic graves, sauf dans les cas où l'affection dont il est atteint expose les tiers à un risque de contamination. Un pronostic fatal ne doit être révélé qu'avec circonspection, mais les proches doivent en être prévenus, sauf exception ou si le malade a préalablement interdit cette révélation ou désigné les tiers auxquels elle doit être faite."
- Il est des dérogations légales prévues à l'obligation du secret : déclarations administratives de naissance, décès, maladies infectieuses à déclaration obligatoire ou de maladies professionnelles. Ailleurs, tout renseignement concernant l'état du malade doit être ignoré de l'administration, en particulier dans le cadre du PMSI.

CODE PENAL

Il décrit les sanctions qui frappent le professionnel qui déroge aux règles déontologiques. L'article 378 isole le secret médical des autres secrets professionnels et stipule (Loi du 21 février 1944, validée par Ord. 28 juin 1945, réactualisée en 1971 et 1975) : "Les médecins, chirurgiens, et autres officiers de santé, ainsi que les pharmaciens, les sages-femmes et toutes autres personnes dépositaires, par état ou profession ou par fonctions

temporaires ou permanentes, des secrets qu'on leur confie, qui, hors le cas où la loi les oblige ou les autorise à se porter dénonciateurs, auront révélé ces secrets, seront punis d'un emprisonnement d'un mois à six mois et d'une amende de 500 à 8000 Frs".

ASPECTS PRATIQUES

A l'heure de l'informatique, dans la grande majorité des hôpitaux, la protection du secret ne fait pas l'objet d'une organisation systématique et standardisée destinée à l'assurer. Par exemple, l'accès aux dossiers papier est relativement aisé à tout bipède revêtu d'une blouse blanche. Cependant la faible maniabilité des documents ainsi que leur décentralisation, s'ils nuisent à leur disponibilité, concourent à leur protection. Il n'en est pas de même de données informatisées qui, si elles sont d'accès moins immédiat au tout-venant, permettent toutes les recherches et sélections dès lors que le système qui assure leur protection est franchi.

D'autre part, dans un hôpital, le responsable du dossier médical n'est pas le patient, ni le ou les médecins, mais l'hôpital représenté par son directeur. C'est donc lui qui est responsable des déclarations des fichiers nominatifs à la CNIL. Toutefois, bien que responsable, il n'a pas accès au contenu médical des dossiers.

INFORMATIQUE ET LIBERTES

La plupart des pays occidentaux ont défini un code d'usage de l'informatique dont l'objectif est de prévenir toute atteinte à la vie privée des personnes. En effet, l'informatique, par nature, fait peser un risque sur le secret de l'information : elle facilite l'organisation et l'extraction des données ; elle en assure la concentration ou la transmission ; elle permet les interconnexions de fichiers de finalité différentes, or par une série de tris et de recoupements sur diverses caractéristiques, il est possible d'identifier un individu ; enfin elle incite à une catégorisation des individus.

La loi française, dite "Informatique, fichiers et liberté" du 6 janvier 1978, crée un nouveau type de secret, le secret informatique, qui s'impose à tout citoyen, et par suite au médecin qui fait usage de fichiers nominatifs, pour qui il se superpose au secret médical. Elle porte essentiellement sur la protection des données nominatives, recueillies sur les individus, faisant l'objet de traitements automatisés, et s'oppose à toute utilisation à l'encontre de ces individus. Elle organise un droit d'accès, impose un devoir d'information préalable et soumet tout projet de traitement à l'avis d'une commission indépendante. Une directive européenne de 1995 reprend pour partie ces dispositions.

DEFINITIONS

- Notion de "nominatif":

Art.4: "Sont réputées nominatives au sens de la présente loi les informations qui permettent, sous quelque forme que ce soit, directement ou non, l'identification des personnes physiques auxquelles elles s'appliquent, que le traitement soit effectué par une personne physique ou une personne morale."

Il faut noter que cette notion d'information nominative est très large: "directement ou non". Il ne suffit pas de supprimer le nom si un numéro le remplace, d'autre part, les possibilités de recoupements multicritère peuvent lever l'anonymat. Il est interdit de mettre en mémoire, sauf autorisation explicite de l'intéressé, des informations en rapport avec l'appartenance politique, syndicale, religieuse ou ethnique.

• Notion de "Traitement automatisé":

Art.5: "Est dénommé traitement automatisé d'informations nominatives au sens de la présente loi tout ensemble d'opérations réalisées par des moyens automatiques, relatif à la collecte, l'enregistrement, l'élaboration, la modification, la conservation et la destruction d'informations nominatives ainsi que tout ensemble d'opérations de même nature se rapportant à l'exploitation de fichiers ou bases de données et notamment les interconnexions ou rapprochements, consultations ou communications d'informations nominatives ."

En fait, les dispositions concernant la collecte, l'enregistrement et la conservation des informations nominatives sont applicables aux fichiers non informatisés ou mécanographiques (Art. 45), soit aux dossiers médicaux traditionnels. Relève également de la loi toute procédure de transmission ou de téléconsultation, quel que soit le moyen de télécommunication (minitel, internet).

LE DROIT D'ACCES ET DE RECTIFICATION

C'est la possibilité pour un individu de connaître les informations que l'on possède à son sujet.

a) L'accès aux informations.

Le principe de base, "avoir accès à ce que l'on sait sur vous", est explicité dans l'article 3 :

Art.3: "Toute personne a le droit de connaître et de contester les informations et les raisonnements utilisés dans les traitements automatisés dont les résultats lui sont opposés."

La mise en oeuvre pratique figure dans les articles 34 et 35, nous ne citons que le 34 pour les cas généraux :

Art.34: "Toute personne justifiant de son identité a le droit d'interroger les services ou organismes chargés de mettre en oeuvre les traitements automatisés dont la liste est accessible au public en application de l'article 22 ci-dessus en vue de savoir si ces traitements portent sur des informations nominatives la concernant et, le cas échéant, d'en obtenir communication."

Mais, pour les informations médicales, une clause particulière figure dans l'article 40. Cet article a été rédigé pour tenir compte de l'intérêt moral du patient tel qu'il est défini dans l'article 35 du code de déontologie:

Art.40: "Lorsque l'exercice du droit d'accès s'applique à des informations à caractère médical, celles-ci ne peuvent être communiquées à l'intéressé que par l'intermédiaire d'un médecin qu'il désigne à cet effet."

b) La rectification des informations.

Le droit de connaître (droit d'accès aux informations) n'est pas suffisant pour garantir les libertés individuelles. S'il y a déformation de la vérité, il est nécessaire d'en obtenir rectification. C'est l'article 36 qui régit cette possibilité :

Art.36: "Le titulaire du droit d'accès peut exiger que soient rectifiées, complétées, clarifiées, mises à jour ou effacées les informations le concernant qui sont inexactes, incomplètes, équivoques, périmées ou dont la collecte, ou l'utilisation, la communication ou la conservation est interdite.

Lorsque l'intéressé en fait la demande, le service ou organisme concerné doit délivrer sans frais copie de l'enregistrement modifié."

L'OBLIGATION D'INFORMATION PREALABLE

Elle est instituée par l'article 27.

Art.27: "Les personnes auprès desquelles sont recueillies des informations nominatives doivent être informées :

- du caractère obligatoire ou facultatif des réponses ;
- des conséquences à leur égard d'un défaut de réponse ;
- des personnes physiques ou morales destinataires des informations ;
- de l'existence d'un droit d'accès et de rectification.

Lorsque de telles informations sont recueillies par voie de questionnaires, ceux-ci doivent porter mention de ces prescriptions.

Ces dispositions ne s'appliquent pas à la collecte des informations nécessaires à la constatation des infractions."

Cette obligation d'information a pour conséquence le refus possible de l'individu que décrit l'article 26.

Art.26: "Toute personne physique a le droit de s'opposer, pour des raisons légitimes, à ce que des informations nominatives la concernant fassent l'objet d'un traitement."

Ce droit ne s'applique pas aux traitements, limitativement désignés dans l'acte réglementaire prévu à l'article 15."

LA C.N.I.L. : Commission Nationale Informatique et Liberté

Elle est créée par les articles 6 et suivants de la loi. Elle est indépendante, bien que rattachée au Ministère de la Justice.

Les fonctions sont les suivantes:

- Donner un avis (acceptation ou refus) sur toute constitution de fichiers nominatifs.
- Assurer la diffusion du contenu de la loi, c'est-à-dire informer les citoyens de leurs droits et devoirs.

- Vérifier la conformité des traitements des fichiers nominatifs, avec les déclarations préliminaires faites auprès de la C.N.I.L.
- Instruire toute affaire relative aux plaintes dont elle serait saisie.
- Emettre des recommandations.

Dans le domaine médical, un certain nombre d'avis et de textes sont importants:

- Avis autorisant la constitution des fichiers du "programme de médicalisation du système d'information" (PMSI) dans le cadre des hôpitaux.
- la délibération n° 85-07 relative à l'utilisation d'informations médicales nominatives pour la recherche médicale (du 19 février 1985) précise que les objectifs de la recherche doivent être déclarés à l'avance, que le recueil et le traitement doivent être sous la responsabilité des seuls médecins et que le matériel informatique utilisé doit être dédié à cette fonction.
- La loi du 1er juillet 1994 avec son décret d'application (n° 95-682 du 9 mai 1995) institue une demande d'avis préalable et supplémentaire auprès d'un comité consultatif spécialisé dans la recherche pour les registres et les banques de données, qui jusqu'alors se trouvaient dans une situation difficile car collectant des données médicales nominatives mais sans participer directement à l'acte de soins.

PROTECTION DES DONNEES MEDICALES INFORMATISEES

La protection des données médicales a une double dimension :

- intégrité et disponibilité d'une part ;
- protection du secret d'autre part.

RISQUES

En effet, les risques qui pèsent sur l'informatique sont de deux ordres : dégradation et accès illicite. Ils peuvent se combiner tant au niveau des causes que des conséquences et toucher aussi bien le matériel que les programmes ou les données :

- destruction accidentelle ou volontaire conduisant à une perte de données ou à une utilisation impossible ;
- manipulation erronée ou falsification conduisant à des données non valides ;
- erreurs de programmes ou pannes matérielles de conséquences identiques ;
- vol des supports d'information ou des machines conduisant à une utilisation délictueuse ;
- accès délictueux, par consultation illicite ou utilisation frauduleuse des moyens de communication, ayant les mêmes conséquences ;
- accès non-intentionnel induisant une utilisation non-intentionnelle.

MESURES PREVENTIVES

Le responsable de l'organisation informatique est tenu de mettre en œuvre les mesures de nature à assurer la protection des fichiers comme la sécurité des données.

Mesures matérielles de protection physique

- protection des salles informatiques contre l'intrusion, l'incendie et autres dégâts ;
- regroupement des matériels dans les salles protégées ;
- redondance des équipements sensibles ;
- sécurisation de l'alimentation électrique ;
- sauvegardes automatiques régulières des données et des programmes ;
- conservation des sauvegardes dans des locaux indépendants.

Mesures logicielles de protection

- identification et authentification des utilisateurs (mot de passe, carte individuelle) ;
- hiérarchisation des données, des utilisateurs et des actions autorisées ;
- cloisonnement des ressources informatiques ;
- recherche systématique et prévention des virus informatiques ;
- programmes de surveillance ;
- cryptographie.

Mesures organisationnelles

- désignation d'un responsable de la sécurité des données ;
- établissement de procédures de sécurité ;
- formation et sensibilisation des utilisateurs ;
- audits de contrôle ;
- séparation stricte des données médicales nominatives avec utilisation d'un matériel qui lui soit dévolu exclusivement.

CONCLUSION

Le respect des prescriptions de ces lois et règles de fonctionnement est primordial. Une dégradation des données lors d'une transmission télématique peut être à l'origine d'erreurs d'interprétation. Un défaut de sécurité peut faciliter l'usage illicite de données confidentielles. De nombreux groupes de pression ou organismes ont intérêt à violer le secret médical pour mieux "connaître" les individus à qui ils ont affaire. Si le secret médical est violé, les libertés individuelles n'existent plus.

D'autre part, un autre aspect de la responsabilité concerne l'utilisation de logiciels ou de bases de données dans le domaine de la santé. Ainsi, si un programme fournit une mauvaise information au médecin, la responsabilité de ce dernier est-elle engagée ou faut-il se tourner vers le constructeur de la machine ou le développeur du logiciel ?

OBJECTIFS

- Connaître les notions de secret médical et les limites posées par la loi.
- Connaître les notions de fichier nominatif et les limites posées par la loi.
- Citer les techniques permettant d'assurer l'intégrité des données.
- Citer les techniques permettant d'assurer la sécurité des données.

Serment d'HIPPOCRATE (400 ans avant J-C)

"Je jure par Apollon, médecin, Asklépios, Hygéia et Panakéia, prenant à témoin tous les dieux et toutes les déesses, d'accomplir, selon mon pouvoir et mon jugement, ce serment et cet engagement écrit.

Je jure de considérer à l'égal de mes parents celui qui m'aura enseigné l'art de la médecine; de partager avec lui ma subsistance et de pourvoir à ses besoins, s'il est dans la nécessité; de regarder ses fils comme des frères, et s'ils veulent étudier cet art, de le leur apprendre sans salaire ni contrat; de communiquer les préceptes généraux, les leçons orales et tout le reste de la doctrine à mes fils, à ceux de mon maître et aux disciples enrôlés et assermentés suivant la loi médicale, mais à aucun autre.

Je ferai servir le régime diététique à l'avantage des malades selon mon pouvoir et mon jugement; pour leur dommage et leur mal ... non. Je ne donnerai pas, quiconque m'en prierait, une drogue homicide ni ne prendrai l'initiative de pareille suggestion ; de même, je ne donnerai à aucune femme un pessaire abortif.

Par la chasteté et la sainteté, je sauvegarderai ma vie et ma profession. Je ne taillerai pas les calculeux, et je laisserai cette pratique à des professionnels. En quelque maison que je doive entrer, je m'y rendrai pour l'utilité des malades, évitant tout méfait volontaire et corrupteur et, très particulièrement, les entreprises lascives sur le corps de femmes ou d'hommes, qu'ils soient libres ou esclaves.

Les choses que, dans l'exercice ou même hors de l'exercice de mon art, je pourrai voir ou entendre sur l'existence des hommes et qui ne doivent pas être divulguées au dehors, je les tairai, estimant que ces choses-là ont droit au secret des Mystères.

Si j'accomplis, jusqu'au bout ce serment et lui fais honneur, qu'il me soit donné de jouir des fruits de la vie et de cet art, honoré à jamais parmi tous les hommes. Mais si je la viole et si je me parjure, qu'il m'arrive tout le contraire !"

Le Serment d'Hippocrate actuel :

"En présence des Maîtres de cette Faculté, de mes chers condisciplines, devant l'effigie d'Hippocrate,

Je promets et je jure, au nom de l'Etre Suprême, d'être fidèle aux lois de l'honneur et de la probité dans l'exercice de la Médecine.

Je donnerai mes soins gratuits à l'indigent et n'exigerai jamais un salaire au-dessus de mon travail, je ne participerai à aucun partage clandestin d'honoraires.

Admis dans l'intérieur des maisons, mes yeux ne verront pas ce qui s'y passe, ma langue taira les secrets qui me seront confiés, et mon état ne servira pas à corrompre les mœurs, ni à favoriser le crime.

Je ne permettrai pas que des considérations de religion, de nation, de race, ou de classe sociale viennent s'interposer entre mon devoir et mon patient.

Je garderai le respect absolu de la vie humaine, dès la conception.

Même sous la menace, je n'admettrai pas de faire usage de mes connaissances médicales contre les lois de l'humanité.

Respectueux et reconnaissant envers mes Maîtres, je rendrai à leurs enfants l'instruction que j'ai reçue de leur père.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses.

Que je sois couvert d'opprobre et méprisé de mes Confrères si j'y manque. "

BIBLIOGRAPHIE

INFORMATIQUE

- "Architecture et technologie des ordinateurs". P. Zanella, Y. Ligier. Dunod, 1993.
- "L'Informatique éclatée". J.-M. Desainquentin, B. Sauter. Masson, 1991.
- "Le micro ... comment ça marche ?". R. White. Dunod, 1993.
- "Histoire universelle des chiffres". G. Ifrah. Bouquins, 1993.
- "Algèbre de Boole". Que-sais-je. PUF 1246.
- "Logique sans peine". Lewis Carroll. Ed. Hermann.
- "Initiation à l'algorithmique et aux structures de données". J.Courtin, I.Kowalski. Dunod, 1987.
- "Bases de données relationnelles". G.Gardarin, P.Valduriez. Eyrolles, 1985.
- "Intelligence artificielle". A.Bonnet. InterEditions, 1984.
- "Gödel, Escher, Bach". D.Hofstadter. InterEditions, 1985.
- "Y'a un bug !". F. Cointe. Armand Colin, 1990.

INFORMATIQUE MEDICALE

- "Informatique médicale". F. Grémy. Médecine-Sciences Flammarion, 1987.
- "Informatique médicale". P. Degoulet, M. Fieschi. Abrégés Masson, 1994.
- "Traitement de l'information médicale - méthodes et applications hospitalières". P. Degoulet, M. Fieschi. Manuels informatiques Masson, 1991.
- "Le médecin et son ordinateur". M. Goldberg, Ed. Frison-Roche, 1994.
- "Initiation à l'informatique". G.Soula, M.Fieschi et M.Roux. MEDIFAC
- "Intelligence artificielle en médecine". M. Fieschi. Méthodes et programmes Masson, 1986.
- "Evaluation de la décision médicale". B. Grenier. Evaluation et statistique Masson, 1996.
- "Histoire illustrée de 20 ans d'informatique hospitalière". C. Gallet, J.-M. Turreilles. P.G Promotion, 1993.
- "L'information médicale, l'ordinateur et la loi". L. Dusserre, H. Ducrot, F-A. Allaert. Editions Médicales Internationales, 1996.
- "Handbook of Medical Informatics". JH van Bommel, MA Musen. Springer, 1997.
- "ABC of Medical Computing". British Medical Journal ; vols 310-311 : série de 20 articles du 24 juin 1995 au 11 novembre 1995.
- "Guide to the Internet". British Medical Journal ; vol 311 : série de 4 articles du 25

novembre 1995 au 16 décembre 1995.

- "Informatique médicale". Revue du Praticien ; vol 46 : numéro spécial du 1er février 1996

COMPTE RENDUS DE COLLOQUES

- "MEDINFO". Congrès de l'International Medical Informatics Association (IMIA), tous les 3 ans. Elsevier - North Holland.
- "Lecture Notes in Medical Informatics - MIE". Congrès annuel de l'European Federation for Medical Informatics (EFMI). Springer Verlag.
- "Informatique et Santé". Colloques annuels de l'AIM (France). Springer Verlag.
- "Symposium on Computer Application in Medical Care". Congrès annuel de l'American Medical Informatics Association. Mc-Graw Hill.

REVUES : INFORMATIQUE MEDICALE - SANTE PUBLIQUE

- Technique et Science Informatique (TSI), Paris : Dunod.
- Computer and Biomedical Research, San Diego : Academic Press.
- Computer Methods and Programs in Biomedicine, Amsterdam : Elsevier.
- Medical Informatics, London : Taylor & Francis.
- Methods of Information in Medicine, Stuttgart : Schattauer.
- Informatique et Santé, Lyon : P.G. Promotion.
- Revue d'Epidémiologie et de Santé Publique, Paris : Masson.
- Journal of the American Medical Informatics Association (JAMIA).

QUELQUES SITES WEB

- www.chu-rouen.fr : le CHU de Rouen recense la quasi-totalité des serveurs médicaux français et beaucoup d'étrangers
- www.med.univ-rennes1.fr : le site de la Faculté de Rennes avec les bases IconoCERF et ADM
- www.le-pmsi.fr : le site officiel du PMSI
- www.ordmed.org : le site du Conseil de l'Ordre des Médecins
- www.cnil.fr : le site de la Commission Nationale Informatique et Libertés
- www.mieur.nl/mihandbook : le site compagnon du livre "Handbook of Medical Informatics"
- www.ncbi.nlm.nih.gov/PubMed/medline.html : le site permettant d'interroger Medline gratuitement, avec le pointeur vers la NLM et les projets UMLS et Visible Human

