

METHODE + PROGRAMMES

INFORMATIQUE DOCUMENTAIRE

A. DEWEZE

Préface des Pr. J. CORDONNIER et R. BOUCHÉ

3^e édition

MASSON 

BIBLIOTHEQUE DU CERIST

**INFORMATIQUE
DOCUMENTAIRE**

CHEZ LE MÊME ÉDITEUR

Du même auteur :

L'ACCÈS EN LIGNE AUX BASES DOCUMENTAIRES. *ABC des langages*, 1983, 176 pages.

Autres ouvrages :

SQL et DB2. Le relationnel et sa pratique. Par M. KOUTCHOUK, *Méthodes Informatiques et Pratique des Systèmes*, 1989, 224 pages.

COMPRENDRE LES BASES DE DONNÉES. Théorie et pratique. Par A. MESGUICH et B. NORMIER. *Collection MIM*, 1987, 5^e tirage, 216 pages.

BASES DE DONNÉES ET SGDB. De la conception à la mise en œuvre. Démarche pratique. Par F. KRAMARZ et O. PERRAULT. *Méthodes Informatiques et Pratique des Systèmes*, 1986, 200 pages.

LES VIDÉODISQUES. Par G. BROUSSAUD, préface de P. AIGRAIN. *Collection Technique et Scientifique des Télécommunications*. 1986, 216 pages.

LES FICHIERS INFORMATIQUES. Conception et performances. Par O. HANSON, traduit de l'anglais par E. DAVID, *collection MIM*, 1986, 176 pages.

TECHNIQUES DE LA BUREAUTIQUE. Par M. POLITIS, 1985, 2^e édition revue et augmentée, 208 pages.

ORDINATEURS, INTERFACES ET RÉSEAUX DE COMMUNICATION. Par S. COLLIN, *collection MIM*, 1988, 120 pages.

RÉSEAUX ET MICRO-ORDINATEURS. Par P.H. JESTY, traduit de l'anglais par I. LERESTIE, *collection MIM*, 1987, 144 pages.

DICTIONNAIRE D'INFORMATIQUE. Par M. GINGUAY et A. LAURET, 1987, 3^e édition refondue et augmentée, 344 pages.

DICTIONNAIRE DES TÉLÉCOMMUNICATIONS, anglais-français. Par J. DE LUCA, 1988, 408 pages.

LES PARADIS INFORMATIONNELS. Du Minitel aux services de communication du futur. Par M. MARCHAND et le SPES. *Collection Technique et Scientifique des Télécommunications*, 1987, 256 pages.

Liste complète des ouvrages de la collection en pages 2 et 3 de couverture.

METHODE + PROGRAMMES

Collection coordonnée par J.-F. PHELIZON

INFORMATIQUE DOCUMENTAIRE

André DEWEZE

*Docteur de l'Université Claude-Bernard (Lyon)
Sciences Mathématiques*

Préface des Professeurs

Jacques CORDONNIER et Richard BOUCHÉ

Troisième édition actualisée et augmentée

MASSON Paris Milan Barcelone Mexico **1989**

Cet ouvrage a été publié en langue espagnole par Masson Barcelone, en 1988.

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, 1985, 1989

ISBN : 2-225-81702-2

ISSN : 0757-2093

MASSON
MASSON ITALIA EDITORI
MASSON S.A.
MASSON EDITORES

120, bd Saint-Germain, 75280 Paris Cedex 06
Via Statuto 2, 20121 Milano
Balmes 151, 08008 Barcelona
Dakota 383, Colonia Napoles, 03810 Mexico D.F.

PRÉFACE

Dans le secteur de la documentation, la part de l'informatique a beaucoup évolué depuis quelques années. Réservé, au début, aux centres de documentation « riches » des entreprises et des administrations d'une certaine importance, l'ordinateur s'est mis à la portée d'un plus grand nombre d'utilisateurs quand le développement des réseaux de télécommunication en a facilité l'accès à distance. Mais, même à ce stade, la base de données documentaires reste quelque chose d'assez lourd à mettre en œuvre et nécessite un gros ordinateur serveur et une équipe importante d'informaticiens. Le documentaliste doit savoir identifier les serveurs et les bases de données qui lui permettent de trouver la meilleure réponse à une question, mais il continue à gérer sa bibliothèque et son fonds local d'une façon traditionnelle.

L'arrivée de micro-ordinateurs d'une certaine puissance pour un coût qui ne cesse de décroître modifie cette situation. Maintenant, le documentaliste se voit offrir des programmes de gestion de bibliothèque et des logiciels de documentation qui lui permettent de construire ses propres bases de données.

Il doit maintenant être capable de définir ses besoins, d'identifier le matériel qui pourrait le mieux les satisfaire, et choisir les logiciels les mieux adaptés aux applications à mettre en place. Bref, il doit savoir rédiger un cahier des charges qui lui permettra de dialoguer avec les vendeurs d'informatique et de critiquer leurs propositions.

Mais tout documentaliste sait que la définition d'un langage de représentation ou de thésaurus est une opération essentielle pour la gestion de son fonds documentaire. Si un ordinateur est utilisé, incapable d'une quelconque interprétation, il va exiger la plus grande rigueur dans l'expression des éléments du thésaurus. Aucun à peu près ne pourra être de mise et le documentaliste devra en tenir compte. Il y a de fortes chances aussi pour que le centre de documentation s'associe avec d'autres pour former un réseau. Le documentaliste se trouvera alors partie prenante dans un groupe de travail chargé de mettre au point le thésaurus commun.

On constate donc que l'intervention de l'informatique conduit le documentaliste à des tâches d'analyse et de conception qui exigent de lui, à la fois une certaine maîtrise de l'outil informatique, des capacités d'analyse pour la définition du cahier des charges et des connaissances plus fondamentales, utiles à la conception des thésaurus.

Un ouvrage sur l'informatique documentaire ne peut être complet que s'il aborde tous ces aspects et André Deweze était certainement parmi les personnes les mieux placées pour faire un tel livre. Son expérience professionnelle d'ancien

responsable de la documentation de l'entreprise industrielle Merlin-Gerin lui permet de parler en praticien des problèmes d'informatisation d'un centre de documentation et les chapitres sur le cahier des charges et l'évaluation des coûts doivent être d'une grande aide pour le documentaliste confronté à ces difficultés.

Mais André Deweze n'est pas qu'un praticien : très intéressé par les problèmes de recherche, il a été rédacteur en chef pendant 23 ans de la revue « *TA Informations* » spécialisée dans le traitement informatique des langues et il a participé aux travaux du groupe SYDO* ce qui l'a conduit à soutenir une thèse de doctorat de l'Université Claude Bernard à Lyon sur les réseaux sémantiques. Il a également été l'animateur d'un groupe de travail qui a conçu le thésaurus multilingue sur l'électricité et l'électronique : *THÉSÉE*. Il est donc l'homme capable de faire le pont entre la recherche et l'application et son livre apporte au documentaliste un certain nombre de connaissances fondamentales précieuses pour la conception des systèmes d'information documentaire.

Enfin, André Deweze a aussi participé à des enseignements de type DESS (Diplôme d'Études Supérieures Spécialisées) à Lyon et à Grenoble et son ouvrage est incontestablement, pour l'étudiant en Sciences de l'Information, un manuel extrêmement utile qui faisait cruellement défaut jusqu'à présent dans le monde francophone. A ce titre, nous ne pourrions mieux faire que de le préconiser comme manuel de base dans les écoles de l'AIESI**.

Jacques CORDONNIER

Professeur à l'École des Bibliothécaires
de GENEVE (Suisse)
Président de l'AIESI

Richard BOUCHÉ

Professeur en Sciences de l'Information
à l'Université LYON I (France)

* Le groupe SYDO rassemble des équipes de recherche des Universités de Lyon I, Lyon II et Grenoble II (France), Fribourg (Suisse), Mayence (Allemagne) dans le domaine de l'application du traitement automatique des langues aux systèmes d'information documentaire.

** L'AIESI est l'Association Internationale des Écoles en Sciences de l'Information (francophones). Elle se situe dans le cadre de l'AUPFL (Association des Universités Partiellement ou Entièrement de Langue Française).

TABLE DES MATIERES

PRÉFACE	V
INTRODUCTION	1
1. LA CHAÎNE DOCUMENTAIRE	3
1.1. Production, transfert et utilisation de l'information	3
1.2. Place de l'ordinateur dans la chaîne documentaire	5
1.3. Les flux, les caractéristiques et les supports de l'information	7
1.4. Rôle du service de documentation dans la chaîne documentaire	9
2. FONCTIONS ET ORGANES DE L'ORDINATEUR	11
2.1. Unités fonctionnelles de l'ordinateur	11
2.2. Organes d'entrée	11
2.3. Organes de sortie	12
2.4. Représentation de l'information	13
2.4.1. De l'idéogramme au caractère alphabétique et à sa représentation binaire	13
2.4.2. Code des cartes perforées (code Hollerith)	16
2.4.3. Codes binaires	17
2.4.4. Codes alphanumériques à 8 bits	19
2.4.5. Décimal codé binaire (DCB)	21
2.4.6. Système hexadécimal	21
2.5. Supports d'enregistrement de l'information	23
2.5.1. Carte perforée	23
2.5.2. Bande magnétique	25
2.5.3. Disquettes souples	28
2.5.4. Unité de disques «durs»	29
2.5.5. Mémoire de l'Unité Centrale	29
2.5.6. Sauvegarde des informations	31
2.6. Accès à l'information	31
2.6.1. Accès séquentiel	32
2.6.2. Accès sélectif	32
2.7. Traitement de l'information	33
2.7.1. Organisation et rôle de la mémoire centrale	33
2.7.2. L'unité arithmétique et logique	34
2.7.3. L'exécution des instructions	34
2.7.4. Traitement d'une chaîne de caractères	35
2.8. Les modes d'exploitation	38
2.9. Systèmes d'exploitation	38
3. ORDINATEUR ET PRODUCTION DE DOCUMENTS	39
3.1. Imprimantes à impact	40
3.1.1. Imprimantes à sphère ou à marguerite	40
3.1.2. Imprimantes à aiguilles	40
3.1.3. Imprimantes à chaîne	41
3.2. Procédés de composition sans impact	42
3.2.1. Procédé thermique	42
3.2.2. Procédé par jet d'encre	42
3.2.3. Procédé électrostatique	42
3.2.4. Photocomposition optique avec disque ou tambour	43
3.2.5. Photocomposition avec tube cathodique	43
3.2.6. Photocomposition par laser	44

3.2.7. Sortie d'ordinateur sur microforme (COM)	45
3.2.8. Choix du mode opératoire	45
3.3. Photocomposition programmée par ordinateur	49
3.4. Production informatique d'outils documentaires	51
3.4.1. Bulletin analytique "Analyses"	51
3.4.2. Index auteurs ou index mots-clés	51
3.4.3. Bulletin d'index pour les recherches bibliographiques	55
3.5. Traitement de texte	58
 4. ORDINATEUR ET ARCHIVAGE DE DOCUMENTS	59
4.1. Indexation des microformes et COM	59
4.2. Le disque optique numérique	60
4.2.1. Représentation et lecture de l'information	60
4.2.2. Analyse de l'image - Compression des données	61
4.3. Système d'exploitation des disques optiques	63
4.4. Etat de la technique - Perspectives d'évolution	64
4.5. Les applications du disque optique	64
4.5.1. Archives électroniques	64
4.5.2. Diffusion de banques de données	65
4.5.3. Télécopie de documents primaires	66
4.5.4. Diffusion d'ouvrages par les éditeurs	66
 5. INFORMATISATION D'UNE BIBLIOTHEQUE	67
5.1. Procédures d'approvisionnement des documents	69
5.2. Formalisation des opérations de bibliothèque	77
5.3. Représentation des opérations de bibliothèque par un système de relations entre les éléments d'ensembles de données	79
5.4. Structure et représentation d'un fichier	83
5.5. Caractéristiques d'un fichier lecteurs	89
5.6. Caractéristiques du fichier documents	89
5.7. Gestion des abonnements	97
5.8. Gestion des arrivées de périodiques	101
5.8.1. Objectifs du contrôle	101
5.8.2. Simulation à l'aide d'un kardex manuel	101
5.8.3. Ecran de contrôle des arrivées d'une revue	103
5.8.4. Procédure de dépistage des anomalies de réception	106
5.9. Circulation des revues - Prêt des livres	109
5.10. Production d'une fiche individuelle par lecteur	111
5.11. Listes de diffusion	111
5.12. Liens entre fichiers	114
5.13. Impression d'un catalogue de bibliothèque	115
5.13.1. Création du fichier CATLIV TRIE	117
5.13.2. Création du fichier CATEDI	119
5.14. Création de modèles d'états avec OMNIS	121
 6. ANALYSE DE CONTENU DES DOCUMENTS	125
6.1. Constitution de la référothèque	125
6.2. Objectifs des référothèques	125
6.3. Procédure d'établissement d'un résumé	128
6.4. Méthodes d'indexation	131
6.4.1. Mise-à-jour automatique du fichier inverse en langage naturel	131
6.4.2. Indexation manuelle avec un lexique contrôlé	132
6.4.3. Indexation à l'aide d'un thésaurus	133
6.4.4. Indexation avec analyseur automatique de texte	133
6.5. Fonctionnement d'un analyseur morphographique	137
6.5.1. Fonctions de l'analyseur morphographique (AM)	137

6.5.2. Structure des fichiers de l'analyseur morphographique	139
6.5.3. Constitution de l'antidictionnaire	145
6.5.4. Algorithme de reconnaissance d'un mot	146
6.5.5. Exploitation des résultats de l'analyseur morphographique	147
7. CONCEPTION ET REALISATION D'UN THESAURUS	150
7.1. Présentation du thésaurus	150
7.1.1. Accès par liste alphabétique	150
7.1.2. Accès par champs sémantiques	151
7.1.3. Accès par arborescence hiérarchique	151
7.2. Représentation informatique du thésaurus	152
7.2.1. Fichier dictionnaire	152
7.2.2. Fichier des relations hiérarchiques	153
7.2.3. Fichier des relations associatives	156
7.2.4. Fichier des substitutions	159
7.2.5. Table des champs sémantique	160
7.3. Collecte et structuration des descripteurs pour l'élaboration d'un thésaurus	160
7.3.1. Collecte du vocabulaire dans un domaine de spécialité	160
7.3.2. Structure du vocabulaire dans un domaine de spécialité	161
7.4. Analyseur morphographique, utilisation et construction de thésaurus	162
7.5. Polyémie et thésaurus multilingue	164
8. CONCEPTION ET UTILISATION DES BASES DE DONNEES BIBLIOGRAPHIQUES	165
8.1. Système de saisie de la référothèque	165
8.1.1. Modes de saisie	165
8.1.2. Structure d'un fichier référothèque	167
8.1.3. Spécifications des contrôles de saisie	167
8.1.4. Contrôle des mots-clés	170
8.1.5. Choix de la structure d'un fichier	171
8.1.6. Ecrans de saisie-validation des données	175
8.2. Organisation des fichiers inverses	177
8.3. Exploitation de la référothèque	183
8.4. Critères de performances d'un système d'interrogation	187
8.5. Brève comparaison des logiciels d'interrogation	188
9. CAHIER DES CHARGES D'UNE APPLICATION INFORMATIQUE	189
9.1. Définition de l'application - Rôle du cahier des charges	189
9.2. Analyse de l'existant	190
9.3. Description sommaire de l'application - Objectifs poursuivis	191
9.4. Présentation des résultats	192
9.4.1. Catalogue des ouvrages	193
9.4.2. Liste des acquisitions du mois	193
9.4.3. Fiche de prêt des ouvrages	193
9.4.4. Fiche individuelle de lecteur	194
9.4.5. Dossier d'étudiant	194
9.5. Données nécessaires	195
9.5.1. Provenance des données - Codes à utiliser	195
9.5.2. Présentation des données - Modèles d'écran	195
9.5.3. Volume des données	196
9.6. Inventaire des fichiers	198
9.6.1. Date système	199
9.6.2. Fichier lecteurs (étudiants et personnel) et tables associées	199
9.6.3. Fichiers documents et tables associées	200
9.6.4. Fichiers de relations	202

9.6.5. Fichier "kardex" (KARD)	203
9.6.6. Fichier "commentaires" (COMM)	203
9.6.7. Lexique de termes contrôlés ou thésaurus (THES)	203
9.6.8. Fichier "traitement de texte" (TEXT)	204
9.6.9. Tables d'Index	204
9.7. Traitements - Calculs	205
9.8. Contrôles des données	205
9.8.1. Nature et responsables des contrôles	205
9.8.2. Procédure de contrôle avant la création d'un enregistrement dans le fichier documents	206
9.8.3. Procédure de contrôle avant la création d'un enregistrement dans le fichier document	208
9.9. Sauvegarde et durée de conservation	210
9.9.1. Sauvegarde du disque dur - Sauvegarde des disquettes	210
9.9.2. Durée de conservation	212
9.10. Exploitation	213
9.10.1. Dossiers d'exploitation	213
9.10.2. Catégories d'utilisateurs - Mots de passe	214
9.10.3. Mode opératoire de consultation	215
9.10.4. Mode opératoire de maintenance	217
9.10.5. Implantation des terminaux	219
9.11. Déroulement du projet	219
9.11.1. Préparation des données	219
9.11.2. Création des fichiers et programmation	220
9.11.3. Saisie des données	221
9.12. Développements ultérieurs	221
9.13. Conclusions sur la généralisation de l'étude	222
10. COUTS D'UNE APPLICATION INFORMATIQUE	223
10.1. Coûts de réalisation d'une application informatique	223
10.1.1. Frais de personnel pour la réalisation de l'application	224
10.1.2. Maintenance des logiciels - Amortissement des études	225
10.1.3. Frais d'équipement pour la réalisation de l'application	225
10.2. Coûts d'exploitation d'une application informatique	225
10.2.1. Frais de personnel pour l'exploitation de l'application	225
10.2.2. Frais d'équipement pour l'exploitation de l'application	228
10.2.3. Frais de fourniture pour l'exploitation de l'application	228
10.2.4. Frais de postes et télécommunications	228
10.2.5. Frais logistiques	228
10.3. Coûts de préparation des données d'entrée	228
10.3.1. Coûts d'acquisition des documents primaires	228
10.3.2. Coûts d'analyse et d'indexation	229
10.3.3. Coûts de saisie des données	229
11. CHOIX DES SOLUTIONS	229
11.1. Critères de choix d'un équipement	229
11.1.1. Equipement propre au service de documentation	230
11.1.2. Partage des ressources d'un ordinateur central	231
11.2. Critères de choix d'un logiciel	232
11.2.1. Choix d'un progiciel spécifique	232
11.2.2. Développement d'une application à partir de logiciels et de modules banalisés	233
11.2.3. Les logiciels documentaires pour micro-ordinateurs	234
11.3. Considérations de rentabilité	235

12. CREATION D'UNE BASE DE DONNEES RELATIONELLE POUR MICRO-ORDINATEURS	236
12.1. Caractéristiques d'une base de données relationnelle	237
12.1.1. Propriétés d'un fichier informatique	237
12.1.2. Propriétés d'une base de données relationnelle	239
12.1.3. Caractéristiques du SGBD relationnel 4ème dimension	240
12.1.4. Recherche de fiches - Temps d'accès	240
12.1.5. Economie de 4D sur le disque	241
12.2. Exemple: Création d'un annuaire d'entreprise	242
12.2.1. Objectifs de l'application	242
12.2.2. Création de la base - Analyse préliminaire	243
12.2.3. Création des procédures	247
12.3. Création d'une base de données documentaires	252
12.3.1. Objectifs de l'application	252
12.3.2. Création de la base - Analyse préliminaire	254
12.4. Gestion des commandes et des abonnements	257
12.4.1. Définitions	257
12.4.2. Objectifs de l'application	261
12.4.3. Analyse organique - Définition des fichiers	263
12.4.4. Commandes	269
12.4.5. Abonnements	269
12.5. Basethèque: Base relationnelle pour applications documentaires	269
12.5.1. Fonctions de Basethèque	269
12.5.2. Saisie des documents	271
12.5.3. Recherche de documents dans la base et diffusion sélective	271
12.5.4. Visualisation ou impression des résultats	274
12.5.5. Gestion de Bibliothèque	274
BIBLIOGRAPHIE COMMENTÉE	275
INDEX	279