



MÉTHODES DE PROGRAMMATION ET ALGORITHMIQUE

- INTRODUCTION A LA PROGRAMMATION
 1. Algorithmique et langages. J. Biondi et G. Clavel.
 2. Structures de données. G. Clavel et J. Biondi.
 3. Exercices corrigés. G. Clavel et F.B. Jorgensen.
- SCHEMAS ALGORITHMIQUES FONDAMENTAUX. P.C. Scholl et J.P. Peyrin.
- MATÉRIEL ET LOGICIEL. L. Nolin
 1. Cours de licence avec exercices.
 2. Cours de maîtrise avec exercices.
- INTRODUCTION A LA PROGRAMMATION SYSTÉMATIQUE. N. Wirth.
- ALGORITHMIQUE ET REPRÉSENTATION DES DONNÉES
 1. Files, automates d'états finis. M. Lucas, J.P. Peyrin et P.C. Scholl.
 2. Évaluations, arbres, graphes, analyse de texte. M. Lucas.
 3. Récursivité et arbres. P.C. Scholl.
- ALGORITHMIQUE Conception et analyse. G. Brassard et P. Bratley.
- GÉNIE LOGICIEL. M. Thorin.
- MAINTENANCE DU LOGICIEL. R.L. Glass et R.A. Noiseux.
- LA TRANSPORTABILITÉ DU LOGICIEL. O. Lecarme et M. Pellissier.
- ANALYSE FORMELLE D'ALGORITHMES. Raisonnements et erreurs dans des algorithmes. R. Lesuisse.
- PROCESSUS CONCURRENTS. Introduction à la programmation parallèle. M. Ben-Ari.
- PROCESSUS SÉQUENTIELS COMMUNICANTS. C.A.R. Hoare.
- LA COMPRESSION DES DONNÉES. Méthodes et applications. G. Held.
- PRINCIPES DE PROGRAMMATION FONCTIONNELLE. H. Glaser, C. Hankin et D. Till.
- PROGRAMMATION EN LOGIQUE. C.J. Hogger.
- MISE EN ŒUVRE DES LANGAGES FONCTIONNELS DE PROGRAMMATION. S.L. Peyton Jones.
- CONSTRUCTION ET VÉRIFICATION DE PROGRAMMES. R. Backhouse.
- OUTILS LOGICIELS POUR LA PROGRAMMATION SYSTÈME. T.J. Biggerstaff.

LES LANGAGES ET LEUR TRAITEMENT

- LES LANGAGES DE PROGRAMMATION. Concepts essentiels, évolution et classification. J. Lonchamp.
- INTRODUCTION AU LANGAGE ADA. D. Price.
- MANUEL ADA. Langage normalisé complet. M. Thorin.
- ADA. Une introduction. H. Ledgard.
- APPRENDRE ET APPLIQUER LE LANGAGE APL. B. Legrand.
- APL et GDDM. Travail en plein écran. B. Legrand.
- PROGRAMMATION EN ASSEMBLEUR. Initiation à partir du Fortran. J.F. Phelizon.
- LE LANGAGE C. B.W. Kernighan et D.M. Ritchie.
- LE LANGAGE C. Solutions. C.L. Tondo et S.E. Gimpel.
- LANGAGE C norme ANSI. Vers une approche orientée objet. Ph. Drix.
- LANGAGE C. Problèmes et exercices. A.R. Feuer.
- PROGRAMMER EN C++ S. C. Dewhurst et K.T. Stark.
- CONSTRUCTION LOGIQUE DE PROGRAMMES COBOL. Mise à jour COBOL 85. M. Koutchouk.
- COBOL. Perfectionnement et pratique. M. Koutchouk.
- (COMMON) LISP. Une introduction à la programmation. H. Wertz.
- OCCAM 2. Manuel de référence. INMOS.
- MODULA-2. A.B. Fontaine.
- LES LANGAGES DE PROGRAMMATION. Pascal, Modula, Chili, Ada. Ch. Smedema, P. Medema, M. Boasson.
- LE LANGAGE PASCAL. J.M. Crozet et D. Serain.
- LANGAGE PL/1. Initiation. Perfectionnement. R.P. Balme.
- POP 11. Un langage adapté à l'intelligence artificielle. R. Barrett, A. Ramsay et A. Sloman.
- TRAITEMENT DES LANGAGES ÉVOLUÉS. Compilation, Interprétation, Support d'exécution. Y. Noyelle.
- APPRENDRE PASCAL ET LA RÉCURSIVITÉ. Avec exemples en TURBO PASCAL. R. Romanetti.

-
- Cours rédigé et enseigné par un professeur francophone

Suite page 3 de couverture

1327659

67

Traitement de l'information médicale

30V

103067

CHEZ LE MÊME ÉDITEUR

Des mêmes auteurs :

INTELLIGENCE ARTIFICIELLE EN MÉDECINE, des systèmes experts, par M. FIESCHI, préface de M. ROUX. *Collection Méthode + Programmes*, 1986, 2^e édition, 240 pages.

Autres ouvrages :

LES SCHÉMAS DIRECTEURS : DÉMARCHE PRATIQUE, par G. BALANTZIAN. *Collection Méthodes Informatiques et Pratique des Systèmes*, 1990, 3^e édition, 272 pages.

L'ÉVALUATION DES SYSTÈMES D'INFORMATION ET DE COMMUNICATION, par G. BALANTZIAN. *Collection Méthodes Informatiques et Pratique des Systèmes*, 1990, 272 pages.

CONCEPTION ASSISTÉE DES SYSTÈMES D'INFORMATION. Méthode, modèles, outils, par F. BODART et Y. PIGNEUR. *Collection Méthodes Informatiques et Pratique des Systèmes*, 1989, 2^e édition, 336 pages.

MÉTHODE GÉNÉRALE D'ANALYSE DES APPLICATIONS INFORMATIQUES, par X. CASTELLANI :

- Tome 0 — Macro-analyse, étude préalable et analyse conceptuelle des systèmes d'information, 1987, 696 pages.
- Tome 1 — Étapes et points fondamentaux de l'analyse fonctionnelle, 1986, 7^e édition, 328 pages.
- Tome 2 — Étapes et points fondamentaux de l'analyse organique et de la programmation, 1986, 6^e édition, 268 pages.

INGÉNIERIE DES DONNÉES. Bases de données, systèmes d'informations, modèles et langages, par E. PICHAT et R. BODIN. *Collection Manuels Informatiques Masson*, 1990, 432 pages.

MAÎTRISER LA MODÉLISATION CONCEPTUELLE, par R. PLANCHE, préface de J.L. PEAUCELLE, *Collection Méthodes Informatiques et Pratique des Systèmes*, 1988, 256 pages.

BIostatistique EN MÉDECINE CLINIQUE, par J.A. INGELFINGER, F. MOSTELLER, L.A. THIBODEAU et J. H. WARE, traduit de l'anglais par C. de CHEVEIGNÉ, 1988, 360 pages.

INTELLIGENCE ARTIFICIELLE ET BON SENS, par la FONDATION F.R. BULL, préface de L. LEPRINCE-RINGUET, 1991, 336 pages.

INFORMATIQUE DOCUMENTAIRE, par A. DEWÈZE, préface de J. CORDONNIER et R. BOUCHÉ. *Collection Méthode + Programmes*, 1989, 3^e édition, 296 pages.

MANUELS INFORMATIQUES MASSON

NC

Traitement de l'information médicale méthodes et applications hospitalières

Patrice DEGOULET

Professeur à la faculté de médecine
Broussais Hôtel-Dieu
de l'université de Paris-VI

Marius FIESCHI

Professeur à la faculté de médecine
de l'université d'Aix-Marseille II

préface de
François GRÉMY

MASSON Paris Milan Barcelone Bonn 1991

Tous droits de traduction, d'adaptation et de reproduction par tous procédés, réservés pour tous pays.

Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des pages publiées dans le présent ouvrage, faite sans l'autorisation de l'éditeur, est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part, les reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, et d'autre part, les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (loi du 11 mars 1957 art. 40 et 41 et Code pénal art. 425).

Des photocopies payantes peuvent être réalisées avec l'accord de l'éditeur. S'adresser au : Centre français du copyright, 6 bis, rue Gabriel-Laumain, 75010 Paris, tél. : 48.24.98.30.

© Masson, Paris, 1991

ISBN : 2-225-82514-9

ISSN : 0249-6992

MASSON
MASSON S.p.A.
MASSON S.A.
DÜRR und KESSLER



120, bd Saint-Germain, 75280 Paris Cedex 06
Via Statuto 2/4, 20121 Milano
Balmes 151, 08008 Barcelona
Maarweg, 30, 5342, Rheinbreitbach b. Bonn

Préface

Ce "Traitement de l'information médicale", avec son sous-titre "Méthodes et applications hospitalières", est une nouvelle étape d'une longue marche commencée à la fin des années 1960. Il n'est pas si loin, le temps où ceux qui plaidaient pour l'émergence et une certaine autonomie de ce qu'on appelait alors *Informatique Médicale*, étaient considérés comme des imposteurs, des gêneurs ou de dangereux utopistes. Les arguments qu'on invoquait contre eux étaient variés, convergents, mais parfois contradictoires :

- Qu'avons-nous besoin de nous occuper d'information médicale pour faire fonctionner nos hôpitaux ? Réservez nos efforts à l'informatisation des activités administratives et comptables : les stocks, la facturation, le personnel. N'est-ce pas suffisant pour assurer la bonne et honnête gestion qu'on attend de nous ? (discours de l'administration hospitalière).
- La médecine est beaucoup trop subtile, trop "incommunicable", trop "informalisable", pour être coincée dans des procédures à base de tout ou rien (discours des médecins).

Et, pour ceux qui admettaient que l'informatisation des données médicales est faisable et utile :

- Pourquoi vouloir ériger l'informatique médicale en une discipline nouvelle, alors qu'il suffit de mettre en présence des demandeurs (éventuels) : les médecins, et des techniciens : les informaticiens ?

C'est dire qu'il a fallu faire preuve d'une certaine combativité pour obtenir la création d'unités de recherche à la Salpêtrière, puis à Nancy, la mise en place de quelques Services d'Informatique Médicale dans quelques hôpitaux, et l'individualisation d'une discipline universitaire : *Biostatistique et Informatique Médicale*. Depuis les débuts héroïques et confus, quelques idées-force se sont imposées, qui ont volatilisés les objections des premiers jours :

- D'abord, une rectification de la part des promoteurs de l'informatique médicale. D'une logique instrumentale (que faire de cet instrument dans le domaine de la santé ?), ils sont passés à une logique de finalité : quels sont les problèmes d'information qui se posent en Santé et en Médecine ? Quels sont les méthodes et outils qui permettent d'atteindre les objectifs ainsi définis ? Et, parmi ces outils, quelle place donner à l'informatique dans ses développements matériels, logiciels et méthodologiques ?
- Ensuite, le fait qu'un certain nombre de cliniciens et de médecins de laboratoire, plus lucides que d'autres, se sont engagés dans une recherche sur les apports possibles de l'informatique à leur pratique. Sans vouloir être exhaustif, les noms de

Henry DUCROT, Marcel LEGRAIN, Joël MÉNARD, Pierre DUSSEY, Etienne MARTIN, sont parmi ceux qui nous viennent les premiers à l'esprit pour illustrer cette évolution...

- La réticence des hôpitaux "atténuée, mais toujours actuelle" à s'engager vers l'informatisation des unités de soins, et de l'action médicale, en restreignant les réalisations concrètes, a amené à approfondir la réflexion méthodologique et épistémologique : jusqu'où est-il possible de formaliser la connaissance et l'action médicales ? Quels sont l'intérêt et les limites de ces tentatives ? C'était poser le problème de l'aide à la décision et du langage médical... C'est cette réflexion qui est à l'origine du manuel "Informatique Médicale", que l'auteur de ces lignes a commis dans le début des années 1980.
- A la même période, le Directeur des Hôpitaux de l'époque, Jean de KERVAIS-DOUÉ, a affirmé qu'on avait fait fausse route jusque là, en négligeant l'apport d'un bon usage de l'information médicale à la gestion de l'hôpital. Il a lancé le Projet puis le Plan de Médicalisation des Systèmes d'Information. L'administration, grâce à lui, reconnaissait qu'il était difficile de gérer une entreprise et une institution, quelle qu'elle soit, en faisant abstraction de ses finalités et de sa production. De là, l'ouverture des Départements de l'Information Médicale dans plusieurs centaines d'hôpitaux. De là, aussi, l'impulsion donnée aux méthodes d'évaluation de l'hôpital dans son fonctionnement quotidien. Ici encore, ce sont les premiers pas d'une longue marche, où l'Informatique Médicale hospitalière va jouer un rôle décisif.
- Enfin, la spécialisation d'Informatique Médicale est apparue et n'est plus discutée. La maîtrise simultanée des concepts et du langage de la médecine, des outils de l'informatique et de la statistique, donne à ceux qui possèdent cette maîtrise une silhouette spécifique et irremplaçable.

Les auteurs de cet ouvrage sont parmi les meilleurs représentants de cette nouvelle génération de jeunes maîtres. En Patrice DEGOULET, je suis heureux de reconnaître l'un des plus doués, des plus créatifs, et des plus brillants de ceux que je m'honore d'appeler mes élèves... Quant à Marius FIESCHI, il est le plus brillant représentant de l'Ecole de Marseille : il s'est imposé sur le plan international comme l'un des meilleurs spécialistes de l'intelligence artificielle appliquée à la médecine. Les lecteurs apprécieront de cet ouvrage la clarté, mais aussi le caractère technique, ainsi que la profonde réflexion sur la pensée et les concepts de la Médecine qui sous-tend toutes ces pages. Ce livre marque donc bien, à n'en pas douter, une nouvelle étape dans cette évolution prometteuse et irrévocable, vieille aujourd'hui de quelque vingt-cinq ans. Le temps des prophètes et des philosophes s'estompe. Arrivent les spécialistes pointus. Espérons qu'absorbés par l'objet technique, ils continueront de se rappeler la signification et la finalité de celui-ci : servir les hommes et leur santé ; mieux comprendre comment fonctionne, et ne fonctionne pas, l'intelligence de l'Homme.

François GRÉMY,

Professeur à la faculté de médecine de Montpellier

Table des matières

Préface	v
Avant-propos	xv
Introduction	1
I Méthodologie générale	5
1 Modèles d'information	7
1.1 Données, informations et connaissances	7
1.2 Modèles	8
1.2.1 De la représentation à la sémantique des informations	8
1.2.2 Modèles généraux et modèles spécifiques	9
1.2.3 Le système d'information	9
1.3 Concepts de modélisation	10
1.3.1 Objets et données élémentaires	10
1.3.2 Liaisons et catégories	10
1.3.3 Ensembles, domaines et attributs	11
1.3.4 Notion d'abstraction et héritage de propriétés	12
1.3.5 Primitives sémantiques	14
1.3.6 Syntaxe et sémantique d'une représentation	16
1.3.7 Standardisation et structuration	17
1.4 Structures	17
1.4.1 Structures de relations et d'enregistrements	17
1.4.2 Trames et scripts	20
1.4.3 Objets structurés	21
1.5 Opérations	22
1.5.1 Sélection d'un sous-ensemble ou d'un élément d'un modèle	22
1.5.2 Les langages navigationnels	23
1.5.3 Spécification d'opérations et langages assertionnels	24
1.6 Contraintes	27
1.6.1 Définition	27
1.6.2 Contraintes inhérentes, explicites et implicites	27
1.6.3 Contraintes sur les attributs d'un objet	28
1.6.4 Contraintes sur les liaisons entre 2 ensembles	29
1.6.5 Contraintes de dépendance	31
1.6.6 Contraintes de classes	33
1.7 Résumé et conclusions	33

2	Conduite de projets informatiques	39
2.1	Introduction	39
2.2	Les étapes de la conduite de projets	40
2.2.1	Les méthodes en cascade	40
2.2.2	Les méthodes de prototypage rapide	43
2.2.3	Les systèmes à base de connaissances	43
2.3	Le schéma directeur et la méthode RACINES	43
2.3.1	Introduction	43
2.3.2	Les acteurs du schéma directeur et leur rôle	44
2.3.3	La mise en place du schéma directeur	45
2.3.4	Les étapes du schéma directeur	46
3	Analyse et conception des systèmes d'information	51
3.1	Introduction	51
3.2	La méthode d'analyse structurée SADT	52
3.2.1	Introduction	52
3.2.2	Actigrammes et datagrammes	53
3.2.3	Liaisons entre les diagrammes	53
3.2.4	Organisation hiérarchique des diagrammes	53
3.2.5	Textes explicatifs et glossaire	54
3.2.6	Décomposition des diagrammes	54
3.2.7	Formulaires SADT	56
3.2.8	Modélisation conceptuelle	56
3.3	La méthode d'analyse structurée SASS	56
3.3.1	Introduction	56
3.3.2	Les diagrammes de flux de données	57
3.3.3	Les étapes de la méthode SASS	57
3.3.4	Discussion	58
3.4	Le modèle entité-association	59
3.4.1	Introduction	59
3.4.2	Structure	59
3.4.3	Contraintes	61
3.4.4	Opérations	62
3.4.5	Discussion	63
3.5	Le modèle relationnel binaire et la méthode NIAM	64
3.5.1	Introduction	64
3.5.2	Structure et symbolisme	64
3.5.3	Contraintes	66
3.5.4	Discussion	67
3.6	Les approches orientées objets	68
3.6.1	Introduction	68
3.6.2	Concepts de bases	68
3.6.3	Les étapes de l'approche objet	69
3.6.4	Discussion	71
3.7	Notes de lecture	72

4	Systèmes de gestion de bases de données	75
4.1	Introduction	75
4.2	Objectifs d'une base de données	77
4.2.1	Aide à la prise de décisions	77
4.2.2	Partage des données	77
4.2.3	Diminution des redondances	77
4.2.4	Diminution des incohérences	77
4.2.5	Standardisation	78
4.2.6	Structuration	78
4.2.7	Indépendance logique	78
4.2.8	Indépendance physique	79
4.3	Architecture générale d'une base de données	79
4.3.1	Les trois niveaux de description des données	79
4.3.2	Correspondances entre les niveaux de description des données	81
4.4	Objets manipulés	81
4.4.1	Les objets	81
4.4.2	Les items ou atomes d'information	82
4.4.3	Clés d'accès	82
4.4.4	Valeurs nulles	83
4.4.5	Agrégats d'items	83
4.5	Structures logiques	83
4.5.1	Structure d'arbre	83
4.5.2	Structure en réseau ("network" ou "plex structure")	83
4.5.3	Le modèle relationnel	84
4.5.4	Les modèles sémantiques	85
4.5.5	Les modèles orientés objet	85
4.6	Les langages de définition des données	86
4.6.1	Caractéristiques générales	86
4.6.2	Le langage de description du schéma conceptuel	86
4.6.3	Le langage de description des schémas externes	87
4.6.4	Le langage de description du modèle interne	87
4.7	La description des contraintes	87
4.8	Les langages de manipulation des données	88
4.8.1	Caractéristiques générales	88
4.8.2	Primitives de manipulation des données	89
4.8.3	Procédures et transactions	89
4.9	Le système de gestion de bases de données	89
4.9.1	Définition	89
4.9.2	Transferts de données	90
4.9.3	Protection des données	90
4.9.4	Monitoring de la base de données	90
4.10	Dictionnaire des données et métabase	91
4.11	Les utilisateurs d'une base de données	91
4.12	Discussion et conclusion	92

5 Réseaux et communications	97
5.1 Introduction	97
5.2 Typologie des réseaux	97
5.2.1 Les supports physiques de communication	97
5.2.2 Réseaux locaux et réseaux longue distance	98
5.2.3 Topologie des réseaux	98
5.3 Les standards de communication	99
5.4 Le modèle ISO/OSI	101
5.4.1 Les architectures de réseaux	101
5.4.2 Les sept couches du modèle ISO/OSI	101
5.4.3 La couche physique (couche 1)	102
5.4.4 La couche liaison de données (couche 2)	102
5.4.5 La couche réseau (couche 3)	104
5.4.6 La couche transport (couche 4)	104
5.4.7 La couche session (couche 5)	104
5.4.8 La couche présentation (couche 6)	105
5.4.9 La couche application (couche 7)	105
5.5 La transmission des données médicales	105
6 Interfaces homme-machine	109
6.1 Introduction	109
6.2 Le matériel	109
6.2.1 L'interaction homme-machine : entrée des données	109
6.2.2 L'interaction machine-homme : sortie des résultats	111
6.3 Principe de fonctionnement des interfaces	111
6.3.1 Les composants fonctionnels d'une interface	111
6.3.2 Les modèles architecturaux	112
6.3.3 Les styles d'interaction	113
6.4 Les logiciels de réalisation d'interfaces	113
6.4.1 Les systèmes de fenêtrage	114
6.4.2 Les boîtes à outils	115
6.4.3 Les systèmes de développement d'interfaces	116
II Systèmes d'information médicaux	119
7 Données et décisions médicales	121
7.1 Introduction	121
7.2 Les probabilités et leur estimation	122
7.2.1 Probabilités conditionnelles	122
7.2.2 Valeur prédictive d'un test	123
7.3 Valeur diagnostique d'un signe ou d'un test	124
7.4 Structuration des décisions cliniques sous incertitude : les arbres de décision	128
7.5 Représentation de concepts imprécis : les ensembles flous	128

8 Langage médical et systèmes de classification	131
8.1 Introduction	131
8.2 Les langages classificatoires	131
8.3 Exemples de classifications	134
8.3.1 La Classification internationale des Maladies et ses extensions	135
8.3.2 Le CANDO médical et pharmaceutique	137
8.3.3 La nomenclature SNOMED	138
8.3.4 Le thésaurus MeSH	140
8.4 Vers un langage médical unifié, le projet UMLS	141
8.5 Discussion	142
9 Systèmes d'information hospitaliers	147
9.1 Introduction	147
9.2 Le système d'information de l'hôpital	147
9.2.1 Les différents niveaux du système d'information	147
9.2.2 Objectifs, structures et fonctions	148
9.3 Stratégies et solutions techniques	150
9.3.1 L'approche verticale : approche SIH traditionnelle	150
9.3.2 L'approche horizontale : approche produits	153
9.3.3 L'approche distribuée : verticale et horizontale	153
9.4 Les ressources nécessaires	155
9.4.1 L'allocation des ressources et la prévision des coûts	155
9.4.2 Les ressources humaines	156
9.5 Discussion et conclusion	157
10 Informatisation des dossiers médicaux	161
10.1 Introduction	161
10.2 Les fonctions du dossier médical et les bénéfices attendus de l'informatisation	162
10.3 La modélisation des informations médicales	164
10.3.1 Comment standardiser la terminologie médicale ?	164
10.3.2 Comment structurer le dossier médical ?	164
10.4 Mise en œuvre des dossiers médicaux informatisés	168
10.4.1 Les logiciels généraux de gestion de bases de données	168
10.4.2 Les logiciels spécialisés de gestion de données médicales	170
10.5 Perspectives et limites des dossiers informatisés	171
10.5.1 La connexion de systèmes experts et de bases de connaissances à un système de gestion de dossiers	171
10.5.2 Génération d'informations structurées et standardisées à partir de textes libres	171
10.5.3 Génération automatique de nouvelles connaissances à partir d'un dossier structuré et standardisé	172
10.6 Indications des dossiers informatisés	172
10.6.1 Le dossier minimum et les résumés de dossier	172
10.6.2 Les dossiers médicaux spécialisés	173
10.6.3 Les registres multicentriques de maladies	174
10.7 Résumé et conclusions	174

11 Informatisation des unités de soins	179
11.1 Introduction	179
11.2 Le système d'information de l'unité de soins	179
11.2.1 Les acteurs de l'unité de soins	179
11.2.2 La démarche de soins	180
11.2.3 Les événements	180
11.3 Les fonctions d'un système de gestion de l'unité de soins	181
11.3.1 Les fonctions de prise en charge et de suivi des patients . . .	181
11.3.2 Les fonctions de gestion de l'unité de soins	183
11.3.3 Enseignement et recherche	184
11.4 Intégration dans le système d'information hospitalier	184
11.4.1 L'estimation des besoins des utilisateurs	185
11.4.2 Stratégies de mise en œuvre et solutions techniques	185
12 Informatisation des laboratoires	191
12.1 Introduction	191
12.2 Le système d'information biologique	191
12.3 Les fonctions d'un système de gestion de laboratoires	192
12.3.1 La demande d'examens	192
12.3.2 Le prélèvement	193
12.3.3 La réception des échantillons	193
12.3.4 Analyse et acquisition des résultats	193
12.3.5 Validation	193
12.3.6 Edition des résultats, aide à la décision	194
12.3.7 Transmission et archivage des résultats	194
12.3.8 Comptabilité des actes et facturation	194
12.3.9 Statistiques, contrôle de qualité	194
12.4 Les différents systèmes de laboratoire	195
12.4.1 Biochimie et hématologie	195
12.4.2 Banque du sang	195
12.4.3 Microbiologie et immunologie	195
12.4.4 Histologie et anatomo-pathologie	195
12.5 Intégration dans le système d'information	195
13 L'imagerie médicale	199
13.1 Introduction	199
13.2 L'imagerie médicale	199
13.2.1 Concepts fondamentaux de l'imagerie digitalisée	199
13.2.2 Quelques techniques d'imagerie	200
13.3 Le système d'information d'imagerie	202
13.4 Applications de l'informatique en imagerie	203
13.5 PACS et systèmes d'information hospitaliers	204
13.5.1 Définition et objectifs d'un PACS	204
13.5.2 Liens d'un PACS avec le système d'information hospitalier .	205

14 Aide à la décision médicale	211
14.1 Introduction	211
14.2 L'aide à la décision médicale	212
14.2.1 Les objectifs et les questions posées	212
14.2.2 Typologie des systèmes d'aide à la décision	213
14.3 Méthodologie de l'aide à la décision	215
14.3.1 L'approche purement numérique de l'aide à la décision	215
14.3.2 L'approche intelligence artificielle	215
14.3.3 Les différents modèles de connaissances	216
14.3.4 Les méthodes de raisonnement	218
14.3.5 Les formalismes de représentation des connaissances	219
14.3.6 Logique et théorie de la décision pour la représentation de l'incertain : le raisonnement sous incertitude	222
14.4 Exemples de systèmes d'aide à la décision	224
14.4.1 INTERNIST, CADUCEUS et QMR	225
14.4.2 Le système MYCIN	225
14.4.3 Le système expert SPHINX	227
14.5 Intégration dans les systèmes d'information hospitaliers	230
14.5.1 Intérêt et problèmes posés	230
14.5.2 L'exemple du système HELP	230
14.5.3 Le système CARE	234
14.6 Discussion et conclusion	235
15 Banques de données et d'informations médicales	241
15.1 Introduction	241
15.2 Aspects méthodologiques	241
15.2.1 Recherche de données et recherche d'information	241
15.2.2 Principaux modèles de recherche d'information	242
15.2.3 Performances et efficacité des systèmes documentaires	243
15.3 Organisation et mise en œuvre	244
15.3.1 Banques de données bibliographiques	244
15.3.2 Banques d'information	245
15.4 Discussion et conclusion	247
III Evaluation	249
16 Evaluation des systèmes d'information	251
16.1 Introduction	251
16.2 Présentation du problème	252
16.2.1 Les dimensions de la validation des systèmes d'information	252
16.2.2 Classification des critères d'évaluation	252
16.3 Méthodologie d'évaluation	254
16.4 Evaluation d'un environnement logiciel	254
16.5 Validation des systèmes experts	256
16.5.1 Les problèmes liés à la validation	256
16.5.2 Méthodologie d'évaluation	258
16.5.3 Un exemple : l'évaluation du système SPHINX	260

16.5.4	Les méthodes de validation automatique	263
16.5.5	L'intégration des systèmes experts	264
17	Mesure et analyse de l'activité médicale	267
17.1	Introduction	267
17.2	Méthodes et principes de l'évaluation	267
17.3	Mesure et analyse de l'activité médicale à l'hôpital	268
17.3.1	La cotation des actes	268
17.3.2	Les groupes homogènes de malades	269
17.3.3	L'utilisation des résumés de sortie standardisés	271
17.4	Discussion et conclusion	274
18	Protection des données, aspects éthiques et juridiques	277
18.1	Introduction	277
18.2	La protection des données nominatives	277
18.2.1	Les bases légales et déontologiques	277
18.2.2	Mesures matérielles, logicielles et organisationnelles	281
18.3	Exemples	282
18.3.1	Les études multi-centriques	282
18.3.2	La responsabilité médicale et l'utilisation de systèmes experts	284
18.4	Résumé et conclusions	284
18.5	Notes de lecture	285
	Annexe	287
	Index	289

Avant-propos

Le concept de système d'information s'est particulièrement développé et affiné dans le domaine médical depuis une vingtaine d'années. Plus récemment, il s'est retrouvé au centre de la démarche entreprise par de nombreux centres hospitaliers et l'on peut s'attendre à ce qu'il permette une meilleure approche des processus de soins.

Ce livre s'adresse à tous ceux, professionnels ou étudiants qui sont confrontés ou intéressés par des problèmes de conception ou de réalisation de systèmes d'information médicaux.

Notre principal but, en écrivant ce texte, est de présenter et d'illustrer les principaux concepts de l'informatique médicale en montrant que cette discipline ne se limite pas à la mise en œuvre d'ordinateurs dans le domaine médical.

Les étudiants de maîtrise et de troisième cycle trouveront dans ce livre les éléments indispensables pour la compréhension des problèmes du traitement de l'information médicale. Les professionnels y trouveront, par le texte présenté et les références bibliographiques nombreuses qui lui sont annexées, un moyen d'actualiser leurs connaissances.

Nous remercions chaleureusement tous ceux qui nous ont apporté leurs critiques et suggestions : Françoise Aimé, Geneviève Botti, Dominique Fieschi, Isabelle Fofol, François Christophe Jean, Marie-Christine Jaulent, Michel Joubert, Marion Lavril, Dominique Sauquet et Gérard Soula. Sylvie Bonay et Muriel Yvars ont participé à frappe du manuscrit. Les icones de la couverture ont été réalisés par Mohamed Ben Said

Introduction

Plusieurs évolutions majeures ont marqué la pratique médicale des vingt dernières années. L'*accroissement des connaissances médicales* est explosif et l'augmentation du nombre de paramètres nécessaires à la prise en charge des patients se fait en parallèle. Le développement ininterrompu de nouvelles méthodes d'investigations comme les techniques d'imagerie, les explorations fonctionnelles ou les techniques dérivées du génie biologique contribue à cette augmentation des connaissances. Il devient de plus en plus difficile, pour un médecin, de maîtriser les connaissances nécessaires à la pratique des soins, voire simplement d'y accéder en temps utile.

Le dossier médical ne se résume plus aux notes cliniques du médecin mais devient la compilation d'analyses de plus en plus sophistiquées. Ainsi par exemple dans le domaine de l'imagerie, alors qu'en 1970 le médecin ne disposait pratiquement que des appareils de radiologie classique, plusieurs dizaines de techniques faisant largement appel aux moyens informatiques sont maintenant disponibles (échographie, scanner, angiographies digitales, imagerie par résonance magnétique, etc.). D'une façon plus générale, la gestion d'objets complexes est devenue nécessaire à la prise de décisions. Ces objets recouvrent, en plus des images, des textes en langage clair, des signaux (électriques, acoustiques, électro-physiologiques, etc.) mais aussi des ensembles de concepts reliés entre eux (connaissances physiopathologiques) ou des procédures devant être activées sous certaines conditions (protocoles de calcul d'une dose optimale d'un médicament, protocoles de soins infirmiers, etc.).

Les *maladies chroniques* occupent une place croissante dans l'activité médicale et la durée moyenne de surveillance des patients augmente régulièrement. La probabilité pour un patient donné d'être suivi par un seul médecin est quasi nulle. La prise en compte de l'évolution d'un malade demande un effort de synthèse toujours plus important.

La prise en charge des patients n'est plus le fait d'un médecin isolé ou d'un nombre réduit de médecins mais d'une *équipe de soins* partageant des ressources et des compétences complémentaires. Le dossier du patient doit être immédiatement accessible et communicable aux différents membres de l'équipe de soins.

Les différents partenaires du système de santé prennent progressivement conscience des *contraintes économiques et éthiques* qui viennent peser sur le développement des techniques de pointe, en même temps que des risques engendrés par l'apparition de thérapeutiques plus efficaces mais aussi plus dangereuses. Le nombre, la richesse et la complexité des techniques peuvent avoir des conséquences négatives sur le développement de la médecine et de sa pratique et amener des coûts de plus en plus difficiles à supporter pour la société.

La recherche en informatique médicale vise, dans ce cadre, à développer des méthodes et des outils pour faciliter le traitement rationnel, rapide et fiable des informations médicales, c'est-à-dire l'amélioration directe ou indirecte de la qualité

des soins. Mémorisation, communication, aide à la décision et évaluation représentent les mots clés de cette recherche.

La *mémorisation* concerne tout particulièrement les données des patients, stockées dans les dossiers médicaux et les connaissances médicales de nature pratique comme les protocoles de prise en charge de problèmes médicaux individualisés, ou plus générales comme des connaissances anatomiques, physiologiques ou pharmacologiques. Elle concerne également la connaissance de l'environnement de travail comme les ressources humaines ou matérielles disponibles à un moment donné ou le cadre juridique de l'activité de soin. L'accessibilité, la fiabilité et la pertinence des données concernant un malade ou les connaissances nécessaires à la prise de décisions conditionnent la qualité des soins. D'essentiellement diagnostique et pronostique, l'exercice médical devient également stratégique : choix du meilleur site de traitement, choix des examens complémentaires ou des thérapeutiques les plus appropriées.

L'activité médicale suppose un effort de *communication*. La mémorisation des informations nécessaires à la prise de décisions n'a de sens que si les informations stockées sont accessibles immédiatement par l'ensemble de l'équipe de soin. Savoir, par exemple, qu'un patient amené inconscient à l'hôpital est diabétique ou hémophile peut être d'un précieux secours pour les premiers gestes à pratiquer. Mais les informations concernant un patient donné sont habituellement dispersées dans une ou plusieurs institutions et il importe de fournir aux différents utilisateurs une vision unique d'un dossier physiquement distribué.

L'*aide à la décision* par ordinateur a longtemps été considérée par la communauté médicale comme relevant du strict domaine de recherche, par opposition aux possibilités bien assimilées de mémorisation et de communication. Il est vrai que l'enthousiasme initial suscité par la réalisation des premiers systèmes experts a fait rapidement place à une vision plus tempérée de la décision assistée par ordinateur. Cette modération n'est pas toujours justifiée aujourd'hui. Des méthodes simples d'aide à la décision, comme des alarmes déclenchées automatiquement par la mise à jour d'un dossier médical, ont rapidement fait la preuve de leur utilité. Les indications des systèmes experts ont pu être mieux précisées et surtout leur complémentarité avec des techniques plus classiques basées sur l'exploitation statistique ou probabiliste de bases de dossiers médicaux bien mise en évidence.

Les contraintes économiques mais aussi l'éthique professionnelle imposent la mise en œuvre de procédures d'*évaluation* des technologies biomédicales. La recherche et les développements répondant à cet objectif peuvent prendre des aspects très divers comme par exemple l'analyse du contenu informatif ou décisionnel des données médicales, l'évaluation de procédures diagnostiques et thérapeutiques utilisées en routine ou l'évaluation de nouvelles technologies.

Le présent ouvrage est consacré essentiellement aux méthodes et applications hospitalières du traitement de l'information médicale. Sont ainsi exclus les aspects purement économiques ou financiers de l'informatique, l'informatisation des cabinets médicaux ou l'enseignement assisté par ordinateur. Ce choix ne préjuge bien entendu pas de l'importance de ces domaines. De même, les méthodes générales exposées dans la première partie de l'ouvrage peuvent être utilisées à profit dans ce type d'applications comme d'ailleurs dans des domaines extra-médicaux. Les méthodologies nécessaires au développement de systèmes d'information efficaces font un large appel à l'informatique, mais également aux méthodes d'organisation, de communication, à l'épidémiologie, aux techniques de documentation ou à la socio-

logie de l'entreprise. En d'autres termes, il ne s'agit plus d'informatiser une activité en prenant en compte le seul aspect technique et technologique, mais *d'intégrer* le traitement automatique de l'information dans les activités de l'équipe soignante prenant en charge le patient. Il importe de tenir compte des attentes de chacun et de distribuer les bénéfices de cette mise en place aux différents acteurs impliqués.

La première partie de l'ouvrage est consacrée à la présentation de méthodes générales utilisées dans la conception et le développement de systèmes et sous-systèmes d'information hospitaliers. Les trois premiers chapitres abordent des aspects conceptuels concernant la modélisation des informations à traiter, la conduite, l'analyse et la conceptualisation des systèmes d'information. Les trois chapitres suivants exposent des aspects plus techniques relatifs aux systèmes de gestion de bases de données, aux architectures de communication et à l'interface homme-machine. Alors que les années 80 ont été marquées par l'évolution des matériels et des réseaux de communication, il est probable que les années 90 verront des bouleversements considérables dans le domaine des logiciels, en particulier grâce à la diffusion de systèmes d'exploitation ouverts, au développement de nouvelles interfaces, de bases de données multimédias ou d'ateliers de génie logiciel. Il devient alors possible de réduire les coûts de développement des applications médicales et d'en améliorer leur qualité et leur convivialité.

La deuxième partie est consacrée à la mise en place et à la description des systèmes d'information médicaux à l'hôpital. Les caractéristiques des données médicales sont rappelées au chapitre 7. Le langage médical et les principaux langages de classification utilisés dans le codage des concepts médicaux font l'objet du chapitre 8. Le clivage fréquemment observé entre le monde soignant et le monde gestionnaire ne résiste pas à une analyse visant à améliorer l'efficacité globale du système hospitalier. Ceci est d'autant plus évident qu'un tel clivage peut se retrouver à l'intérieur même des structures administratives ou médico-techniques de l'hôpital. L'activité de soin est entièrement dépendante du bon accès aux informations concernant les patients et produites par les différents acteurs (administrateurs, médecins, biologistes, personnels de laboratoire, personnels infirmiers, secrétaires, étudiants). La même information peut être administrative, médicale ou infirmière suivant l'usage qui en est fait. L'amélioration de la fiabilité des informations, des temps d'accès à ces informations ou de leur disponibilité conduit à concevoir des systèmes d'information intégrés. Cette idée d'intégration est à la base du développement des systèmes d'information hospitaliers (SIH) qui se proposent, en favorisant la communication et le partage de l'information, de mieux utiliser les ressources humaines et matérielles de l'hôpital. Différents modèles et approches sont présentés dans le chapitre 9.

Le dossier médical du patient est un support permettant de mémoriser à long terme les informations recueillies sur un patient au cours de ses différents séjours et/ou consultations à l'hôpital. L'automatisation du dossier médical regroupe toutes les techniques visant à faciliter le recueil, l'enregistrement, la conservation et la restitution des informations contenues dans le dossier médical. Elle repose essentiellement, mais non uniquement, sur l'utilisation des méthodes informatiques. Elle n'est pas un objectif en soi mais un moyen de faciliter la tâche du médecin, de l'épidémiologiste ou du spécialiste de santé publique. Différentes approches, méthodes et techniques font l'objet du chapitre 10. Les applications du traitement des données dans les unités de soins et les laboratoires de biologie font l'objet des chapitres 11 et 12. Certaines disciplines biologiques comme la biochimie ou l'hématologie ont bénéficié de l'approche

informatique dès l'apparition des premiers ordinateurs. Les systèmes actuels pilotent des automates d'analyses capables de réaliser plusieurs centaines de dosages à l'heure et effectuent des travaux de secrétariat qui allègent considérablement la charge du personnel. Ils permettent d'acheminer, en conjonction avec les réseaux hospitaliers, les résultats dans l'unité de soins et peuvent, dans certains cas, générer des alarmes ou suggérer des diagnostics médicaux. Un autre aspect de l'application de l'informatique en instrumentation médicale mérite une mention spéciale: c'est celui du traitement automatique des images, qu'il s'agisse d'identification de cellules, d'images radiologiques, scintigraphiques ou d'applications telles que la tomодensitométrie ou la résonance magnétique nucléaire. Ces applications sont présentées succinctement dans le chapitre 13 en même temps que le problème de leur intégration dans un SIH.

Dès l'apparition des premiers ordinateurs, les médecins ont compris l'intérêt que pouvait présenter l'informatique pour les aider dans leurs tâches de décisions diagnostiques ou thérapeutiques. De très nombreux systèmes d'aide à la décision ont été réalisés dans ce domaine. Ils permettent, après la saisie d'informations relatives à un patient particulier, de proposer des hypothèses diagnostiques ou des attitudes thérapeutiques. Le chapitre 14 présente différents exemples de systèmes d'aide à la décision médicale.

L'insuffisance d'informations peut être responsable d'erreurs diagnostiques ou de thérapeutiques inappropriées. Très rapidement des banques de données informatisées ont été développées, qu'il s'agisse de banques de références ou de banques de connaissances médicales. Les banques de références ne contiennent pas l'information recherchée, mais seulement sa localisation (titre de l'article, nom de la revue scientifique...). Les banques de connaissances fournissent au contraire directement l'information recherchée, comme les propriétés ou le mode d'action d'un médicament. Ces systèmes, décrits dans le chapitre 15, permettent au clinicien de parfaire sa formation et son information dans les meilleures conditions.

La troisième partie permet d'aborder le problème de l'évaluation selon deux points de vue. Le chapitre 16 est consacré à l'évaluation des systèmes d'information et de leurs composantes considérés comme de nouvelles technologies. L'évaluation d'un environnement logiciel permettant de développer des applications médicales et l'évaluation des systèmes experts d'aide à la décision servent d'exemples. Le chapitre 17 montre au contraire comment des systèmes informatiques peuvent servir de support à la mesure et à l'analyse de l'activité médicale. Le dernier chapitre est consacré à la protection des données, aux aspects législatifs, déontologiques et éthiques du traitement de l'information médicale.

Chapitre 1

Modèles d'information

1.1. Données, Informations et connaissances

Le processus de la connaissance est une partie importante de l'activité humaine. La connaissance est le résultat de l'expérience et de l'apprentissage.

Partie I

Méthodologie générale

La méthodologie générale est une discipline qui étudie les méthodes de recherche scientifique. Elle vise à établir des principes et des règles qui régissent la conduite de la recherche. Elle est essentielle pour garantir la rigueur et la validité des résultats obtenus. Les chercheurs doivent donc maîtriser ces méthodes pour pouvoir explorer de nouvelles connaissances et résoudre des problèmes complexes. La méthodologie générale s'applique à tous les domaines de la science, de la physique à la biologie, en passant par les sciences humaines et sociales. Elle permet de structurer une recherche, de définir des objectifs clairs, de choisir des méthodes adaptées et d'analyser les résultats de manière objective. C'est un cadre de référence indispensable pour tout chercheur souhaitant contribuer à l'avancement de la connaissance.

La méthodologie générale est une discipline qui étudie les méthodes de recherche scientifique. Elle vise à établir des principes et des règles qui régissent la conduite de la recherche. Elle est essentielle pour garantir la rigueur et la validité des résultats obtenus. Les chercheurs doivent donc maîtriser ces méthodes pour pouvoir explorer de nouvelles connaissances et résoudre des problèmes complexes.

La méthodologie générale est une discipline qui étudie les méthodes de recherche scientifique. Elle vise à établir des principes et des règles qui régissent la conduite de la recherche. Elle est essentielle pour garantir la rigueur et la validité des résultats obtenus. Les chercheurs doivent donc maîtriser ces méthodes pour pouvoir explorer de nouvelles connaissances et résoudre des problèmes complexes. La méthodologie générale s'applique à tous les domaines de la science, de la physique à la biologie, en passant par les sciences humaines et sociales. Elle permet de structurer une recherche, de définir des objectifs clairs, de choisir des méthodes adaptées et d'analyser les résultats de manière objective. C'est un cadre de référence indispensable pour tout chercheur souhaitant contribuer à l'avancement de la connaissance.

Chapitre 1

Modèles d'information

1.1 Données, informations et connaissances

La poursuite de la *connaissance* est une partie importante de l'activité humaine, en tout cas l'un de ses moteurs. La connaissance revêt un double aspect : d'une part celui de l'accumulation de données (érudition) ; d'autre part celui d'une compréhension des phénomènes observés, résultats de l'analyse des relations entre les faits connus et de la vérification d'hypothèses explicatives par la recherche de nouvelles données (expérience).

“Nous acquérons des connaissances par le moyen des sens et par la réflexion que l'esprit fait sur lui-même conséquemment à l'impression des objets qui ont affecté les sens. Parmi le grand nombre des objets qui se présentent sur le vaste théâtre du monde, les sens en saisissent autant qu'il leur est possible, et confient le dépôt de ces impressions à la mémoire. Or j'appelle *matière brute* la collection de ces impressions des sens, ou les idées simples que les sens nous fournissent alors. L'esprit compare, dispose et lie ces idées simples acquises par les sens, aperçoit leurs rapports, et en forme des idées composées. De ces idées, il déduit et établit des principes pour en tirer ensuite des conclusions qui découlent naturellement des principes simples et certains, ou qui sont la conséquence de plusieurs principes compliqués ; et dans ce dernier cas ce sont les facultés réunies de l'esprit, qui agissent.”[Zimmerman 1774]

Par *données* on entend des faits élémentaires enregistrés sur des phénomènes du monde extérieur et qui peuvent être considérés comme acquis (ou presque). Le mot phénomène est compris au sens large. Un axiome ou un postulat peut être considéré comme une donnée d'une connaissance mathématique ou logique.

L'*information* peut être considérée comme l'augmentation de la connaissance qui peut être déduite d'un ensemble de données, de cette “matière brute” que constituent les faits [Langefors 1974, 1977]. La distinction entre données, informations et connaissances apparaît ainsi toute relative. La connaissance est la résultante des ajouts successifs d'informations aux données de départ, informations qui sont elles-mêmes considérées comme des données pour les inférences de niveau supérieur (figure 1.1).

Un *processus intelligent* est un processus qui a la faculté de connaître et de comprendre c'est-à-dire de manipuler des données et d'inférer de nouvelles connaissances.

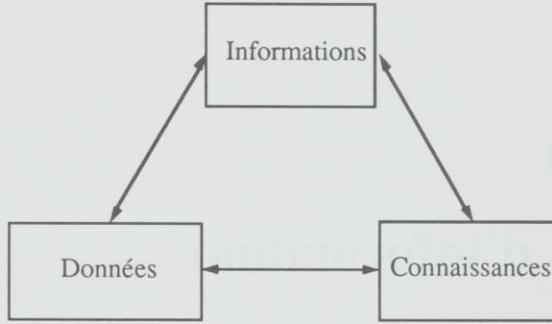


Figure 1.1 : Données, informations et connaissances

1.2 Modèles

1.2.1 De la représentation à la sémantique des informations

Le mode habituel et le plus ancien de représentation et de transmission des connaissances est le langage naturel. Ce n'est pas le seul. Depuis longtemps les hommes fixent leurs connaissances sur des supports divers (pierre, papier...) en utilisant des techniques variées (écriture, dessin...). En lui-même le langage (ou l'écriture) n'est pas la connaissance mais une suite de phonèmes (ou de caractères), c'est-à-dire seulement un mode de représentation des données et des connaissances, une *notation*. Pour comprendre cette représentation il faut pouvoir associer un sens à chacun des termes utilisés, les symboles et à leur combinaison, c'est-à-dire disposer d'un modèle de représentation des connaissances. De nombreux modèles spécifiques ont été développés en fonction des besoins propres. Ainsi, une carte de géographie utilise un modèle graphique de représentation des connaissances géographiques. La notation mathématique est basée sur un modèle de représentation des concepts mathématiques.

L'informatique et plus particulièrement les langages de programmation ont longtemps engendré une séparation nette entre les données enregistrées et leur signification [Tsichritzis 1982]. Par exemple l'expression en langage naturel "Pierre est brun et mesure 1,74m" est interprétée comme "la couleur des cheveux de Pierre (qui est une personne de sexe masculin) est brune et sa taille est de 1 mètre 74 centimètres". Les concepts de personne, de cheveux, de couleur de cheveux, de taille sont implicites pour le processeur d'information qu'est l'être humain. Au contraire, un programme informatique pourra stocker le chiffre 1,74 et les chaînes de caractères "Pierre" et "brun" sans se soucier de la signification de ces termes. Les concepts utilisés ne sont pas *a priori* reliés dans le programme.

La réalisation d'un *modèle de données ou de connaissances* essaie de combler ce fossé en associant aux données enregistrées un modèle de représentation puis de signification des données. Elle représente une évolution naturelle de l'informatique. Initialement limités par les capacités de stockage des premiers ordinateurs, les traitements informatisés se réduisaient souvent à la manipulation de données numérisées, à charge pour l'utilisateur d'en interpréter les résultats et d'en définir la signification. Le développement des techniques de stockage et de traitement du langage, mais aussi des images ou des sons permet désormais de stocker et de manipuler une connaissance plus élaborée.

1.2.2 Modèles généraux et modèles spécifiques

Chaque utilisateur d'un système informatique utilise habituellement un modèle spécifique, propre à ses besoins. La réalisation d'un modèle général vise à coordonner, à intégrer des modèles plus spécifiques. L'objectif du concepteur d'un système informatique est de définir le modèle le plus général possible. La puissance d'un modèle général peut être définie par sa capacité d'intégration de modèles plus spécifiques de la réalité telle qu'elle est vue ou sa puissance de représentation de la réalité telle qu'elle existe. De même qu'une décision politique se réfère à un modèle politique, une décision informatique se réfère à un modèle de données et/ou de connaissances. En l'absence de modèle les décisions deviennent aléatoires.

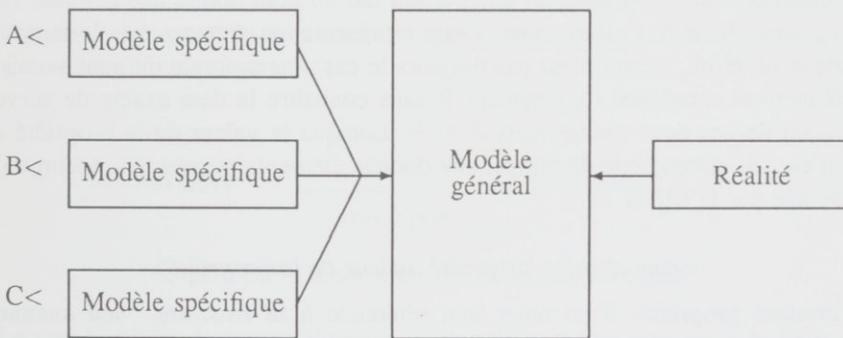


Figure 1.2 : Modèles généraux et spécifiques

1.2.3 Le système d'information

J. de Rosnay définit dans le *Macroscope* un système comme "un ensemble d'éléments en interaction dynamique, organisés en fonction d'un but . . . L'analyse des systèmes consiste à définir les limites du système, à modéliser, à identifier les éléments importants et les types d'interactions entre ces éléments, puis à déterminer les liaisons qui les intègrent en un tout organisé" [de Rosnay 1975].

Un système d'information est "un système qui a pour objectifs de rassembler, de traiter, de manipuler et de fournir les informations nécessaires à certaines activités" [Planche 1988]. Il peut donc comporter des éléments manuels et des éléments informatisés.

Un modèle de système d'information est une représentation d'un système, cherchant à en faciliter la compréhension. Un modèle est nécessairement une simplification de la réalité, une abstraction cachant certains détails pour en mettre en valeur d'autres. Pour Tardieu "la construction d'un système d'information se situera en permanence entre :

- l'analyse de l'organisation et de son système d'information perçu comme un objet naturel,
- la conception d'un objet artificiel ayant pour but de représenter cette organisation pour en améliorer l'efficacité [Tardieu 1984].

1.3 Concepts de modélisation

1.3.1 Objets et données élémentaires

Une *donnée élémentaire* représente la plus petite quantité d'information manipulable, l'atome d'information. Langefors [Langefors 1977] la définit par exemple par le quadruplet :

<nom d'objet, propriété de l'objet, valeur de la propriété, temps>

Le mot *objet* est compris ici au sens large, puisqu'un objet peut avoir (objet concret) ou ne pas avoir (objet abstrait) d'existence physique. Le terme *propriété* de l'objet représente une caractéristique quelconque de l'objet, par exemple sa longueur, caractéristique qui prend pour un objet défini par un nom donné une certaine valeur à un certain moment d'observation. Cette représentation suppose que le temps soit parfaitement défini, ce qui n'est pas toujours le cas. Par exemple on peut savoir que l'événement A a précédé l'événement B sans connaître la date exacte de survenue de chacun de ces deux événements A et B. Lorsque la valeur de la propriété d'un objet n'est pas dépendante du temps, une donnée élémentaire peut, par exemple, être représentée par le triplet :

<nom d'objet, propriété, valeur de la propriété>

Certaines propriétés d'un objet font référence à sa structure, "son anatomie". On parlera de *propriétés structurales*. Un objet physique est fait de constituants. Il a une taille, une couleur, etc. D'autres propriétés font référence à sa fonction, "sa physiologie". Le cœur fait circuler le sang, les globules rouges amènent l'oxygène aux tissus, un hôpital permet de soigner des malades, un programme informatique effectue certains nombres d'opérations. On parle alors de *propriétés comportementales*.

1.3.2 Liaisons et catégories

Un moyen graphique simple pour représenter des données et des connaissances consiste à faire figurer chacun des objets dans un réseau et à représenter les *liaisons* deux à deux entre les objets. Considérons par exemple la phrase "Mr Dupont, âgé de 40 ans est traité par de l'aspirine". Cette phrase peut être représentée sous forme du graphe de liaisons de la figure 1.3.

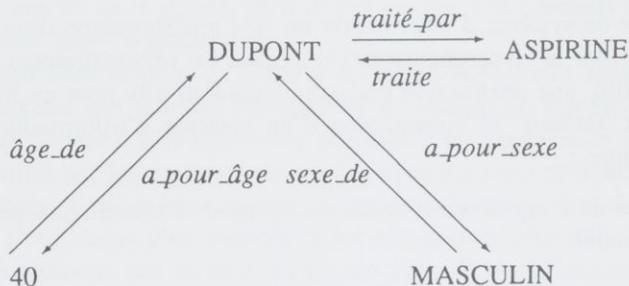


Figure 1.3 : Graphe de liaisons

Un autre moyen, complémentaire du premier consiste à regrouper des objets similaires en *catégories* (par exemple une catégorie de patients, une catégorie de médicaments). Dans ce cas l'appartenance d'un objet à une catégorie définit implicitement une liaison entre l'objet et sa catégorie d'appartenance. Par exemple le fait que Dupont soit considéré comme un objet de la catégorie *PATIENT* est équivalent à la représentation de la figure 1.4.

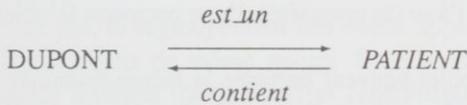


Figure 1.4 : Appartenance d'un objet à une catégorie

Il devient alors possible de représenter globalement les liaisons entre les catégories d'objets ainsi définies (figure 1.5).

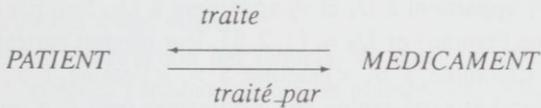


Figure 1.5 : Liaison entre deux catégories d'objets

La combinaison de façons diverses de ces deux techniques est à la base de la plupart des modèles de données, comme par exemple le modèle sémantique binaire décrit par Abrial [Abrial 1974]. Dans ce modèle, chaque nœud du graphe correspond à une catégorie. Un arc entre deux catégories définit un lien atomique entre des paires d'objets appartenant à chaque catégorie. Une catégorie est définie par un nom de catégorie et chaque lien par une double application, une application f_1 de la catégorie C_1 vers la catégorie C_2 et une application f_2 réciproque de f_1 de la catégorie C_2 vers la catégorie C_1 .

D'une façon générale, un modèle de données *strictement typé* est un modèle dans lequel chaque donnée appartient à une catégorie bien définie et homogène. Dans ce modèle, les propriétés des données peuvent être inférées des propriétés des catégories auxquelles elles appartiennent. A l'inverse, un modèle *faiblement typé* est un modèle qui ne fait aucune hypothèse sur les catégories de données et peut donc mélanger données et catégories de données. La plupart des modèles de données utilisés en informatique sont des modèles de données fortement typés.

1.3.3 Ensembles, domaines et attributs

Les notions d'ensemble et de domaine se rapportent directement à la notion de catégorie. Un *ensemble* (*set*) est une collection finie ou infinie, non ordonnée, d'objets proprement identifiés et caractérisés par une condition d'appartenance. Par exemple, les entiers positifs $\{1, 2, 3, \dots\}$ représentent un ensemble infini et les malades hospitalisés dans un établissement de soins un ensemble fini.

L'*extension* ou réalisation d'un ensemble spécifie une occurrence observée (*token*). Par exemple l'ensemble $\{\text{bleu, blanc, rouge}\}$ représente une extension, parmi les nombreuses extensions possibles, de l'ensemble *COULEUR*.

L'intension¹ d'un ensemble définit toutes les occurrences possibles de l'ensemble par spécification des conditions d'appartenance. L'intension représente en quelque sorte la réunion de toutes les extensions possibles d'un domaine. Elle supporte ainsi une partie de la signification de cet ensemble.

Un *domaine* dans un modèle de données est défini comme l'ensemble homogène des valeurs qui peuvent être prises par une propriété d'un objet [Childs 1968, Childs 1977]. Par exemple le domaine des entiers de 1 à 100 constitue l'ensemble des valeurs qui peuvent être prises par l'âge (la propriété) d'une personne (l'objet considéré dans cet exemple).

Deux propriétés d'un objet peuvent partager le même domaine (par exemple le domaine des dates ou celui des entiers). Il apparaît alors important d'associer un nom appelé *attribut* à chacune des significations de ces domaines. L'attribut représente alors l'intension du domaine et les valeurs prises par l'attribut pour un ensemble particulier d'objets, l'extension du domaine. Par exemple 70 représente la valeur de l'attribut âge pour un patient donné.

Le *produit cartésien* de 2 domaines D_1 et D_2 est l'ensemble des doublets $\langle v_1, v_2 \rangle$ tels que v_1 appartient à D_1 et v_2 appartient à D_2 . Soit par exemple les domaines $D_1 = \{\text{Homme, Femme}\}$ et $D_2 = \{1, 2, 3\}$, leur produit cartésien est $D_1 \times D_2 = \{\langle \text{Homme}, 1 \rangle, \langle \text{Homme}, 2 \rangle, \langle \text{Homme}, 3 \rangle, \langle \text{Femme}, 1 \rangle, \langle \text{Femme}, 2 \rangle, \langle \text{Femme}, 3 \rangle\}$. Le produit cartésien de ces 2 domaines peut être représenté comme un tableau à 2 dimensions où chaque ligne correspond à l'un des doublets définis précédemment (figure 1.6).

Homme	1
Homme	2
Homme	3
Femme	1
Femme	2
Femme	3

Figure 1.6 : Produit cartésien de deux domaines

Le produit cartésien de n domaines est l'ensemble des n -uplets ou *tuples* d'ordre n $\langle v_1, v_2, \dots, v_n \rangle$ tels que, pour toute valeur de i comprise entre 1 et n , v_i appartient à D_i . Il peut être représenté comme un tableau à n colonnes, dans lequel chaque ligne est un n -uplet et chaque colonne correspond à l'un des domaines.

1.3.4 Notion d'abstraction et héritage de propriétés

D'une façon générale, le concept d'*abstraction* peut se définir comme la facilité de cacher le détail pour se concentrer sur les propriétés générales et communes d'un ensemble d'objets. En établissant ces propriétés générales, l'abstraction permet d'établir des relations entre différentes catégories d'objets. La généralisation et l'agrégation sont les deux modalités d'abstraction les plus couramment utilisées [Smith 1977a, Smith 1977b].

La *généralisation* permet le regroupement d'éléments ou de types d'éléments proches. Elle exprime les concepts *EST_UN* (*IS_A*) et *SORTE_DE* (*A_KIND_OF*).

¹Le terme intension est écrit dans ce texte avec un *s* et non un *t* comme dans intention. Sa signification est celle introduite par Gottlob Frege et Alonzo Church [Frege 1949, Apostel 1967].

Par exemple le regroupement de différentes molécules de pénicilline *A, G, M* permet de constituer le concept de *PENICILLINE* ou type d'objet *PENICILLINE*. Le type d'objet *PENICILLINE* et le type d'objet *CEPHALOSPORINE* constituent le type d'objet *ANTIBIOTIQUE*.

pénicilline_A et pénicilline_G *SONT_DES* Pénicillines
pénicillines et céphalosporines *SONT_UNE_SORTE_DE* antibiotique

Dans le premier cas, le regroupement des objets appelés pénicillines correspond à un regroupement d'objets de même nature. Dans le cas suivant, il s'agit du regroupement de types d'objets (pénicillines et céphalosporines), définis à un niveau d'abstraction supérieur. La distinction entre ces deux niveaux d'abstraction est bien sûre relative. L'inverse de la généralisation est la *spécialisation*

Suivant les cas, un type obtenu par généralisation peut hériter, soit de l'ensemble, soit d'une partie seulement des propriétés de ses constituants. Par exemple, le fait que toutes les personnes ont un nom et un âge est hérité du fait que chaque personne a un nom et un âge. Au contraire des propriétés peuvent être spécifiques d'un type, par exemple le concept d'âge moyen.

L'*agrégation* est l'abstraction par laquelle un objet est construit à partir de ses constituants (ex : appareil cardiovasculaire ou pulmonaire). Elle exprime le concept *EST_UN_ELEMENT_DE (PART_OF)*. Elle permet de représenter la structure interne d'un objet. Dans l'exemple de la figure 1.7, le concept de personne est obtenu par agrégation des concepts de nom, adresse, sexe, âge, etc. Le nom est considéré comme un constituant du concept de personne. Le concept de personne est abstrait à partir des concepts de malades et de médecins. L'adresse d'un malade a la même signification sémantique que l'adresse d'un médecin ; le concept d'adresse peut être rattaché au concept de personne. La figure 1.8 donne une extension possible d'une partie de ce modèle.

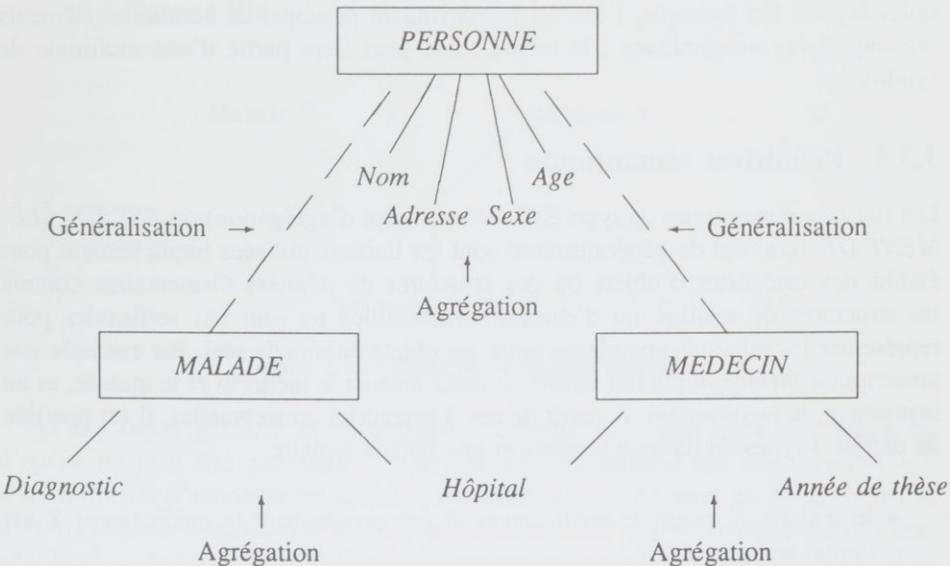


Figure 1.7 : Agrégation et généralisation

