



HAL
open science

Gestion et interrogation d'une base de données communales sur micro ordinateur

Agnès Jauneau

► **To cite this version:**

Agnès Jauneau. Gestion et interrogation d'une base de données communales sur micro ordinateur. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 1983. Français. NNT: . tel-00308109

HAL Id: tel-00308109

<https://theses.hal.science/tel-00308109>

Submitted on 29 Jul 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée à

l'Institut National Polytechnique de Grenoble

pour obtenir le grade de
DOCTEUR DE 3ème CYCLE
«Génie informatique»

par

Agnès JAUNEAU



**GESTION ET INTERROGATION
D'UNE BASE DE DONNEES COMMUNALES
SUR MICRO-ORDINATEUR.**



Thèse soutenue le 29 juin 1983 devant la commission d'examen.

C. DELOBEL] Président Examineurs
L. BOLLIET	
J. COURTIN	
S. LAVALLEE	



INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

Année universitaire 1982-1983

Président de l'Université : D. BLOCH

**Vice-Président : René CARRE
Hervé CHERADAME
Marcel IVANES**

PROFESSEURS DES UNIVERSITES :

ANCEAU François	E.N.S.I.M.A.G.
BARRAUD Alain	E.N.S.I.E.G.
BAUDELET Bernard	E.N.S.I.E.G.
BESSON Jean	E.N.S.E.E.G.
BLIMAN Samuel	E.N.S.E.R.G.
BLOCH Daniel	E.N.S.I.E.G.
BOIS Philippe	E.N.S.H.G.
BONNETAIN Lucien	E.N.S.E.E.G.
BONNIER Etienne	E.N.S.E.E.G.
BOUVARD Maurice	E.N.S.H.G.
BRISSONNEAU Pierre	E.N.S.I.E.G.
BUYLE BODIN Maurice	E.N.S.E.R.G.
CAVAIGNAC Jean-François	E.N.S.I.E.G.
CHARTIER Germain	E.N.S.I.E.G.
CHENEVIER Pierre	E.N.S.E.R.G.
CHERADAME Hervé	U.E.R.M.C.P.P.
CHERUY Arlette	E.N.S.I.E.G.
CHIAVERINA Jean	U.E.R.M.C.P.P.
COHEN Joseph	E.N.S.E.R.G.
COUMES André	E.N.S.E.R.G.
DURAND Francis	E.N.S.E.E.G.
DURAND Jean-Louis	E.N.S.I.E.G.
FELICI Noël	E.N.S.I.E.G.
FOULARD Claude	E.N.S.I.E.G.
GENTIL Pierre	E.N.S.E.R.G.
GUERIN Bernard	E.N.S.E.R.G.
GUYOT Pierre	E.N.S.E.E.G.
IVANES Marcel	E.N.S.I.E.G.
JAUSSAUD Pierre	E.N.S.I.E.G.
JOUBERT Jean-Claude	E.N.S.I.E.G.
JOURDAIN Geneviève	E.N.S.I.E.G.
LACOUME Jean-Louis	E.N.S.I.E.G.
LATOMBE Jean-Claude	E.N.S.I.M.A.G.

.../...

LESSIEUR Marcel	E.N.S.H.G.
LESPINARD Georges	E.N.S.H.G.
LONGQUEUE Jean-Pierre	E.N.S.I.E.G.
MAZARE Guy	E.N.S.I.M.A.G.
MOREAU René	E.N.S.H.G.
MORET Roger	E.N.S.I.E.G.
MOSSIERE Jacques	E.N.S.I.M.A.G.
PARIAUD Jean-Charles	E.N.S.E.E.G.
PAUTHENET René	E.N.S.I.E.G.
PERRET René	E.N.S.I.E.G.
PERRET Robert	E.N.S.I.E.G.
PIAU Jean-Michel	E.N.S.H.G.
POLOUJADOFF Michel	E.N.S.I.E.G.
POUPOT Christian	E.N.S.E.R.G.
RAMEAU Jean-Jacques	E.N.S.E.E.G.
RENAUD Maurice	U.E.R.M.C.P.P.
ROBERT André	U.E.R.M.C.P.P.
ROBERT François	E.N.S.I.M.A.G.
SABONNADIERE Jean-Claude	E.N.S.I.E.G.
SAUCIER Gabrielle	E.N.S.I.M.A.G.
SCHLENKER Claire	E.N.S.I.E.G.
SCHLENKER Michel	E.N.S.I.E.G.
SERMET Pierre	E.N.S.E.R.G.
SILVY Jacques	U.E.R.M.C.P.P.
SOHM Jean-Claude	E.N.S.E.E.G.
SOUQUET Jean-Louis	E.N.S.E.E.G.
VEILLON Gérard	E.N.S.I.M.A.G.
ZADWORNY François	E.N.S.E.R.G.

PROFESSEURS ASSOCIES

BASTIN Georges	E.N.S.H.G.
BERRIL John	E.N.S.H.G.
CARREAU Pierre	E.N.S.H.G.
GANDINI Alessandro	U.E.R.M.C.P.P.
HAYASHI Hirashi	E.N.S.I.E.G.

PROFESSEURS UNIVERSITE DES SCIENCES SOCIALES (Grenoble II)

BOLLIET Louis
Chatelin Françoise

PROFESSEURS E.N.S. Mines de Saint-Etienne

RIEU Jean.
SOUSTELLE Michel

CHERCHEURS DU C.N.R.S.

FRUCHART Robert
VACHAUD Georges

Directeur de Recherche
Directeur de Recherche

.../...

ALLIBERT Michel	Maître de Recherche
ANSARA Ibrahim	Maître de Recherche
ARMAND Michel	Maître de Recherche
BINDER Gilbert	
CARRE René	Maître de Recherche
DAVID René	Maître de Recherche
DEPORTES Jacques	
DRIOLE Jean	Maître de Recherche
GIGNOUX Damien	
GIVORD Dominique	
GUELIN Pierre	
HOPFINGER Emil	Maître de Recherche
JODD Jean-Charles	Maître de Recherche
KAMARINOS Georges	Maître de Recherche
KLEITZ Michel	Maître de Recherche
LANDAU Ioan-Dore	Maître de Recherche
LASJAUNIAS J.C.	
MERMET Jean	Maître de Recherche
MUNIER Jacques	Maître de Recherche
PIAU Monique	
PORTESEIL Jean-Louis	
THOLENCE Jean-Louis	
VERDILLON André	

CHERCHEURS du MINISTERE de la RECHERCHE et de la TECHNOLOGIE (Directeurs et Maîtres de Recherches, ENS Mines de St. Etienne)

LESBATS Pierre	Directeur de Recherche
BISCONDI Michel	Maître de Recherche
KOBYLANSKI André	Maître de Recherche
LE COZE Jean	Maître de Recherche
LALAUZE René	Maître de Recherche
LANCELOT Francis	Maître de Recherche
THEVENOT François	Maître de Recherche
TRAN MINH Canh	Maître de Recherche

PERSONNALITES HABILITEES à DIRIGER des TRAVAUX de RECHERCHE (Décision du Conseil Scientifique)

ALLIBERT Colette	E.N.S.E.E.G.
BERNARD Claude	E.N.S.E.E.G.
BONNET Rolland	E.N.S.E.E.G.
CAILLET Marcel	E.N.S.E.E.G.
CHATILLON Catherine	E.N.S.E.E.G.
CHATILLON Christian	E.N.S.E.E.G.
COULON Michel	E.N.S.E.E.G.
DIARD Jean-Paul	E.N.S.E.E.G.
EUSTAPOPOULOS Nicolas	E.N.S.E.E.G.
FOSTER Panayotis	E.N.S.E.E.G.

.../...

GALERIE Alain	E.N.S.E.E.G.
HAMMOU Abdelkader	E.N.S.E.E.G.
MALMEJAC Yves	E.N.S.E.E.G. (CENG)
MARTIN GARIN Régina	E.N.S.E.E.G.
NGUYEN TRUONG Bernadette	E.N.S.E.E.G.
RAVAINE Denis	E.N.S.E.E.G.
SAINFORT	E.N.S.E.E.G. (CENG)
SARRAZIN Pierre	E.N.S.E.E.G.
SIMON Jean-Paul	E.N.S.E.E.G.
TOUZAIN Philippe	E.N.S.E.E.G.
URBAIN Georges	E.N.S.E.E.G. (Laboratoire des ultra-réfractaires ODEILLON)
GUILHOT Bernard	E.N.S. Mines Saint Etienne
THOMAS Gérard	E.N.S. Mines Saint Etienne
DRIVER Julien	E.N.S. Mines Saint Etienne
BARIBAUD Michel	E.N.S.E.R.G.
BOREL Joseph	E.N.S.E.R.G.
CHOVET Alain	E.N.S.E.R.G.
CHEHIKIAN Alain	E.N.S.E.R.G.
DOLMAZON Jean-Marc	E.N.S.E.R.G.
HERAULT Jeanny	E.N.S.E.R.G.
MONLLOR Christian	E.N.S.E.R.G.
BORNARD Guy	E.N.S.I.E.G.
DESCHIZEAU Pierre	E.N.S.I.E.G.
GLANGEAUD François	E.N.S.I.E.G.
KOFMAN Walter	E.N.S.I.E.G.
LEJEUNE Gérard	E.N.S.I.E.G.
MAZUER Jean	E.N.S.I.E.G.
PERARD Jacques	E.N.S.I.E.G.
REINISCH Raymond	E.N.S.I.E.G.
ALEMANY Antoine	E.N.S.H.G.
BOIS Daniel	E.N.S.H.G.
DARVE Félix	E.N.S.H.G.
MICHEL Jean-Marie	E.N.S.H.G.
OBLÉD Charles	E.N.S.H.G.
ROWE Alain	E.N.S.H.G.
VAUCLIN Michel	E.N.S.H.G.
WACK Bernard	E.N.S.H.G.
BERT Didier	E.N.S.I.M.A.G.
CALMET Jacques	E.N.S.I.M.A.G.
COURTIN Jacques	E.N.S.I.M.A.G.
COURTOIS Bernard	E.N.S.I.M.A.G.
DELLA DORA Jean	E.N.S.I.M.A.G.
FONLUPT Jean	E.N.S.I.M.A.G.
SIFAKIS Joseph	E.N.S.I.M.A.G.
CHARUEL Robert	U.E.R.M.C.P.P.
CADET Jean	C.E.N.G.
COEURE Philippe	C.E.N.G. (LETI)

.../...

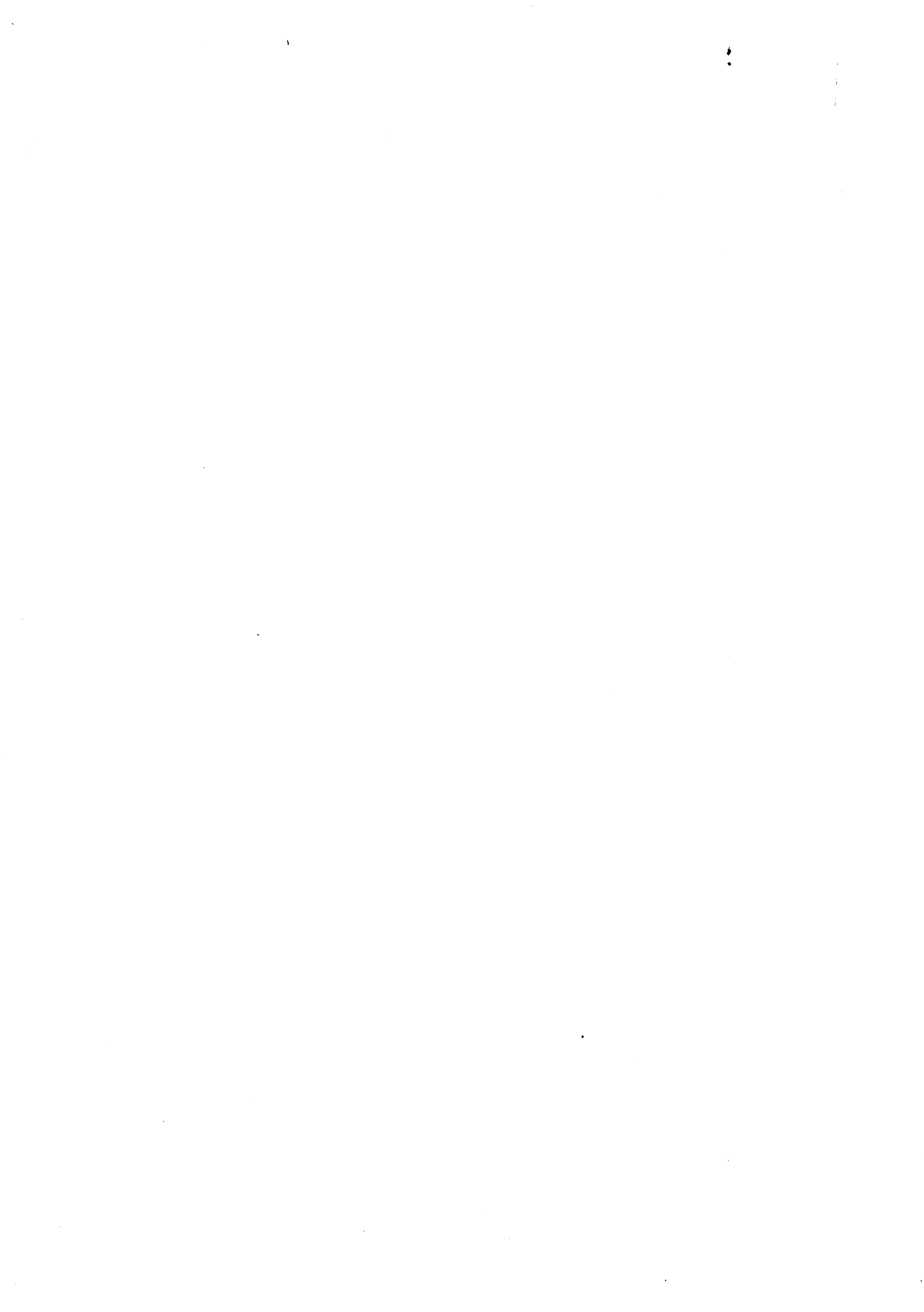
DELHAYE Jean-Marc
DUPUY Michel
JOUVE Hubert
NICOLAU Yvan
NIFENECKER Hervé
PERROUD Paul
PEUZIN Jean-Claude
TAIEB Maurice
VINCENDON Marc

C.E.N.G. (STT)
C.E.N.G. (LETI)
C.E.N.G. (LETI)
C.E.N.G. (LETI)
C.E.N.G.
C.E.N.G.
C.E.N.G. (LETI)
C.E.N.G.
C.E.N.G.

LABORATOIRES EXTERIEURS

DEMOULIN Eric
DEVINE
GERBER Roland
MERCKEL Gérard
PAULEAU Yves
GAUBERT C.

C.N.E.T.
C.N.E.T. (R.A.B.)
C.N.E.T.
C.N.E.T.
C.N.E.T.
I.N.S.A. Lyon



A mon mari,

à mes enfants,



Je tiens à remercier,

Monsieur Claude DELOBEL, Professeur à l'Université de Grenoble 1, qui a bien voulu me faire l'honneur de présider le jury de cette thèse,

Monsieur Louis BOLLIET, Professeur à l'Université de Grenoble 2, qui m'a accueilli dans son GIS de Mini et Micro Informatique, et a dirigé mon travail,

Monsieur Jacques COURTIN, Professeur à l'Université de Grenoble 2, qui a bien voulu accepter de faire partie du jury,

Monsieur Sylvestre LAVALLEE, de la Société ICARE, Chef du Projet CATHI, qui a guidé le projet et a bien voulu participer au jury,

Mesdemoiselles Maryse BAUDRAS et Isabelle ARNAUD, et Monsieur Yves MONTVIGNIER, qui ont pris une part active dans la réalisation du logiciel,

Madame Clotilde CHALAND, qui a assuré la dactylographie rapide et soignée du texte, malgré des contraintes de temps extrêmement serrées, me permettant ainsi de soutenir ma thèse,

Mesdames DURAND-EVRARD et DURAND et tout le Service de Documentation de la Mairie de Grenoble, qui ont constitué le thesaurus et assuré la saisie des fiches, je les remercie pour la démonstration le jour de la soutenance,

enfin tous ceux et celles qui à des titres divers ont participé au travail et m'ont apporté soutien et conseil.



Note explicative :

Société ICARE : filiale de la SCET (Société Centrale pour l'Equipement du Territoire) et des principales collectivités locales Rhône-Alpes (Courly, Ville de Lyon, Saint Etienne, Grenoble, Valence, Villeurbanne, Bourg en Bresse) et de l'Etablissement Public Régional Rhône-Alpes, c'est une Société d'Economie Mixte, qui intervient sur toutes les applications de l'informatique dans les municipalités.

Projet CATHI : Consultation Automatisée d'un THésaurus et Information documentaire.



S O M M A I R E

Introduction - Contexte général	p. 0
1. Evolution des systèmes documentaires	p. 1
1.1. Les anciennes classifications	p. 1
1.2. Définition générale d'un thésaurus	p. 1
1.3. Opposition à l'analyse du langage naturel	p. 3
2. Progrès technologiques et conséquences sur les systèmes documentaires	p. 4
Chapitre 1 - Présentation du projet	p. 7
1. Historique du projet	p. 8
2. Objectifs initiaux du projet	p. 9
3. La base de données	p. 11
4. Fonctions et caractéristiques générales du projet CATHI	p. 12
4.1. Fonctions du logiciel	p. 12
4.2. Caractéristiques générales du logiciel	p. 16
5. Mise en oeuvre, expérimentation et évolution future	p. 17
Chapitre 2 - Le fonds documentaire	p. 19
1. Les sources	p. 20
1.1. Sélection des sources	p. 20
1.2. Organisation des sources	p. 20
1.3. Les textes pris en compte	p. 21
2. Description de la fiche documentaire	p. 22
2.1. Numéro de fiche	p. 22
2.2. Le code bordereau	p. 22
2.3. Les références	p. 22
2.4. Analyse du texte	p. 23
2.5. Descripteurs du texte	p. 24

Chapitre 3 - L'outil documentaire : le thesaurus	p. 29
1. Description d'une fiche thesaurus	p. 30
1.1. Premier champ : le mot-clé de référence	p. 30
1.2. Deuxième champ : le numéro d'ordre	p. 30
1.3. Troisième champ : le type	p. 31
1.4. Quatrième champ : la note d'application	p. 31
1.5. Les autres champs : les mots-clés liés	p. 31
2. Structure d'ensemble du thesaurus	p. 34
3. Les synonymes	p. 36
3.1. Première caractéristique	p. 36
3.2. Seconde caractéristique	p. 36
4. Les non-synonymes ou mots-clés principaux	p. 37
4.1. La relation de hiérarchie	p. 37
4.2. La relation d'analogie	p. 38
5. Problèmes posés par la suppression des mots-clés	p. 38
6. Ensemble de contrôles effectués par le système	p. 39
6.1. Contrôles concernant les descripteurs	p. 40
6.2. Existence des mots-clés liés	p. 41
6.3. Contrôles concernant les synonymes	p. 42
6.4. Inexistence du mot-clé principal associé à un mot-clé synonyme	p. 42
7. Anomalies et contrôles de cohérence du thesaurus	p. 43
Chapitre 4 - Description technique de la base de données	p. 51
1. Les fichiers relatifs au thesaurus	p. 53
1.1. Explications sur l'organisation des fichiers du thesaurus	p. 54
1.2. Fichiers de données concernant le thesaurus	p. 57
1.3. Les fichiers de travail concernant le thesaurus	p. 66
1.4. Fichier ANOMALIE	p. 71
2. Les fichiers relatifs au fonds documentaire	p. 73
2.1. Les fichiers de données	p. 74
2.2. Les fichiers de travail	p. 82
3. Fichier INVERSE	p. 84
4. Volume occupé par les fichiers pendant la phase expérimentale	p. 86

Chapitre 5 - Les fonctions du logiciel	p. 89
1. Remarques générales sur les programmes	p. 91
1.1. Les sous-programmes communs	p. 91
1.2. Les commandes	p. 92
1.3. Les écrans	p. 94
2. Etude des principaux programmes	p. 99
2.1. Les programmes de création de la base documentaire	p. 101
2.2. Contrôle et réorganisation du thesaurus	p. 102
2.3. Gestion du thesaurus : GESTDICO	p. 103
2.4. Gestion du fonds documentaire : INTEX, GESTEX1, GESTEX2	p. 112
2.5. Consultation de la base	p. 119
 Chapitre 6 - Etude comparative d'autres logiciels du même type	 p. 141
 Chapitre 7 - Critiques et conclusions	 p. 155
1. Champ d'application du logiciel	p. 156
2. Taille des programmes	p. 158
3. Au niveau de la consultation de la base	p. 159
4. D'un point de vue technique	p. 159
5. Conclusion	p. 160
 Bibliographie	 p. 161
 Annexes	 p. 165



INTRODUCTION

CONTEXTE GÉNÉRAL



1. ÉVOLUTION DES SYSTÈMES DOCUMENTAIRES

1.1. Les anciennes classifications

Les anciennes classifications hiérarchiques et thématiques se sont révélées fort contraignantes dans le contexte de "l'explosion des connaissances" des dernières années, leur principal défaut étant de prétendre donner une structure rigide à l'ensemble du savoir. Par ailleurs, "l'interdisciplinarité" a remis en cause la validité des "classifications a priori". C'est pourquoi ces dernières années ont vu l'apparition de "thesauri" de plus en plus nombreux.

1.2. Définition générale d'un thesaurus

Un thesaurus est une liste de termes significatifs dans laquelle certains mots, les *descripteurs*, sont choisis de préférence à d'autres pour caractériser (traduire) le contenu d'un document.

Les autres termes, qui ne sont pas retenus pour indexer les documents, les *synonymes*, renvoient au descripteur choisi.

Les descripteurs ont entre eux des relations sémantiques. C'est pourquoi on peut définir un *thesaurus* comme étant un *vocabulaire contrôlé et structuré*.

La recherche documentaire est le processus inverse de celui de l'indexation : en utilisant les mots-clés et les relations établies entre eux, l'utilisateur cherche à remonter jusqu'aux documents.

Le vocabulaire d'un thesaurus doit être :

- . non ambigu,
- . évolutif : en particulier, il doit permettre d'incorporer de nouvelles notions, se perfectionner avec les interrogations et de ne rien perdre de son adéquation par rapport aux informations analysées.

La plupart des thesauri possèdent les trois types de relations suivantes :

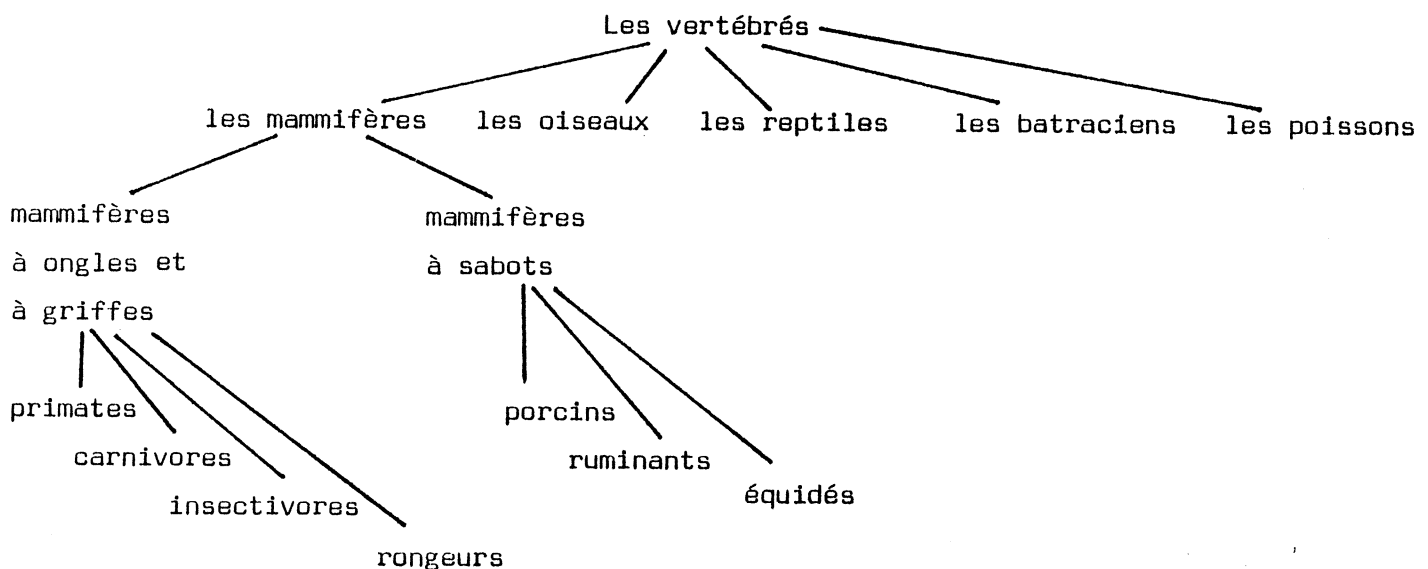
1.2.1. Relation d'équivalence

On regroupe dans une même classe d'équivalence les termes qui expriment une même notion. Un seul terme sera choisi pour représenter tous les autres et indexer les documents ; les autres seront ses synonymes.

1.2.2. Relations de hiérarchie

Les relations de hiérarchie correspondent aux classifications hiérarchiques traditionnelles. Les descripteurs du thesaurus sont reliés entre eux par des relations verticales (ascendantes et descendantes) qui permettent de représenter le thesaurus sous une forme arborescente : quand on descend dans l'arbre, on passe de notions générales à des notions plus spécifiques.

Exemple d'arbre hiérarchique :



1.2.3. Les relations d'analogie ou de voisinage

Elles mettent en évidence la signification voisine entre deux descripteurs sans que ceux-ci puissent être déclarés équivalents. Ces relations sont les plus difficiles à déterminer ; elles ont pour but d'indiquer à un utilisateur les termes voisins à consulter en élargissant son domaine d'investigation, si les réponses obtenues ne l'ont pas satisfait.

Les relations d'analogie sont des relations horizontales (par opposition aux relations verticales de hiérarchie).

Par rapport aux systèmes hiérarchisés "à inclusion", chaque classe englobant la suivante, les systèmes avec thesaurus appartiennent aux *systèmes combinatoires*.

1.3. Opposition à l'analyse du langage naturel

Les systèmes orientés thesaurus s'opposent à ceux qui reposent sur une analyse du langage naturel (par exemple celui des titres, des résumés ou des textes) qui est un langage libre. L'interrogation de bases de données en langue naturelle constitue un axe très prometteur pour l'avenir, mais cela exige des logiciels complexes et des moyens informatiques plus puissants, bien que des expériences dans ce domaine soient faites sur des mini ordinateurs (cf. PIAF).

2. PROGRÈS TECHNOLOGIQUES ET CONSÉQUENCES SUR LES SYSTÈMES DOCUMENTAIRES

Les progrès technologiques considérables dans les domaines de l'informatique et des télécommunications, ont permis de franchir un pas supplémentaire dans la recherche d'informations.

En effet, face à la masse toujours croissante de documents, les systèmes manuels se trouvent de moins en moins performants et seule l'informatique par les capacités qu'elle offre

- . de mémorisation,
- . de traitement d'un grand nombre de données,
- . et d'accès rapide aux enregistrements,

apporte une solution satisfaisante.

Le développement de réseaux nationaux et internationaux de transmission de données, permet de consulter des fichiers très éloignés en temps réel. Cependant, bien des progrès restent à faire, en particulier en ce qui concerne les langages d'interrogation qui ne sont pas normalisés et nécessitent sinon l'intermédiaire de spécialistes, du moins un apprentissage plus ou moins long.

Logiciels documentaires

Il existe à l'heure actuelle une multitude de logiciels documentaires que nous pouvons schématiquement classer en deux grandes catégories :

1. Les logiciels "à menu"

Ces logiciels sont basés sur une recherche arborescente : l'utilisateur est entièrement guidé par une succession de "menus" qui s'affichent, chaque page du menu présentant différentes options. Le choix de l'une d'entre elles provoque l'affichage d'un menu "plus spécialisé". Ce procédé a le mérite d'être simple d'utilisation, mais l'inconvénient d'être long et fastidieux, l'utilisateur ne pouvant échapper au cheminement logique et prédéterminé qui lui est imposé. En outre, ces systèmes ne peuvent s'adapter qu'à des documents présentant une structure hiérarchisée "à inclusion".

2. Les logiciels documentaires proprement dits

Il en existe une grande variété, qui reposent tous sur l'existence d'un dictionnaire et sur la notion de fichier inverse. Le dictionnaire peut être créé a priori ou au moment de la saisie des documents. L'indexage des documents est fait soit automatiquement, soit par l'utilisateur. Chaque terme "descripteur" a un enregistrement dans le fichier inverse, contenant la liste des adresses des documents sur lesquels il renvoie.

On comprend qu'une telle organisation optimise la recherche. La combinaison des termes par les opérateurs ET, OU, SAUF, permettra d'obtenir plus rapidement le(s) document(s) désiré(s).

Ces logiciels sont très performants au niveau d'une recherche conversationnelle ; par contre, la mise à jour en temps réel des fichiers est plus difficile.

Parmi ces logiciels, on distingue :

2.1. Les logiciels basés sur l'analyse du texte intégral

La recherche s'effectue sur tous les mots du texte, à l'exception des mots vides (articles, par exemple un, des, etc ..). D'une grande souplesse d'utilisation, ils supposent malgré tout que l'utilisateur connaisse bien le domaine concerné, puisque le vocabulaire qu'il emploie n'est pas contrôlé ; par ailleurs, ces logiciels exigent des moyens de mémorisation importants.

2.2. Les logiciels basés sur un système d'analyse syntaxique et statistique

En pleine évolution, leur principe est de permettre une recherche en langage naturel par des personnes ne connaissant pas a priori la base. Au moment de la saisie d'un document, le titre ou le texte font l'objet d'une analyse syntaxique : élimination des mots vides, détection des fautes d'orthographe, analyse grammaticale pour lever les ambiguïtés, recherche d'une forme normalisée (exemple : masculin, singulier pour adjectifs, infinitif pour verbes, etc ..). Par ailleurs, toujours au moment de la saisie, les termes retenus

font l'objet d'une analyse statistique pour déterminer leur fréquence d'utilisation. Lors d'une interrogation, chaque question est analysée pour déterminer les mots forts (c'est-à-dire les moins utilisés lors de l'indexation de tous les documents) ; ceux-ci sont en effet supposés caractériser le mieux le document, en étant les plus sélectifs.

Testés à l'heure actuelle, ces logiciels nécessitent des moyens informatiques relativement puissants.

2.3. Les logiciels utilisant un thesaurus

Par rapport aux systèmes documentaires précédents, ils présentent une solution intermédiaire, entre les "logiciels à menu" et les logiciels utilisant un vocabulaire non contrôlé (langage naturel).

Ces logiciels orientés thesaurus sont moins rigides et contraignants que les premiers, mais plus que les seconds. Par contre, moins gourmands en moyens informatiques que les seconds, ils commencent à se développer sur des systèmes micro-informatiques. CATHI en est un exemple. Nous verrons par la suite, à propos de cette application, comment nous avons essayé de limiter les inconvénients d'un vocabulaire contrôlé et structuré d'une part, et d'autre part comment nous avons essayé de rendre les procédures d'accès à la base suffisamment simples pour être accessibles à des non-spécialistes.

CHAPITRE 1

PRÉSENTATION DU PROJET

1. HISTORIQUE DU PROJET

Créé en 1960, le Service de Documentation de la Mairie de Grenoble, adopta, pour l'ensemble de ses documents constitués d'informations concernant les communes, une classification alphanumérique.

En 1978, ce fonds documentaire, unique en France, comportait 35 000 fiches. C'est alors que, rattaché au Conservateur des Archives, le Service de Documentation abandonné l'ancienne classification et adopta le système d'indexation des documents par mots clés. Un thesaurus qui fait l'originalité de la base fut élaboré par Madame DURAND, documentaliste. Ce thesaurus couvre maintenant tous les domaines d'intervention des communes : urbanisme, construction, organisation administrative, finances, pouvoirs du maire, marchés publics....

Suite à la publication du rapport LENOIR PROT qui mettait à jour la faiblesse du marché de l'information en France, le Ministère de l'Industrie décida en 1977 de créer une véritable industrie des bases de données, pour répondre aux offensives américaines, et notamment assurer à la France une indépendance en matière d'information scientifique et technique. C'est ainsi que la MIDIST lança un programme prévoyant la création de 50 bases de données en 1985.

En 1980, répondant à cet appel d'offre et connaissant l'intention de la Mairie de Grenoble d'informatiser la réglementation du personnel communal, Monsieur LAVALLE, chef de projet à la Société ICARE, élaborera un projet. La MIDIST donna très vite son accord, à condition que le projet ne soit pas uniquement grenoblois. Le projet fut donc reformulé à partir d'un groupe régional Grenoble - Lyon - La Courly (communauté urbaine de l'agglomération lyonnaise + St Etienne, Valence, Grenoble, Bourg en Bresse, Villeurbanne).

Les cahiers des charges furent élaborés avec la collaboration du groupe régional et les services de documentation des municipalités concernées. Enquêtes et questionnaires permirent de redéfinir les besoins, tout en préparant le personnel des services à l'introduction de l'informatique.

On décida alors de réaliser le projet sur un micro-ordinateur : le GOUPIL 2 avec 64 K de mémoire, une unité de disques durs : CII 4B de deux fois 10 millions d'octets dont 10 amovibles et une imprimante Honeywell de 100 cps.

Deux partenaires locaux ont été demandés pour participer à la réalisation du logiciel, à savoir la Société CEDIS, société lyonnaise de distribution du matériel GOUPIL, et le GIS de MINI et MICRO INFORMATIQUE du CNRS.

N.B. Le projet d'informatisation ne concerne que le domaine du personnel communal.

2. OBJECTIFS INITIAUX DU PROJET

Les concepteurs du projet visaient un certain nombre d'objectifs. Il s'agissait :

- . *De concevoir un outil professionnel* pour les documentalistes, de façon à rendre plus fiable l'information. En effet, la maîtrise du suivi de la législation (abrogations, modifications, compléments de textes) est difficile à assurer avec les systèmes manuels traditionnels.
- . *De concevoir un outil télématique* permettant l'accès rapide à l'information, non seulement pour les documentalistes, mais aussi pour les élus locaux et les services du personnel. Ceux-ci interrogeront la base avec des terminaux Videotex (MINITEL). Il fallait donc que le langage d'interrogation soit accessible à des non-spécialistes : non-informaticiens, non-documentalistes.
- . *De favoriser une plus grande coopération intercommunale* : le traitement des sources (textes législatifs et réglementaires, revues d'administrations locales) analysées dans tous les services, est une charge de plus en plus lourde, compte tenu de l'accroissement des informations.
Exemple : 4 796 pages au Journal Officiel en 1960
15 256 pages au Journal Officiel en 1980.

C'est pourquoi on a envisagé une répartition des tâches entre les villes, chaque ville se spécialisant dans un domaine précis. Les villes qui ne possèdent pas de service de documentation, peuvent avec un GOUPIL ou un terminal, accéder à la documentation des autres villes.

Enfin, une des objectifs prioritaires était :

- . *de minimiser les demandes d'édition systématiques et l'accumulation de documents* que l'on conserve pour un besoin éventuel sans avoir toujours vérifié l'exactitude de l'information. Cette façon de procéder se révèle lourde et coûteuse dans la pratique : ainsi chaque service se constitue une documentation parallèle à base de photocopies.

Si les deux premiers objectifs ont été ou seront atteints sans trop de difficultés, les deux derniers constituent un pari : en effet, ces objectifs impliquent une nouvelle organisation dans le travail et une nouvelle conception de la recherche d'information, et on doit tenir compte des réticences psychologiques et sociologiques.

Nous allons donner un aperçu général de la base de données et du logiciel, dont nous reparlerons avec plus de détails dans la suite de ce document.

3. LA BASE DE DONNÉES

La base de données est composée :

- . *d'un fonds documentaire* : ensemble de fiches documentaires. Chaque fiche documentaire contient :

- les *références* qui permettent de retrouver le texte

Ainsi a-t-on bien affaire à une base de données et non pas à une banque de données.

Par abus de langage, dans la suite il nous arrivera de parler de "texte" au lieu de "fiche documentaire".

- des *descripteurs* qui sont les mots-clés choisis dans le lexique pour indexer le texte. Indexer un texte signifie choisir un certain nombre de mots-clés dans le lexique pour définir et caractériser le texte.

- . *D'un thesaurus*, ensemble de mots clés (le *lexique*) et de *relations sémantiques* entre les mots-clés. Les mots-clés peuvent être classés en deux groupes disjoints :

- les *descripteurs* : ils correspondent aux termes employés par les documentalistes et ont entre eux des relations de hiérarchie ascendante et descendante, et d'analogie ;

- les *synonymes* : ce sont les mots-clés employés ou susceptibles d'être employés à la place des descripteurs, par les non-spécialistes, lors de l'interrogation de la base. Ils ne peuvent indexer les textes, mais grâce à la relation existant entre chaque synonyme et le descripteur qui lui est associé, l'utilisateur pourra prendre connaissance de toutes les fiches documentaires indexées par ce descripteur.

4. FONCTIONS ET CARACTÉRISTIQUES GÉNÉRALES DU PROJET CATHI*

4.1. Fonctions du logiciel

CATHI se compose de deux modules :

- . le premier comprend les programmes qui ne serviront qu'une fois ou épisodiquement,
- . le second comprend l'ensemble des programmes utilisés en régime de croisière.

4.1.1. Premier module

a. Programme de création

Ce programme permet la *saisie des fiches thesaurus* en mode conversationnel et leur consultation afin que les documentalistes puissent vérifier le travail effectué.

b. Autre programme de création

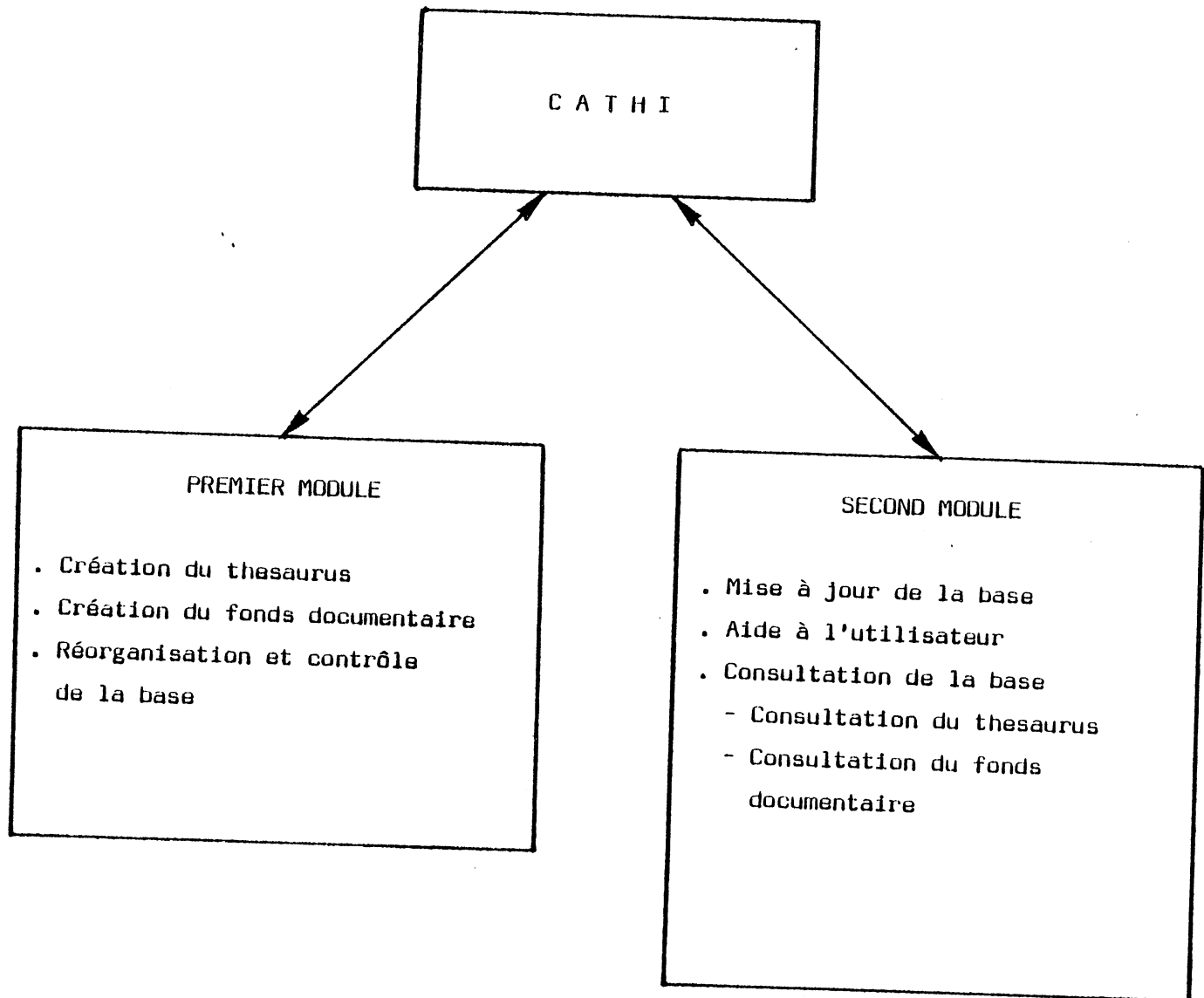
Ce programme permet la *saisie conversationnelle des fiches documentaires* et leur visualisation.

Ces deux programmes n'auront servi qu'une fois. Ce sont les documentalistes qui ont assuré le chargement des fiches thesaurus et documentaires.

c. Le troisième programme

Il sera exécuté à la demande des documentalistes. Sa fonction est d'assurer un certain nombre de contrôles de cohérence et de réorganiser le fichier contenant le lexique des mots-clés, de manière à assurer un accès rapide aux enregistrements des mots-clés et par là de permettre la consultation et la mise à jour de la base en temps réel. A l'heure actuelle, les temps de réponse sont inférieurs à 2 secondes. Ce programme de réorganisation et de contrôle peut être lancé directement ou à partir de la page d'accueil du deuxième module, en tapant un numéro spécial.

* Consultation Automatisée d'un THésaurus et Information documentaire



Présentation générale de CATHI

4.1.2. Second module

a. Les programmes de mise à jour de la base

(gestion du thesaurus et du fonds documentaire).

b. Un programme d'auto-formation

destiné à fournir des explications à l'utilisateur sur le fonctionnement du système. Ces explications constitueront une page écran. Ce programme n'est pas réalisé actuellement.

c. Le programme de consultation de la base

. *Consultation du thesaurus* (lexique et relations sémantiques entre les mots-clés).

. *Consultation du fonds documentaire* : celle-ci comporte une logique de recherche peu sophistiquée : une question se compose de plusieurs mots-clés tapés par l'utilisateur, ceux-ci sont reliés par l'opérateur logique ET. Une première réponse indique le nombre de textes sélectionnés. L'utilisateur peut alors diminuer la cible obtenue,

- soit en utilisant l'opérateur SAUF sur un seul mot-clé,
- soit grâce à des filtres sur la date et(ou) sur la nature juridique.

L'utilisateur peut également utiliser les filtres directement sans passer par la phase "recherche à partir des mots-clés", mais pour le moment cette recherche directe sur la date et(ou) sur la nature juridique est séquentielle.

En général, pour sélectionner un texte, 3 à 4 mots-clés sont utilisés, mais la question peut en comprendre 6 ; aucune contrainte technique ne nous empêche d'augmenter ce nombre, mais l'expérience montre que c'est inutile.

Les programmes du second module s'enchaînent les uns aux autres par l'intermédiaire d'une page d'accueil (voir schéma). Les programmes de réorganisation et de mise à jour ne sont accessibles qu'après la frappe d'un mot de passe.

4.2. Caractéristiques générales du logiciel

L'objectif du projet n'était pas de concevoir un logiciel sophistiqué, mais plutôt de concevoir un outil professionnel adapté à une organisation de la documentation pré-existante et destiné à la fois à un public de documentalistes et de non spécialistes (non informaticiens, non documentalistes). C'est pour cela que l'on a voulu au maximum supprimer les barrières techniques et simplifier les procédures.

Le choix d'un micro-ordinateur, interdisait par ailleurs de prendre en compte tous les besoins d'un service de documentation et il fallait laisser aux documentalistes la maîtrise de leur travail : en aucun cas le système ne devait les supplanter dans leurs tâches les plus intéressantes : c'est pourquoi, le système ne se charge pas de vérifier la cohérence totale du thesaurus.

La mise à jour des fiches thesaurus et documentaires se fait en temps réel et est grandement facilitée grâce à un gestionnaire d'écran assez performant. Par certains aspects, le logiciel se rapproche d'un logiciel de traitement de texte.

La mise à jour et la consultation de la base se font en mode conversationnel : le langage interactif utilisé est le même que dans le projet CLAIRE^{*}. Un des objectifs d'un fonds documentaire est de pouvoir utiliser CATHI, pour interroger aussi bien une base de données locale (celle de CLAIRE par exemple), que celles résidant sur un gros serveur éloigné ; CATHI serait utilisé pour les besoins professionnels et CLAIRE pour les besoins "grand public". Encore faut-il que le "mode d'emploi" reste invariant pour ne pas perturber l'utilisateur.

* *La ville de Grenoble, soucieuse de développer la communication entre la Mairie, les administrations, les différentes associations locales et les Grenoblois, a confié en 1980 à un groupe de travail (ville de Grenoble et CEESI), la mission d'élaborer un projet de renseignements et d'informations assisté par la télématique.*

CLAIRE = Catalogue Local Avis Informations Renseignements Electronique.

Le dialogue utilisateur - machine se fait grâce à un certain nombre de commandes que l'on retrouve dans tous les programmes ; ces commandes correspondent aux touches de fonction du clavier Videotex.

Autre originalité du projet et autre point commun avec CLAIRE : le système d'interrogation développé n'est pas un système par arborescence, utilisé dans la plupart des systèmes Videotex : dans les projets CLAIRE et CATHI, au lieu d'être guidé totalement par le système, l'utilisateur procède par tâtonnements en utilisant les mots-clés du thesaurus et les relations entre ces mots-clés.

5. MISE EN OEUVRE, EXPÉRIMENTATION ET ÉVOLUTION FUTURE DU PROJET

La première phase a duré six mois ; elle comprenait la mise au point du logiciel et la saisie des bordereaux, textes et thesaurus ; elle a donné lieu à la mise en place d'une maquette d'expérimentation au Service de Documentation de la Mairie de Grenoble.

Cette première expérience s'est révélée positive puisque les fiches papier ont été complètement abandonnées et que de nombreux utilisateurs (documentalistes ou non) viennent interroger la base et semblent avoir bien adopté le système. Cependant, beaucoup d'habitudes se trouvent bouleversées ; en particulier les Services du Personnel doivent passer par le système et cela a pour conséquence une nouvelle répartition des tâches entre le Service de Documentation et les Services du Personnel. Pour le moment, CATHI fonctionne sur 360 mots-clés et 1 000 textes ; à terme, la base devrait contenir 1 500 mots-clés et 5 ou 6 000 textes.

Au cours d'une deuxième phase il est prévu de grossir le logiciel existant d'un certain nombre de programmes :

- . programme de recherche par mots-clés auquel on ajoutera un programme de recherche par thème, de type arborescent ;
- . des programmes d'édition seront faits : édition par mot-clé, par thème, suivant la nature juridique, la date, ..., ainsi que des programmes d'édition statistique sur la fréquence d'utilisation des mots-clés lors de l'interrogation de la base ;

- . mise au point d'un certain nombre de techniques de visualisation, telles que par exemple l'utilisation de fenêtre permettant de faire ressortir une partie de l'information et de l'agrandir, ainsi qu'un défilement avant arrière de l'écran ;
- . surtout, au terme d'un certain temps de fonctionnement, on passera à un système en grandeur nature, en testant le système dans différentes villes de la région Rhône Alpes. A cette occasion, une répartition du travail dans la production de la documentation pourrait être mise en place entre les villes qui ont été associées au projet. Chaque ville se spécialisant dans un domaine particulier.

A plus long terme, CATHI pourrait participer au projet national d'une banque de données inter-communales, élaboré par la ville de Metz, le Ministère des PTT et la SCET. Si cela aboutissait, CATHI serait un précurseur, car par la simplicité de ses procédures d'accès et de son langage d'interrogation, l'utilisateur peut se passer de l'intermédiaire de spécialistes (informaticiens, documentalistes), ce qui n'est pas encore le cas pour les grandes banques de données nationales.

Cette extension Videotex multiposte permettra la communication entre

- . services de documentation et services du personnel,
- . différentes villes de la région Rhône Alpes,

et l'accès

- . à des bases de données locales (celle de CLAIRE par exemple),
- . puis nationales,

ce qui demeure l'objectif prioritaire du projet.

CHAPITRE 2

LE FONDS DOCUMENTAIRE

1. LES SOURCES

1.1. Sélection des sources

Seules sont prises en compte les sources juridiques externes concernant les activités communales, c'est-à-dire les publications officielles et les revues d'information communale (50 périodiques).

Le traitement des documents élaborés par les services municipaux (délibérations, arrêtés, notes et études), n'est pas pris en compte dans le cadre de l'expérience.

1.2. Organisation des sources

Les sources ont été réparties en douze domaines qui recouvrent soit l'activité des services municipaux, soit les grands domaines d'intervention des collectivités locales. La nature des sources est la même pour tous les domaines :

- . législation : journaux officiels, bulletins des Ministères,
- . jurisprudence : décisions du Conseil Constitutionnel, du Conseil d'Etat et des Tribunaux Administratifs, publiées par les revues communales,
- . doctrine : questions écrites et orales, articles de revues.

Le projet soumis à la MIDIST concernant seulement le domaine du *personnel travaillant pour les communes*.

Ce choix s'explique par les raisons suivantes :

- . il correspond aux recherches les plus nombreuses,
- . il correspond à une clientèle très élargie (services du personnel, agents municipaux, mairies voisines, ...),
- . les textes sont appliqués dès leur parution,
- . la mise à jour des informations est complexe.

1.3. Les textes pris en compte

Les textes pris en compte sont donc :

- . les dispositions réglementaires concernant le personnel communal, et aussi le personnel d'état quand elles sont susceptibles d'être appliquées aux agents communaux,
- . les textes concernant le personnel travaillant pour les communes : vacataires, collaborateurs occasionnels,
- . les mesures sociales concernant les fonctionnaires et auxiliaires,
- . les dispositions concernant le personnel culturel des communes (employés des musées, bibliothèques, archives).

Le thème personnel comprend :

- . les textes de base, les études et jurisprudences de 1962 à 1978,
- . tous les textes parus depuis 1978, soit environ 12 000 fiches documentaires.

CATHI fonctionne pour le moment avec 300 mots-clés et 1 000 textes. A terme, la base comprendra 1 500 mots-clés et 5 à 6 000 textes.

Des extensions sont prévues qui toucheront : les finances communales, les marchés publics et surtout les documents élaborés par les communes (délibérations, arrêtés, notes et études).

Chaque texte constitue une unité documentaire et fait l'objet d'une fiche documentaire.

2. DESCRIPTION DE LA FICHE DOCUMENTAIRE

La fiche documentaire est composée de deux bordereaux :

. Le *premier bordereau* concerne le *texte* lui-même :

- le numéro de fiche
- le code bordereau = 1
- les références : référence du support
référence du texte
- l'analyse de texte.

. Le *second bordereau* concerne les *descripteurs* du texte :

- le numéro de fiche
- le code bordereau = 2
- les descripteurs du texte qui sont les mots-clés indexant le texte.

2.1. Numéro de fiche

ce numéro sert d'identificateur pour la fiche ; il est évidemment le même pour les deux bordereaux. Il correspond au numéro d'ordre du texte dans le fichier et occupe quatre caractères.

2.2. Le code bordereau sert à distinguer le premier bordereau du second ; il occupe un caractère.

2.3. Les références

2.3.1. Références du support

.. *Titre de la publication* : c'est un code mémotech qui occupe deux caractères ; c'est une zone obligatoire. (Cf. liste des code en annexe).

.. *Date du support* : c'est la date de parution de la publication ; c'est une zone obligatoire à la saisie ; elle comprend le jour, le mois et l'année et occupe 3 x 2 caractères.

- .. *Numéro du support* : il occupe quatre caractères au maximum ; ce n'est pas une zone obligatoire.
- .. *Numéro de la page dans le support* : il occupe au maximum cinq caractères ; c'est une zone obligatoire.

2.3.2. Références du texte

- .. *Appellation* : code mnémotechnique et alphanumérique (cf. liste des codes en annexe). C'est une zone obligatoire qui occupe deux caractères.
- .. *Date du texte* : correspond à la date de ratification ; elle comprend le jour, la date, le mois et occupe 3 x 2 caractères ; ce n'est pas une zone obligatoire.
- .. *Numéro du texte dans le cas de la législation* : occupe six caractères au maximum ; ce n'est pas une zone obligatoire.
- .. *Source du droit* : zone obligatoire codée sur un caractère :
 1. jurisprudence,
 2. commentaire,
 3. législation.

2.4. Analyse du texte

C'est une zone libre comprenant :

- le résumé du texte,

- les éléments de modification :

- .. par exemple, les références des textes abrogés par celui qu'on examine ;
- .. les références des textes abrogeants,
- .. les références diverses (textes modificatifs ou rectificatifs).

Les relations entre les textes (abrogés ou abrogeants, modifiés ou modificatifs), seront faites par l'utilisateur au vu de l'analyse. Il n'y a pas d'automatisme. Cette zone est organisée par les documentalistes et comprend 14 x 38 caractères. Pour la législation, on réécrit les références du texte par précaution en cas d'erreur de frappe sur les chiffres.

2.5. Descripteurs du texte

au nombre maximum de 14, ils occupent au maximum 14 x 38 caractères. Ce sont les mots-clés indexant le texte. Pour que ceux-ci soient acceptés comme descripteurs,

- ils doivent exister dans le thesaurus, c'est-à-dire que leur fiche thesaurus a dû être préalablement saisie ;
- ils ne doivent pas avoir été mis hors service, ni être des synonymes (voir description du thesaurus).

Ces descripteurs constituent le champ de recherche lors d'une interrogation de la base. Le choix des descripteurs est fait par les documentalistes. Certains descripteurs sont généraux pour éviter les silences lors d'une interrogation, d'autres sont précis en vue d'une recherche ponctuelle. Les deux types de descripteurs sont nécessaires. L'expérience dans ce domaine tend à prouver que l'utilisateur préfère un trop grand nombre de réponses à un silence complet (aucun texte sélectionné). Certains descripteurs peuvent renvoyer sur une centaine de textes.

BORDEREAU TEXTE

NUMERO DE FICHE

3 7 5
1

CODE BORDEREAU

1
5

REFERENCE DU SUPPORT :

SOURCE DU DROIT : 3

(1 : légis., 2 : jurisp., 3 : commentaire)

TITRE A N DATE J J M M A A
8 10 2 2 0 1 8 0N° de support
16Page 2 0 6
20 24

REFERENCE DU TEXTE

APPELLATION Q E DATE J J M M A A
25 27 0 5 1 2 7 9N° du TEXTE 2 3 3 2 9
33 38

ANALYSE DU TEXTE :

QUESTION ECRITE 2 3 3 2 9 , 5 DECEMBRE 1 9 7 9 44

LE VERSEMENT DE L'INDEMNITE POUR LANGU 44

E ETRANGERE EST SUBORDONNE AU PASSAGE 44

D'UN EXAMEN DANS LA LANGUE PARLEE 44

44

44

44

44

44

44

44

44

44

(le format du texte est libre : il apparaîtra à l'écran avec la même présentation)

BORDEREAU : MOTS CLES (INDEXANT LE TEXTE)

NUMERO DE FICHE

1 3 7 5 4

Code Bordereau

2 5

Date de création

MOTS - CLES

I N D E M N I T E

6 4

6 4

6 4

6 4

6 4

6 4

6 4

6 4

6 4

6 4

6 4

6 4

6 4

6 4

(Ecrire un mot-clé par ligne en commençant en col. 6)

I C A R E

Documentation communale

BORDEREAU TEXTE

NUMERO DE FICHE 0 0 0 6
1

CODE BORDEREAU 1 1
5

REFERENCE DU SUPPORT :

SOURCE DU DROIT : 1
7

(1 : légist., 2 : jurisp., 3 : commentaire)

TITRE 0 2 DATE 1 6 0 2 8 0
8 10

N° de support 0 0 0 0 Page 1 7 5 8
16 20 24

REFERENCE DU TEXTE

APPELLATION 0 5 DATE 2 8 0 1 8 0
25 27

N° du TEXTE 0 0 0 0 0 0
33 38

ANALYSE DU TEXTE :

D I P L O M E S R E Q U I S P O U R C O N C O U R S S U R T I T R
44

E S
44

M O D I F I E A R R E T E 2 6 S E P T E M B R E 1 9 7 3 A N N E X
44

E I
44

44

44

44

44

44

44

44

44

44

44

44

44

44

44

44

44

44

(le format du texte est libre : il apparaîtra à l'écran avec la même présentation)

BORDEREAU : MOTS CLES (INDEXANT LE TEXTE)

NUMERO DE FICHE

0	0	0	6
1			4

Code Bordereau

2
5

Date de création

MOTS - CLES

R E C R U T E M E N T

6

P E R S O N N E L C O M M U N A L

6

A D J O I N T T E C H N I Q U E

6

P E R S O N N E L D E S E T A B L I S S E M E N T S P U B L I C S C

6

O M M U N A U X

6

6

6

6

6

6

6

6

6

6

(Ecrire un mot-clé par ligne en commençant en col. 6)

CHAPITRE 3

L'OUTIL DOCUMENTAIRE : LE THESAURUS

Le thesaurus est l'ensemble des mots-clés et de leurs relations sémantiques

- . synonymie,
- . analogie,
- . hiérarchie.

Le thesaurus est élaboré par les documentalistes qui choisissent les mots-clés et définissent les relations.

A chaque mot-clé correspond une fiche thesaurus.

1. DESCRIPTION D'UNE FICHE THESAURUS

Elle comprend 19 lignes de 40 caractères et est définie par 20 champs.

1.1. Premier champ : le mot-clé de référence

Sur 38 caractères, il identifie la fiche. On accède à une fiche thesaurus en tapant le mot-clé de référence.

1.2. Deuxième champ : le numéro d'ordre

Il correspond au numéro d'ordre du mot-clé dans le fichier trié. Le numéro d'ordre d'une fiche thesaurus peut donc varier, mais il est géré par le système et ne peut être modifié par l'utilisateur au cours d'une transaction écran-clavier. Ce numéro d'ordre a un rôle sur le plan informatique puisqu'il est inscrit dans tous les enregistrements correspondant au mot-clé. Il occupe quatre caractères. On n'accède jamais à une fiche thesaurus en tapant le numéro d'ordre.

1.3. Troisième champ : le type

Il occupe un caractère. C'est un booléen qui indique si le mot-clé est descripteur de textes, c'est-à-dire si le mot-clé indexe ou est susceptible d'indexer des textes. Le type prend la valeur 0 si le mot-clé est un descripteur ; le type prend la valeur 1 si le mot-clé n'est pas un descripteur. Cette valeur est attribuée par les documentalistes ; en effet, la fiche thesaurus d'un descripteur doit être saisie avant les fiches documentaires qu'il indexe.

1.4. Quatrième champ : la note d'application

Elle occupe quatre lignes de 38 caractères chacune. Elle donne une explication du mot-clé et permet ainsi d'éviter les erreurs d'interprétation. C'est une zone rédigée et entièrement gérée par les documentalistes.

1.5. Les autres champs (5 à 20) : les mots-clés liés

Chaque mot-clé peut être relié à 8 autres mots-clés du thesaurus constituant les mots-clés liés au mot-clé de référence ; la relation entre le mot-clé de référence et un mot-clé lié peut être :

- . une relation de synonymie à laquelle correspond un code liaison 1
- . une relation d'analogie à laquelle correspond un code liaison 4
- . une relation de spécificité à laquelle correspond un code liaison 3
- . une relation de hiérarchie à laquelle correspond un code liaison 2

Chaque mot-clé lié occupe 38 caractères. L'ensemble des codes liaisons et des mots-clés (au maximum 8) définit l'environnement et la place du mot-clé de référence dans le thesaurus.

BORDEREAU : THESAURUS

MOT-CLE DE REFERENCE :

N° d'ORDRE

5 _____ 4

TYPE DE MOT CLE (0 = descripteur, 1 = non descripteur)

NOTE D'APPLICATION :

_____ 82

_____ 82

_____ 82

_____ 82

MOT-CLES DIRECTEMENTS LIES AU MOT CLE DE REFERENCE :

_____ 82

_____ 82

_____ 82

_____ 82

_____ 82

_____ 82

_____ 82

_____ 82

Code liaison (1 = synonyme, 2 = générique, 3 = spécifique, 4 = analogie)

BORDEREAU : THESAURUS

MOT-CLE DE REFERENCE :

N° d'ORDRE 1

A B A N D O N D E P O S T E
5 42

TYPE DE MOT CLE 1 (0 = descripteur, 1 = non descripteur)
43

NOTE D'APPLICATION :

6 44 45 82

7 44 45 82

8 44 45 82

9 44 45 82

MOT-CLES DIRECTEMENTS LIES AU MOT CLE DE REFERENCE :

44 45 82

44 45 82

44 45 82

44 45 82

44 45 82

44 45 82

44 45 82

44 45 82

Code liaison (1 = synonyme, 2 = générique, 3 = spécifique, 4 = analogie)

9

2. STRUCTURE D'ENSEMBLE DU THESAURUS

On peut se représenter le thesaurus comme un ensemble de classes :

- chaque classe a un représentant unique : le *mot-clé principal*
Les autres mots-clés de la classe sont les *synonymes* du mot-clé principal. Chaque synonyme n'est relié qu'à son mot-clé principal, cette liaison se faisant dans le sens unique : synonyme vers mot-clé principal ;
- les représentants de classe ou mots-clés principaux ont entre eux des relations de hiérarchie et d'analogie. Ce sont eux qui sont choisis comme descripteurs de textes.

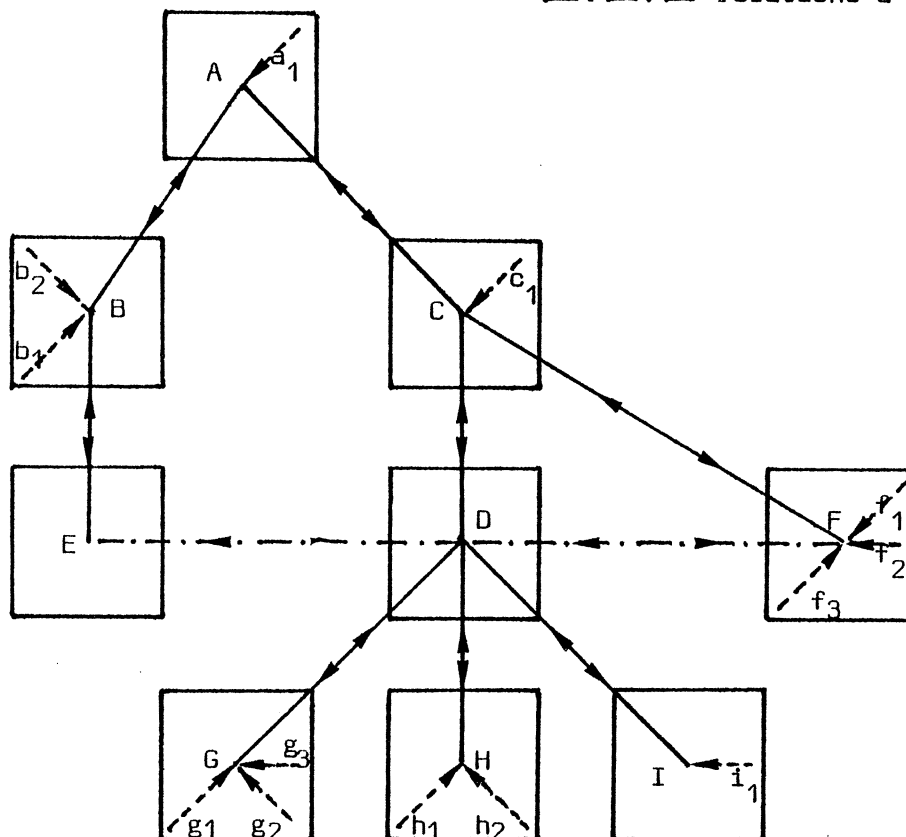
Représentation schématique d'une partie d'un thesaurus :

Chaque classe est représentée par un carré.

———— relations de hiérarchie

----- relations de synonymie

- . . . - relations d'analogie



Explication du schéma du thesaurus :

- . les lettres majuscules A, B, C, .. représentent les mots-clés principaux ;
- . les lettres minuscules a, b, c, .. représentent les synonymes associés à ces mots-clés principaux ;
- . les descripteurs de textes seront choisis dans l'ensemble {A, B, C, D, E, F, G, H, I} ;
- . la relation de synonymie se fait dans le sens unique synonyme vers mot-clé principal ;
- . les relations de hiérarchie se font dans les deux sens : par exemple, si A est générique de C, alors C est spécifique de A ;
- . par contre, une relation analogique peut exister dans les deux sens, comme par exemple D analogue de F et F analogue de D, ou exister en sens unique, comme par exemple D analogue de E, mais E n'est pas l'analogue de D.

Exemples de quelques fiches associées :

Fiche thesaurus de D :

D
3 C
2 G
2 H
2 I
4 F
4 E

Fiche thesaurus de E :

E
3 B

Fiche thesaurus de F :

F
3 C
4 D

Fiche thesaurus de f₁ :

f ₁
1 F

3. LES SYNONYMES

Si A est synonyme de B, la fiche thesaurus de A se présente de la façon suivante :

Exemple : temps de travail est synonyme
de durée hebdomadaire de travail.

A
1B

Rappel : les mots-clés principaux correspondent aux termes employés par les documentalistes, alors que les mots-clés synonymes ont été introduits pour permettre l'interrogation de la base par des non-documentalistes.

3.1. Première caractéristique

Ils ne peuvent indexer des textes (c'est pourquoi le type d'un mot-clé synonyme doit toujours être égal à 1, s'il n'en est pas ainsi, la fiche est refusée).

Grâce à la liaison mot-clé synonyme → mot-clé principal, en tapant un mot-clé synonyme, lors d'une interrogation, l'utilisateur pourra voir toutes les fiches documentaires indexées par le mot-clé principal correspondant. Ainsi en tapant A, il aura toutes les fiches indexées par B ; en tapant temps de travail, il aura toutes les fiches indexées par le mot-clé durée hebdomadaire de travail.

3.2. Seconde caractéristique

Unique liaison vers le mot-clé principal, ce qui signifie :

- . qu'il n'existe pas d'autre relation du synonyme vers un autre mot-clé : c'est pourquoi la présence d'un code liaison égal à 1 sur une fiche thesaurus, interdit toute autre relation sur cette fiche ;
- . aucun mot-clé ne renvoie sur un synonyme : ainsi, un mot-clé synonyme ne doit-il pas apparaître comme mot-clé lié sur une autre fiche. En particulier, il ne figure pas comme mot-clé lié sur la fiche de son mot-clé principal, ceci permet d'associer à un descripteur autant de synonymes qu'on le souhaite.

Cette double particularité des synonymes va permettre de perfectionner le lexique en fonction des interrogations et d'ajuster le thesaurus au langage des non-spécialistes. En effet, nous avons vu que le nombre de synonymes associés à un descripteur n'est pas limité ; par ailleurs, la suppression d'un synonyme est facile puisque cette opération n'exige que la suppression de la fiche correspondante ; aucune autre modification n'est à faire sur les fiches documentaires ou sur les autres fiches thesaurus (puisque un synonyme ne peut ni être un mot-clé lié ni un descripteur).

4. LES NON-SYNONYMES OU MOTS-CLÉS PRINCIPAUX

- . Ce sont eux qui sont choisis pour indexer les textes ;
- . ils ont entre eux des relations de hiérarchie ascendante et descendante et d'analogie.

4.1. La relation de hiérarchie

Exemple de relation hiérarchique :

SERVICE DES SPORTS (PERSONNEL) est un terme générique pour MAITRE NAGEUR ; inversement, MAITRE NAGEUR est un spécifique de SERVICE DES SPORTS (PERSONNEL).

Une relation hiérarchique dans un sens implique l'existence de la relation hiérarchique dans le sens inverse, mais dans ce domaine il n'y aura pas d'automatisme : c'est-à-dire que la présence d'une relation hiérarchique n'entraîne pas la création automatique de la relation hiérarchique inverse. Il en est de même pour la suppression de ce type de relation. Cet automatisme aurait été possible techniquement, mais aurait été en contradiction avec la philosophie générale du projet : "laisser aux documentalistes la maîtrise de leur domaine".

4.2. La relation d'analogie

Exemple de relation d'analogie :

COLLABORATEUR OCCASIONNEL et VACATION ;
 CESSATION DE FONCTION et DEMISSION,
 CESSATION DE FONCTION et LICENCIEMENT,
 CESSATION DE FONCTION et RETRAITE.

Un mot-clé est analogue d'un autre si sa signification est voisine ; l'existence d'une relation d'analogie dans un sens ne signifie pas qu'elle existe nécessairement dans le sens inverse. Il n'y a donc pas de contrôle à faire en ce qui concerne ce type de relation.

5. PROBLÈMES POSÉS PAR LA SUPPRESSION DES MOTS-CLÉS

Nous avons vu que la suppression de mots-clés synonymes ne posait aucun problème ; il n'en est pas de même des autres mots-clés. En effet, ceux-ci peuvent être descripteurs de textes, ou être présents en tant que mots-clés liés dans une autre fiche thesaurus. Leur suppression implique donc la suppression de toutes les liaisons vers eux, c'est-à-dire qu'il faut les supprimer dans les fiches thesaurus où ils apparaissent comme mots-clés liés ; il faut également les supprimer comme descripteurs de textes dans les fiches documentaires. Ce travail a été laissé à l'initiative de la documentaliste.

Mais pour éviter le risque d'avoir des mots-clés indexant un ou plusieurs textes et absents du lexique des mots-clés par suite d'une suppression, il a été décidé qu'aucune suppression effective n'aurait lieu au cours du programme de maintenance du thesaurus : les mots-clés que l'utilisateur veut supprimer sont simplement mis hors service grâce à un booléen ; un message indique à ce moment à l'utilisateur le nombre de liaisons vers ce mot-clé et le nombre de textes qu'il indexe, la suppression effective des enregistrements dans les fichiers ne devant avoir lieu que lors de la prochaine réorganisation du thesaurus, à condition que le mot-clé n'indexe aucun texte et qu'aucun mot-clé ne renvoie sur lui.

A la fin du programme général de contrôle, la liste des mots-clés mis hors service, mais présents comme descripteurs de textes ou comme mots-clés liés, est éditée.

La base de données étant juridique et concernant un domaine bien précis, l'ensemble des descripteurs et de leurs relations évolueront peu. Le thésaurus est très structuré et le noyau composé des descripteurs est rigide, par contre l'ensemble des synonymes pourra évoluer dans le temps, compte tenu de la fréquence d'utilisation de ces synonymes lors des interrogations.

6. ENSEMBLE DE CONTRÔLES EFFECTUÉS PAR LE SYSTÈME

La philosophie qui a déterminé le choix de tel ou tel contrôle était :

- . de signaler ou d'empêcher les erreurs les plus grossières qui auraient pour conséquence un mauvais fonctionnement du système ,
- . par ailleurs, de ne pas se substituer aux documentalistes dans leurs tâches : en ne leur refusant pas, par exemple, une opération de suppression, même si cette opération de suppression du mot-clé est en contradiction avec le fait que ce mot-clé indexe des textes.

Quand les contrôles ont-ils lieu ? Les contrôles sont effectués :

- . soit au moment de la saisie (ou de la modification) des fiches thesaurus ou textes,
- . soit dans le programme général de contrôle qui s'exécute après la réorganisation des fichiers.

On peut regrouper les contrôles en trois grandes catégories :

- . contrôles concernant les descripteurs,
- . contrôles concernant les mots-clés liés,
- . contrôles concernant les mots-clés synonymes.

6.1. Contrôles concernant les descripteurs

6.1.1. Existence dans le thesaurus des descripteurs de textes

Ce contrôle, effectué lors de la saisie (ou de la modification) d'une fiche documentaire, oblige l'utilisateur à saisir préalablement les fiches thesaurus des descripteurs. Cette obligation imposée à l'utilisateur n'est pas très contraignante ; en effet, celui-ci a à sa disposition le programme de création des fiches thesaurus. Cette solution a été jugée préférable à la création automatique des descripteurs dans le thesaurus.

Un deuxième contrôle complète le précédent : il consiste à mettre hors service un mot-clé descripteur au lieu de le supprimer effectivement ; ce deuxième contrôle évite comme le précédent d'avoir des textes indexés par des mots-clés inexistantes (la liste des mots clés mis hors service et indexant des textes est éditée dans le programme général de contrôle).

Par contre, une fiche documentaire comportant un descripteur qui a été mis hors service est refusée à la saisie. De plus, lors de la saisie d'une fiche documentaire, un descripteur mis préalablement hors service est refusé.

6.1.2. Un descripteur ne peut être un synonyme

Le contrôle est fait deux fois :

- . lors de la saisie/modification d'une fiche documentaire, tout descripteur synonyme est refusé,
- . par ailleurs lors de la modification ou de la création d'une fiche thesaurus, un mot-clé descripteur ne peut devenir synonyme.

6.1.3. Le type d'un descripteur doit être égal à zéro

Cette vérification se fait dans le programme général de contrôle avec message en cas d'erreur ; elle est plus un contrôle sur la donnée TYPE qu'un contrôle de cohérence interne au thesaurus.

6.2. Existence des mots-clés liés

Le refus d'une fiche thesaurus comportant un mot-clé lié inexistant ne peut être appliqué, il suffit pour s'en convaincre de considérer deux mots-clés hiérarchiquement dépendants :

Exemple : A est générique de B et B est spécifique de A ; quelle que soit la première fiche saisie (celle de A ou celle de B), elle serait refusée.

Comme dans le cas d'un descripteur inexistant, la solution consistant à créer automatiquement le mot-clé lié, n'a pas été retenue ; en effet, il pouvait s'agir d'une simple faute de frappe et il a été jugé préférable de prévenir l'utilisateur par un message, plutôt que de prendre l'initiative de créer des mots-clés à sa place.

Cependant, il serait possible de laisser à l'utilisateur le choix ou non d'une création automatique dans le thesaurus, d'un descripteur ou d'un mot-clé inexistant, sans que cette création automatique soit obligatoire.

Le risque de mots-clés liés inexistants est peu probable ; en effet, l'utilisateur est prévenu deux fois :

- . au moment de la saisie par un message à l'écran,
- . par un message figurant sur le listing des anomalies, édité par le programme général de contrôle.

Alors, à moins d'une volonté délibérée de la part de l'utilisateur, ce type d'erreurs ne devrait pas exister.

Un deuxième contrôle vient compléter le précédent : un mot-clé qui figure comme mot-clé lié dans d'autres fiches, ne peut être supprimé du lexique, il est simplement mis hors service. La liste des mots-clés mis hors service mais qui figurent toujours comme mots-clés liés, est éditée par le programme général de contrôle.

6.3. Contrôles concernant les synonymes

Nous avons vu que toute *relation d'un mot-clé synonyme vers un autre mot-clé* est refusée dès la saisie, à l'exception de la relation qui le relie à son mot-clé principal.

Par ailleurs, en ce qui concerne les *relations vers un mot-clé synonyme* qui ne doivent pas exister, l'utilisateur est prévenu de ce type d'erreur par un message figurant sur le listing des anomalies édité par le programme général de contrôle.

L'introduction de ce type de contrôles dès la saisie est

- . facile à assurer dans un sens : refus d'un mot-clé lié, si ce dernier est déjà un synonyme ;
- . difficile à assurer dans l'autre sens : refus d'un mot-clé synonyme si celui-ci est déjà mot-clé lié ; en effet, la seule variable indiquant si le mot-clé de référence est un mot-clé lié est une variable NL \bar{X} qui n'est mise à jour qu'au moment de la réorganisation et non en temps réel dans le programme de gestion du thesaurus, ce qui pourrait être envisagé :
 - cela permettrait d'empêcher toute relation vers un mot-clé synonyme,
 - mais augmenterait pas mal le délai de réponse au moment de la saisie.

Le type d'un synonyme doit être égal à 1 : ce contrôle concerne plus la donnée TYPE en elle-même que la cohérence du thesaurus.

6.4. Inexistence du mot-clé principal associé à un mot-clé synonyme

Le mot-clé principal étant le seul mot-clé lié au synonyme, ce contrôle est le même que celui de l'existence d'un mot-clé lié. Cependant, dans le listing des anomalies, le message "mot-clé lié inexistant" est complété par "il a des synonymes."

7. ANOMALIES ET CONTRÔLES DE COHÉRENCE DU THESAURUS

Type d'erreur	Contrôles qui permettent d'éviter ce type d'erreur	Message à l'écran lors de la saisie ou de la modification d'une fiche thesaurus ou documentaire	Message sur papier dans le listing des anomalies édité à la fin du programme général de contrôle	L'erreur entraîne-t-elle, en cas de non correction, le refus d'une fonction demandée par l'utilisateur ?
(1) Inexistence dans le thesaurus d'un descripteur	(a) Lors de la saisie/modification d'une fiche documentaire, contrôle d'existence des descripteurs (b) Lors de la suppression par l'utilisateur d'un mot-clé qui par ailleurs indexe des textes	"Le mot-clé X n'existe pas dans le dictionnaire"	Message sur papier dans le listing des anomalies édité à la fin du programme général de contrôle	Un descripteur inexistant dans le lexique est refusé lors de la saisie/modification d'une fiche documentaire.
(2) Existence de descripteur mis hors service auparavant	(b) Lors de la saisie/modification d'une fiche documentaire, contrôle que les descripteurs n'ont pas été mis hors service	Ce mot-clé indexe Y textes	Le mot-clé X a été mis hors service ; il indexe les textes suivants : ...	Le mot-clé descripteur n'est pas supprimé, il est mis hors service, après confirmation de la part de l'utilisateur de sa volonté de le mettre hors service.
(3) Descripteur synonyme	(a) Lors de la saisie/modification d'une fiche documentaire, vérification qu'aucun descripteur n'est synonyme. (b) Lors de la modification d'une fiche thesaurus, lorsqu'un mot-clé descripteur devient un synonyme	"Le mot-clé X a été mis hors service". "Le mot-clé X est un synonyme". "Le mot-clé X indexe des textes, il ne peut être un synonyme".		Un descripteur préalablement mis hors service est refusé lors de la saisie/modification d'une fiche documentaire. Un descripteur synonyme dans le lexique est refusé lors de la saisie/modification d'une fiche documentaire. Refus de la fiche thesaurus d'un synonyme, si celui-ci indexe des textes.
(4) Inexistence d'un mot-clé lié dans une fiche thesaurus	(a) Lors de la saisie/modification d'une fiche thesaurus, contrôle d'existence des mots-clés liés. (b) Lors de la suppression d'un mot-clé qui est présent comme mot-clé lié dans d'autres fiches thesaurus.	"Le mot-clé lié X n'existe pas". "Il y a Y liaisons vers ce mot-clé."	Le mot-clé lié X n'a pas encore été saisi. Le mot-clé X a été mis hors service, alors qu'il est présent comme mot-clé lié dans Y fiches thesaurus.	La fiche n'est pas refusée ; après le message à l'écran, l'utilisateur peut corriger le mot-clé lié ou le valider. Le mot-clé n'est pas supprimé, il est mis hors service après confirmation de la part de l'utilisateur de sa volonté de le mettre hors service.
(5) Inexistence d'un mot-clé principal (il figure comme mot-clé lié dans les fiches thesaurus de ses synonymes)	(a) Voir 4.a. (b) Voir 4.b.	Voir 4.a. Voir 4.b.	Voir 4.a. Voir 4.b. + le mot-clé X a des synonymes	Voir 4.a. Voir 4.b.
(6) Un mot-clé synonyme a des liaisons vers d'autres mots-clés	(a) Lors de la saisie/modification d'une fiche thesaurus, vérification qu'un mot-clé synonyme n'a qu'un mot-clé lié (le mot-clé principal).	Un seul mot-clé lié permis pour un mot-clé synonyme		La fiche thesaurus d'un synonyme est refusée si d'autres liaisons figurent sur la fiche.
(7) Existence de liaisons vers un mot-clé synonyme	(a) Vérification que les mots-clés liés ne sont pas des	Pas de message à l'écran lors de la saisie/modification	Incohérence : le mot-clé lié X est un synonyme de Y.	

Extrait du thesaurus de CATHI :

SECRETARIAT MEDICAL

H.A. : agent d'exécution chargé d'assurer le secrétariat
d'un service médical

TG SERVICE SOCIAL ET D'HYGIENE (PERSONNEL)

SECURITE DU TRAVAIL

VA C.H.S.

VA MEDECINE DU TRAVAIL

SECURITE SOCIALE

VA ACCIDENT DU TRAVAIL

VA AIDE SOCIALE

VA ALLOCATION D'INVALIDITE

VA ASSURANCE VIEILLESSE

VA MALADIE PROFESSIONNELLE

VA PRESTATION FAMILIALE

SERVICE ADMINISTRATIF (ENCADREMENT)

TG SERVICE ADMINISTRATIF (PERSONNEL)

TS ATTACHE

TS CHEF DE BUREAU

TS DIRECTEUR DES SERVICES ADMINISTRATIFS

TS REDACTEUR

TS SECRETAIRE GENERAL

SERVICE ADMINISTRATIF (EXECUTION)

TG SERVICE ADMINISTRATIF (PERSONNEL)

TS AGENT D'ENQUETE

TS AGENT DE BUREAU

TS APPARITEUR-ENQUETEUR

TS COMMIS

TS STENODACTYLOGRAPHE

VA = analogue

TS = terme spécifique

TG = terme générique

EM = synonyme

LICENCIEMENT

N.A. : action de licencier (motifs et cas) et droits qui en découlent pour les agents titulaires et non titulaires
 TG CESSATION DE FONCTION
 VA CHOMEUR
 VA COMMISSION PARITAIRE COMMUNALE
 VA REVOCATION
 VA SUPPRESSION D'EMPLOI

LIMITE D'AGE

TG STATUT
 VA RECRUTEMENT
 VA RETRAITE

Liquidation de retraite

EM PENSION DE RETRAITE

LISTE D'APTITUDE

N.A. : liste établie au niveau départemental de candidats éventuels à certains emplois communaux à la suite de concours ou au titre de la promotion sociale
 VA AVANCEMENT DE GRADE
 VA CONCOURS
 VA PROMOTION SOCIALE

LOGEMENT DE FONCTION

N.A. : avantage en nature acquis par les instituteurs en 1797 pour compenser un traitement insuffisant
 IG AVANTAGE

MAITRE-NAGEUR

N.A. : agent chargé d'assurer la surveillance des piscines des baignades et de donner des leçons de natation
 IG SERVICE DES SPORTS (PERSONNEL)

CLASSEMENT INDICIAIRE

H.A. : tableau fixant la suite des indices applicables
aux différents emplois
TG DUREE DE CARRIERE

C.N.R.A.C.L.

H.A. : caisse nationale de retraite des agents des collec-
tivités locales, agents titulaires et stagiaires
TG INSTITUTION
VA RETRAITE

COLLABORATEUR OCCASIONNEL

H.A. : Il s'apparente aux vacataires. Il couvre toutes
missions, techniques, médicales, socio-éducatives
TG AGENT NON TITULAIRE
TS ARCHITECTE
TS CONTRACTUEL
TS INGENIEUR DES VILLES
TS JURE DE CONCOURS
TS TECHNICIEN
VA VACATION
VA MISSION D'INGENIERIE ET D'ARCHITECTURE

COMITE MEDICAL

H.A. : Il est appelé à donner son avis sur les conditions
d'aptitude physique aux emplois publics
TG INSTITUTION
VA MEDECINE DU TRAVAIL

COMMS

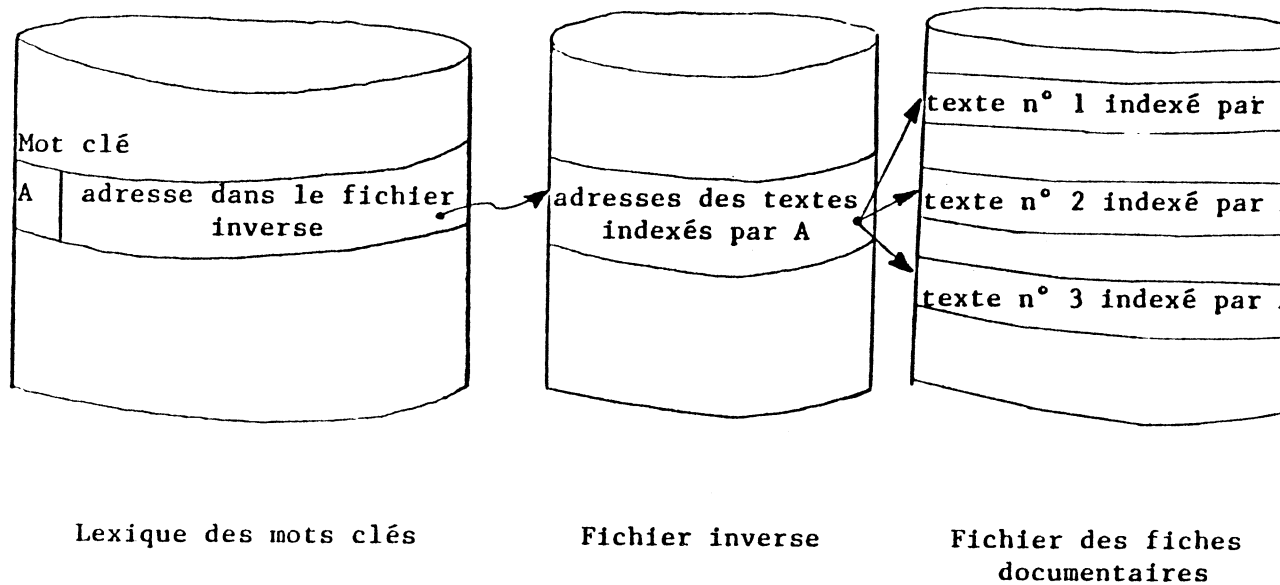
H.A. : agent chargé des tâches administratives d'exécution
TG SERVICE ADMINISTRATIF (Exécution)



CHAPITRE 4

DESCRIPTION TECHNIQUE DE LA 'BASE DE DONNÉES

La conception même du système documentaire : recherche d'un texte à partir de ses mots clés descripteurs, nécessitait l'existence d'un fichier inverse : chaque mot clé descripteur de texte a un ou plusieurs enregistrements dans ce fichier où sont consignées les adresses des fiches documentaires qu'il indexe. Structure que l'on peut représenter par le schéma suivant :



Nous pouvons répartir l'ensemble des fichiers de la base en 3 catégories :

- les fichiers relatifs au thésaurus
- les fichiers relatifs au fonds documentaire
- .- le fichier inverse

Le choix de la structure des fichiers et de la longueur d'enregistrement tient compte de la structure de la machine : l'unité de base dans un fichier est un secteur de 252 octets.

1. LES FICHIERS RELATIFS AU THESAURUS

- Explications sur l'organisation interne des fichiers du thésaurus
- Les fichiers de données
- Les fichiers de travail
- Le fichier des incohérences : "ANOMALIE"

1.1. Explications sur l'organisation des fichiers du thesaurus

La nécessité d'un accès rapide aux enregistrements des mots clés nous a conduit à envisager plusieurs solutions.

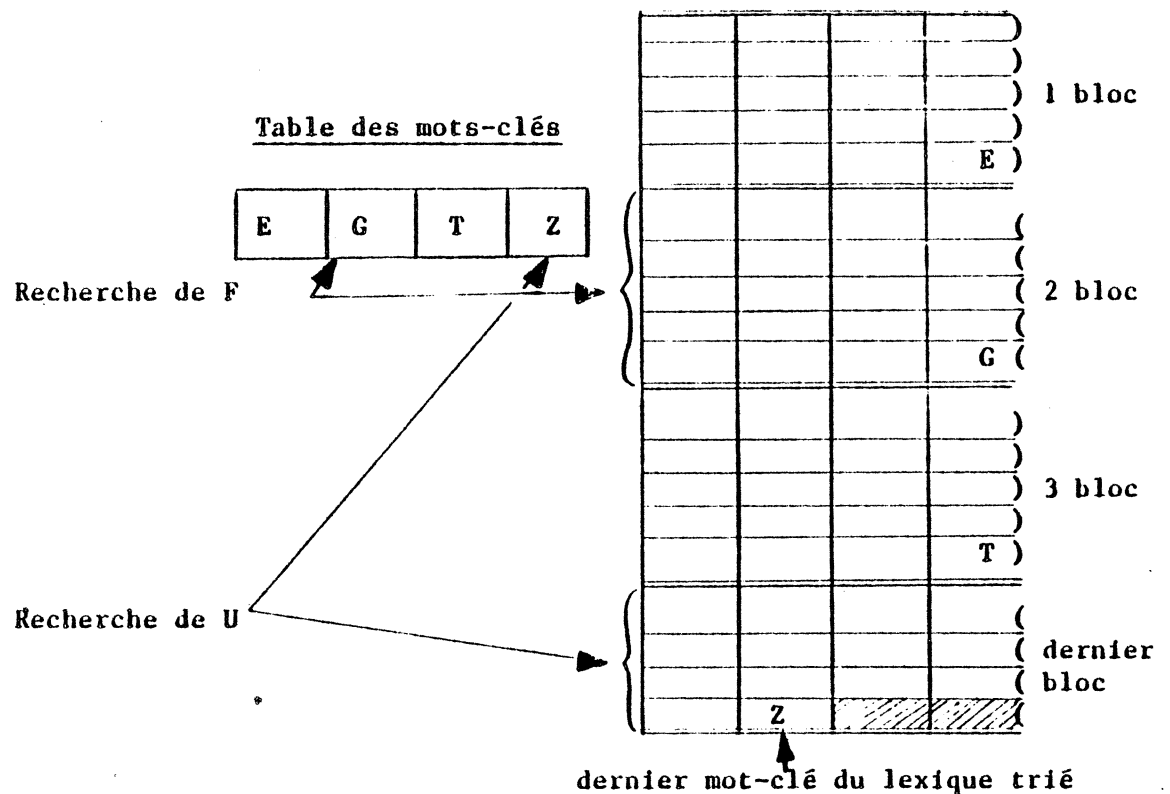
1ère solution : Disposer d'une table en mémoire permettant un accès direct sur l'enregistrement d'un mot clé donné. Cette solution fut écartée d'emblée, en effet, l'occupation mémoire de cette table serait trop importante : $20\text{ K pour }500\text{ mots clés } (38+2) \times 500 = 20\ 000\text{ octets}$, à raison de 38 pour le mot-clé, 2 pour l'adresse.

2ème solution : Faire une recherche dichotomique sur le fichier des mots clés. Avec un lexique de 1 500 Mots clés, il faudrait 8 accès disques pour trouver un mot-clé, et 9 pour avoir accès aux enregistrements correspondants au mot-clé (fichier de données et fichier inverse), ce qui représente un temps de recherche assez long.

3ème solution : La méthode du hash coding a pour inconvénient de faire perdre de la place dans les mémoires de stockage, (il y a des trous dans le fichier) et pose le problème des collisions. Elle ne fut pas retenue.

4ème solution : Celle qui fut adoptée est le séquentiel indexé. Elle consiste à faire une première sélection en utilisant une table en mémoire. Cette table est constituée en prenant un mot-clé tous les 20 dans le lexique trié. On compare le mot-clé recherché à ceux de la table et on détermine ainsi dans quel bloc il se trouve.

Une recherche séquentielle s'effectue dans le bloc sélectionné. Chaque bloc occupe 5 secteurs du disque.

Lexique des mots-clés

Remarques : 20 est un paramètre qui peut être modifié.

Inconvénients de la méthode :

- Avec un thésaurus de 500 mots-clés, la table occupe 950 octets
avec un thésaurus de 1500 mots-clés, la table occupe 2 850 octets
Cette table semble être trop importante compte tenu de la taille de certains programmes du logiciel (prog. recherche en particulier).
- Le lexique des mots-clés doit être trié. Un palliatif a été utilisé : on met les nouveaux mots-clés créés dans une extension à la fin du fichier et on lance le tri quand cette zone devient trop importante.

Propositions :

Diminuer la table en mémoire en ne prenant plus qu'un mot-clé tous les 75 (la table n'occupera plus que 760 octets pour 1500 Mots-clés), et faire une recherche dichotomique sur le bloc sélectionné. Ce bloc conte-

nant 75 mots-clés, soit 15 secteurs (avec 5 enregistrements par secteur), le nombre maximum d'accès pour trouver l'enregistrement d'un mot-clé serait de 4 ($\log_2 15 < \log_2 16 = 4$). Cette solution permettrait d'autre part de diminuer le nombre de comparaisons.

Avec 4 enregistrements par secteur, et en prenant 1 mot-clé tous les 100, il faudrait 5 accès disques maximum pour trouver l'enregistrement correspondant à un mot-clé.

Cette modification est facile à apporter, car :

- la dimension de la table en mémoire est un paramètre
- elle ne suppose qu'une modification partielle du programme de recherche dans DICO

Remarques : Les mots-clés de la partie extension sont consultables mais leur recherche est plus longue. Le délai de réponse pour un mot-clé inexistant est aussi plus long puisqu'on doit s'assurer qu'il n'est pas dans cette zone. Pour remédier à ce genre d'inconvénient, il faut lancer le programme de réorganisation qui comprend le tri, dès que le nombre de fiches thésaurus nouvellement saisies atteint un certain niveau, cela ne représente pas une grosse contrainte car en période de routine le nombre de créations de nouveaux mots-clés est peu élevé, d'autre part, l'exécution du programme de réorganisation ne nécessite pas la présence de l'utilisateur.

1.2. Fichiers de données concernant le thésaurus

Nous avons décidé de répartir les données concernant le thésaurus de la façon suivante, en regroupant les informations indispensables lors d'une recherche de fiche documentaire, dans :

- le fichier "DICO"

- Celui-ci contient
- le mot-clé de référence, le n° d'ordre, le type,
 - l'adresse relative dans le fichier INVERSE, le nombre de textes indexés,
 - l'adresse relative dans DICO du mot-clé principal dans le cas d'un synonyme,
 - le booléen indiquant la mise hors service ou non du mot-clé,
 - l'adresse de chaînage dans le fichier suivant.

- 2ème fichier chaîné au précédent : THESUS1

- Celui-ci contient
- le numéro d'ordre (même valeur que dans DICO)
 - le nombre de mots-clés liés
 - le nombre de mots-clés renvoyant sur lui
 - un booléen indiquant si le mot-clé est un mot-clé principal ou non
 - la note d'application
 - les 2 premiers mots-clés liés avec leurs codes liaisons
 - l'adresse de chaînage dans le fichier suivant. En effet il fallait un troisième enregistrement pour les mots-clés possédant plus de 2 mots-clés liés.

- 3ème fichier chaîné au précédent : THESUS2

- Celui-ci contient - le numéro d'ordre (même valeur que dans DICO et THESUS1)

- les autres mots-clés liés et leurs codes liaisons respectifs. Seul DICO est trié. Cependant ce fichier n'est pas en permanence trié. Comme nous l'avons vu précédemment il est composé de 2 parties : une partie triée et une extension (partie non triée).

D'autre part, le fichier DICO ne comporte jamais de trous (les mots-clés à supprimer ne sont pas recopiés dans DICO après le tri).

Par contre, les fichiers THESUS1 et THESUS2, eux, ne sont jamais triés et peuvent comporter des trous quand des enregistrements sont effacés.

1.2.1. DICO : Premier fichier de données concernant le thésaurus

A) Caractéristiques du fichier

C'est un fichier permanent organisé en accès direct.

Son format d'enregistrement est fixé (voir schéma d'enregistrement)

Longueur de chaque enregistrement : 63 octets

Nombre d'enregistrements par secteur : 4

Volume pour 300 Mots-clés : 75 secteurs, soit 19 200 octets.

B) Ce fichier est créé lors de la première saisie des fiches documentaires.

- Il est mis à jour grâce au programme de maintenance du thésaurus qui assure les fonctions de création, de modification de fiches thésaurus et de mise hors service d'un mot-clé (en effet quand l'utilisateur désire supprimer un mot-clé aucun enregistrement n'est effacé dans ce programme).

- Il est réorganisé à la demande de l'utilisateur ; au cours de cette réorganisation :

. le fichier est trié avec remise à jour des numéros d'ordre,

. a lieu la suppression effective des enregistrements correspondant aux mots-clés mis hors service, à condition que ceux-ci n'indexent aucun texte et qu'aucun mot-clé ne renvoie sur eux.

(Les enregistrements sont tassés, DICO ne comporte pas de trous)

. 2 zones de DICO sont mises à jour : adresse relative dans DICO du mot-clé principal et indice d'enregistrement du mot-clé principal dans son secteur.

- Les variables de DICO sont contrôlées au moment de la saisie et dans le programme général de contrôles qui fait suite à la réorganisation.

Paramètres concernant DICO

PRZ (2) : Adresse relative du dernier mot-clé de DICO (adresse correspondant au secteur)

PRZ (1) : Indice d'enregistrement du dernier mot-clé de DICO dans son secteur

PRZ (5) : Nombre d'enregistrements dans DICO

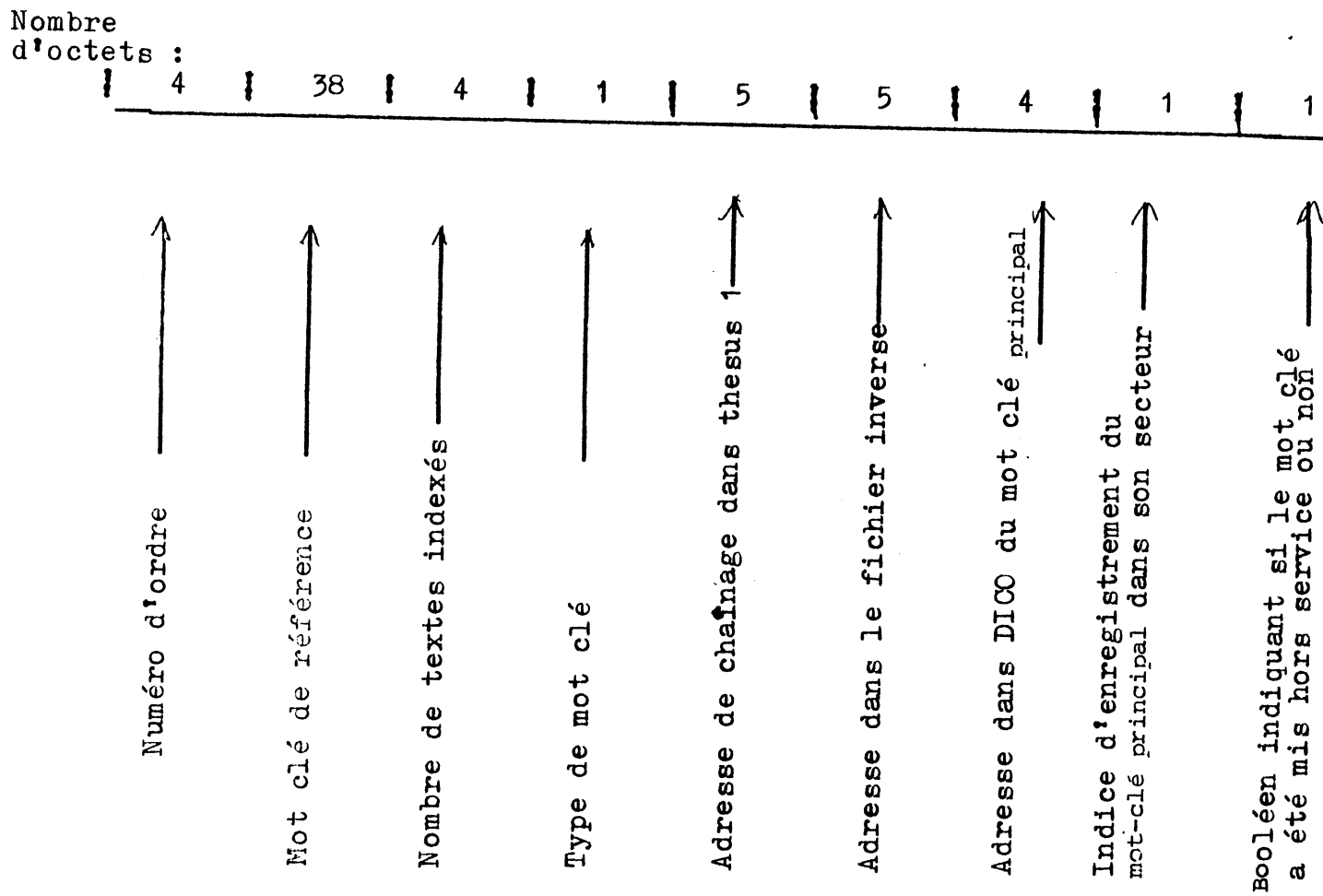
PRZ (7) : Adresse relative du dernier mot-clé de la partie triée de DICO

PRZ (8) : Indice d'enregistrement dans son secteur du dernier mot-clé de la partie triée de DICO

PRZ (9) : Nombre de créations dans DICO depuis le dernier tri

PRZ (12) : Plus grand numéro d'ordre dans DICO

Dessin d'enregistrement de DICO



Voir en annexe la liste des variables et le tableau des variables.

1.2.2. THESUS1 : 2ème fichier de données concernant le thésaurus

A) Caractéristiques du fichier

- C'est un fichier permanent organisé en accès direct
- Son format d'enregistrement est fixe (voir schéma)
- Longueur d'un enregistrement : 246 octets
- 1 seul enregistrement par secteur
- Volume occupé par THESUS1 pour 300 mots-clés : 300 secteurs soit 75 600 octets.

B) Ce fichier est créé et mis à jour en même temps que DICO. Ce fichier n'est jamais trié.

A chaque mot-clé correspond un enregistrement dans THESUS1 et un seul.

Comme pour DICO, quand un mot-clé est mis hors service, l'enregistrement dans THESUS1 correspondant à ce mot-clé n'est pas effacé. L'effacement de l'enregistrement est effectué dans le programme de réorganisation (et uniquement dans ce programme) sous certaines conditions (voir DICO). THESUS1 peut donc comporter des trous. Les enregistrements vacants pourront être réutilisés, lors d'une prochaine création car les adresses correspondantes sont stockées dans un fichier (PILE).

Les variables de THESUS1 sont contrôlées au moment de la saisie (modification) et dans le programme général de contrôles.

Paramètres concernant THESUS1

PR% (4) : Adresse relative au dernier enregistrement dans THESUS1

PR% (10) : Nombre d'enregistrements vacants dans THESUS1

Dessin d'enregistrement de THESUS 1

Nombre d'octets :

4	38	38	38	38	2	1	38	1	38	4	1	5
---	----	----	----	----	---	---	----	---	----	---	---	---

Numéro d'ordre	1ère ligne de la note d'application	2ème ligne de la note d'application	3ème ligne de la note d'application	4ème ligne de la note d'application	Nombre de mots clés liés	1er code de liaison	1er mot clé lié	2ème code de liaison	2ème mot clé lié	Nombre de liaisons vers le mot clé de référence	Booléen indiquant si le mot clé est un mot clé principal ou non	Adresse relative de chaînage dans Thesus 2
----------------	-------------------------------------	-------------------------------------	-------------------------------------	-------------------------------------	--------------------------	---------------------	-----------------	----------------------	------------------	---	---	--

Voir en annexe la liste des variables et le tableau des variables.

1.2.3. THESUS2 : 3ème fichier de données concernant le thésaurus

A) Caractéristiques du fichier

- fichier permanent organisé en accès direct.
- format d'enregistrement fixe (voir schéma)
- longueur d'un enregistrement : 238 octets
- 1 seul enregistrement par secteur
- Volume maximum occupé par le fichier pour 300 Mots-clés : 300 secteurs soit 75 600 octets.

B) Ce fichier est créé et mis à jour en même temps que DICO et THESUS1. Ce fichier n'est jamais trié.

Chaque mot-clé a au maximum 1 enregistrement dans ce fichier (il peut ne pas en avoir un, si son nombre de mots-clés liés est <2)

Ce fichier peut comporter des trous : un enregistrement de ce fichier est effacé dans les 2 cas suivants :

- 1/ au moment de la réorganisation quand un mot-clé est effectivement supprimé des fichiers,
- 2/ au moment de la modification d'une fiche thésaurus (ou de la re-création d'un mot-clé mis hors service) quand le nombre de mots-clés liés qui était >2 devient <2.

Dans les deux cas, l'adresse de l'enregistrement vacant est stockée dans le fichier "PILE" pour être récupérée lors de la création d'un nouvel enregistrement.

Les variables de ce fichier sont contrôlées au moment de la saisie et dans le programme général de contrôles.

Paramètres concernant THESUS2

PR% (3) : adresse relative du dernier enregistrement dans THESUS2

PR% (11) : nombre d'enregistrements vacants dans THESUS1

Dessin d'un enregistrement de THESUS 2

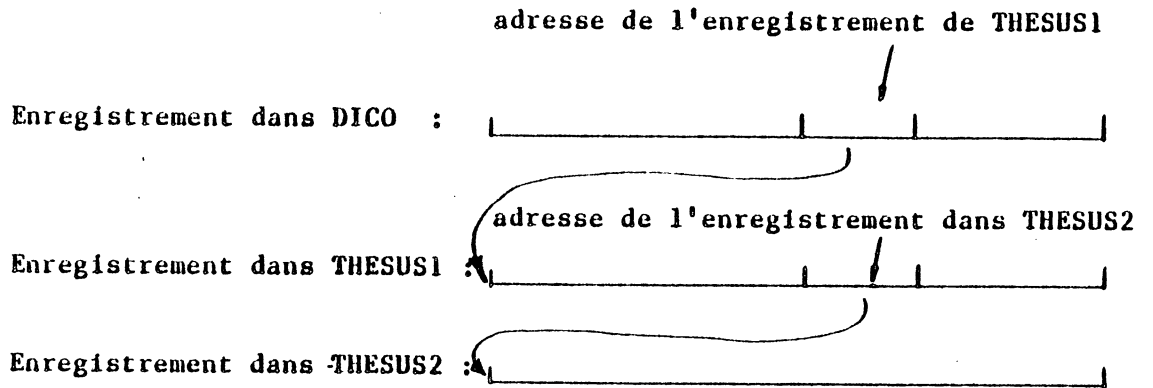
Nombre d'octets :

	4	1	38	1	38	1	38	1	38	1	38	1	38	1	:
	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Numéro d'ordre															
3ème code liaison															
3ème mot clé lié															
4ème code liaison															
4ème mot clé lié															
5ème code liaison															
5ème mot clé lié															
6ème code liaison															
6ème mot clé lié															
7ème code liaison															
7ème mot clé lié															
8ème code liaison															
8ème mot clé lié															

Voir en annexe la liste des variables.

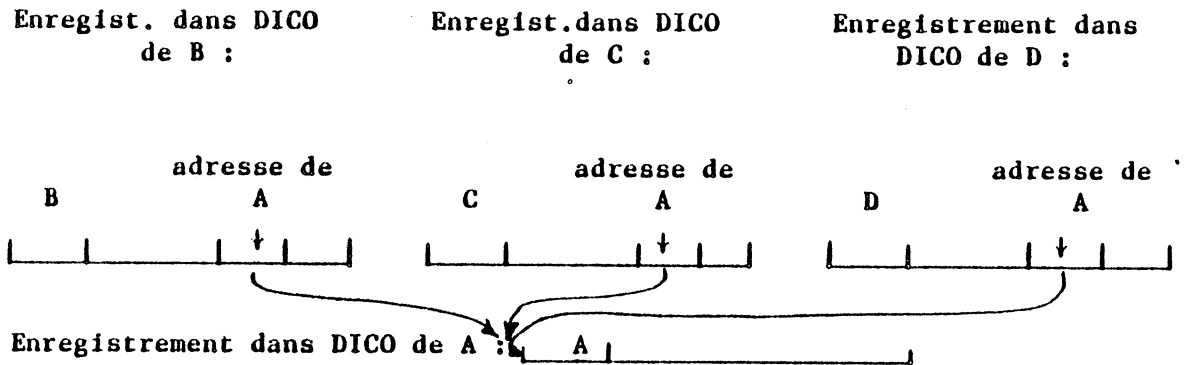
1.2.4. Schéma des différents chaînages

- Pour un même mot-clé les enregistrements dans DICO, THESUS1, THESUS2, sont chaînés entre eux.

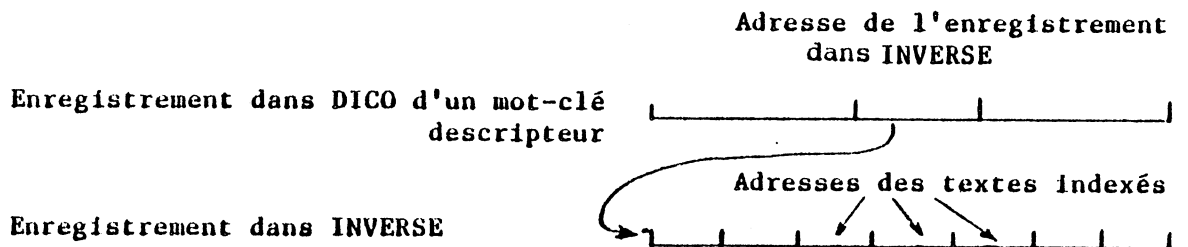


- L'enregistrement dans DICO d'un mot principal est chaîné à celui de ses synonymes.

Ex. si B, C, D, sont les synonymes de A



- Pour un mot-clé descripteur, l'enregistrement dans le fichier inverse est chaîné à celui de DICO



(L'enregistrement dans INVERSE contient les adresses des textes indexés par le mot-clé).

1.3. Les fichiers de travail concernant le thésaurus

Ces fichiers contiennent des informations dont l'utilité est surtout informatique. On en distingue quatre :

- Le fichier des paramètres : PARAM

Ce fichier commun au thésaurus et au fonds documentaire contient la liste des différents paramètres concernant les fichiers de données du thésaurus et du fonds documentaire et le fichier "INVERSE".

- Le fichier contenant la table des mots-clés : TABLEAU
- Le fichier contenant la liste des adresses libres de THESUS1 et THESUS2 : PILE

Ces trois fichiers sont organisés en tableaux virtuels, c'est-à-dire comme des vecteurs : on accède à un élément du tableau par son indice ce qui facilite la programmation.

- Enfin, le fichier intermédiaire créé par le tri : DICOTRIE

1.3.1. 1er fichier de travail concernant le thésaurus

PARAM : fichier des paramètres

A) Caractéristiques du fichier

- Fichier permanent organisé en tableau virtuel : (PR%, dimensionné à 23)
- L'accès est direct et se fait par l'indice
- Volume occupé par ce fichier : 2 x 24 = 48 octets
- Ce fichier est créé par le programme de création du thésaurus
- Il est mis à jour dans les programmes de maintenance du thésaurus et du fonds documentaire, ainsi que dans le programme de réorganisation
- Toutes les valeurs des paramètres sont calculées par programme.

B) Contenu de ce fichier

Les treize premiers entiers (PR% (I), I = 0 à 12) sont les paramètres de gestion des fichiers du thésaurus (voir liste en annexe).

Il y a aussi 9 paramètres de gestion du fonds documentaire (PR% (I), voir liste en annexe, I = 15 à 23). 2 paramètres supplémentaires pourront être ajoutés (PR% (I), I = 13 à 14).

1.3.2. 2ème fichier de travail concernant le thésaurus

TABLEAU : Table des mots-clés

- Ce fichier table permet une recherche rapide d'un mot-clé. Après chaque tri, cette table est recréée en prenant un mot-clé tous les 20 dans le fichier DICO (20 est un paramètre). Ce fichier est écrit en mémoire au début de chaque programme.
- C'est un fichier permanent organisé en tableau virtuel et désigné par TA\$ dans les programmes.
- Volume occupé par ce fichier :
 - . Si on prend 1 mot-clé tous les 20
 - Pour 300 Mots-clés : $38 \times \frac{300}{20} = 570$ octets
 - Pour 500 Mots-clés : $38 \times \frac{500}{20} = 950$ octets
- En dehors du programme de réorganisation du thésaurus, ce fichier n'est jamais modifié.
- Chaque élément du tableau est un mot-clé. PR% (6) est la dimension de TABLEAU.
Le paramètre PR% (6) est recalculé en même temps qu'est recréé TABLEAU

1.3.3. 3ème fichier de travail concernant le thésaurus

PILE : Table des adresses libres de THESUS1 et de THESUS2

Nous avons vu que les fichiers THESUS1 et THESUS2 ne sont jamais tassés ni triés quand un enregistrement est effacé dans ces fichiers. L'adresse de l'enregistrement vacant est stockée dans le fichier PILE en vue d'une prochaine réutilisation.

Caractéristiques du fichier

- Fichier permanent organisé en tableau virtuel (P% (1,50))
- Accès direct par l'indice du tableau
- Volume maximum occupé par ce fichier $2 \times 50 \times 2 = 200$ octets
(On prévoit un maximum de 50 adresses libres dans THESUS1 et 50 adresses libres dans THESUS2).
- La 1ère colonne de PILE (P% (0,-)) contient les adresses libres de THESUS1
La 2ème colonne de PILE (P% (1,-)) contient les adresses libres de THESUS2.
- Ce fichier est créé et modifié dans les programmes de maintenance et de réorganisation du thésaurus.
PR% (10) est le paramètre qui indique le nombre d'adresses libres de THESUS1
PR% (11) est le paramètre qui indique le nombre d'adresses libres de THESUS2.

1.3.4. 4ème fichier de travail concernant le thésaurus

Fichier intermédiaire créé par le TRI : DICOTRIE

Pour trier le fichier DICO, le programme de TRI utilisé ne modifie pas DICO, mais recopie ce dernier en le triant dans un fichier binaire et séquentiel : "DICOTRIE". Ce fichier est bien sûr détruit avant chaque tri et recréé par le tri, il n'est jamais modifié. Après le tri, DICO est recréé à partir de ce fichier : DICOTRIE.

1.4.

FICHER : "ANOMALIE"Caractéristiques du fichier

Fichier permanent organisé en accès direct

Son format d'enregistrement est fixe (voir schéma d'enregistrement)

Il y a un enregistrement par secteur

Son volume est variable et dépend du nombre d'erreurs décelées

Ce fichier est détruit et recréé après le tri quand les contrôles sont effectués.

Il est édité entièrement à la fin du programme de contrôles.

Ce fichier n'est jamais modifié. Il n'est pas trié, car le fichier "DICO" l'étant, les mots-clés erronés, sont recopiés dans "ANOMALIE" dans l'ordre alphabétique, et ils apparaissent triés à l'édition des erreurs.

Le fichier "ANOMALIE" est constitué de 2 parties.

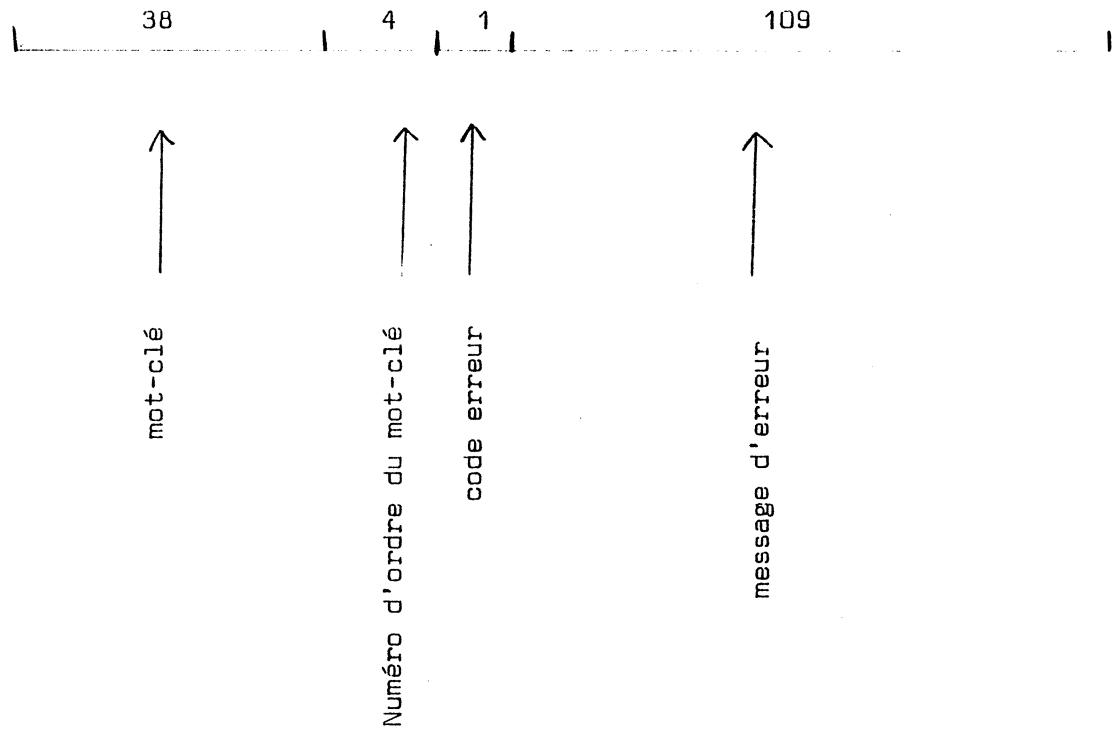
1ère partie qui contient la liste des mots-clés où figure une erreur de type 1, 2 ou (et) 3.

<u>Code erreur</u>	<u>Message d'erreur correspondant</u>
1	"(1) : mot-lié non encore saisi"
2	"(2) : incohérence, mot synonyme de Y"
3	"(3) : incohérence entre type et liaison texte"

2ème partie contient la liste des mots-clés qui ont été mis hors service mais qui ne peuvent être supprimés, le type d'erreur est dans ce cas 4 ou 5.

<u>Code erreur</u>	<u>Message d'erreur correspondant</u>
4	"(4) : on ne peut supprimer ce mot-clé qui indexe les textes suivants"
5	"(5) : ce mot-clé est présent dans X fiches thésaurus"
5	"(5) : ce mot-clé est présent dans X fiches thésaurus, il a des synonymes"

Nombre d'octets :



Voir liste des variables en annexe.

2. LES FICHIERS RELATIFS AU FONDS DOCUMENTAIRE

Ces fichiers peuvent être répartis en 2 groupes :

- Les fichiers de données : TEXTE, ANALYSE, MODES
- Les fichiers de travail : PARAM, PILTEX

2.1. Les fichiers de données

- Le premier fichier accessible soit à partir du numéro de fiche du texte, soit par le fichier "INVERSE" est le fichier : "TEXTE"

Ont été rassemblées dans ce fichier :

- . les informations présentes dans toutes les fiches documentaires : références, source du droit, 2 premières lignes de l'analyse, les 3 premiers descripteurs. (La plupart des fiches documentaires n'ont pas plus de 3 descripteurs).
 - . Les informations de première importance : booléen indiquant si le texte doit être supprimé,* la date de création de la fiche, le nombre de descripteurs.
 - . Ainsi que les adresses de chaînage dans les fichiers MODES et ANALYSE.
- Le fichier "ANALYSE" contient le reste de l'analyse du texte soit 12 lignes.
 - Le fichier "MODES" (Mots DEscripteurs) contient les autres descripteurs (du 4ème au 14ème).

Tous les enregistrements concernant une même fiche documentaire contiennent le numéro de fiche, car ce numéro sert d'identificateur pour une fiche documentaire et de clé d'accès dans le fichier TEXTE.

* Quand l'utilisateur désire supprimer une fiche documentaire, les enregistrements concernant cette fiche ne sont pas effacés, un booléen indique que la fiche documentaire doit être supprimé, la suppression réelle dans les fichiers aura lieu ultérieurement dans un programme spécifique.

2.1.1. ler fichier de données du fonds documentaire : "TEXTE"

A) Caractéristiques du fichier

- fichier permanent,
- fichier organisé en accès direct,
- le format d'enregistrement est fixé (voir schéma d'enregistrement)
la longueur d'enregistrement est 245 octets
il y a un enregistrement par secteur
pour 5000 fiches documentaires, ce fichier occupe 5000 secteurs,
soit : 1 260 000 octets.

En effet chaque fiche documentaire a 1 enregistrement et 1 seul dans ce fichier. On dira qu'il contient l'entête du texte.

B) Ce fichier est créé par le programme de création des fiches documentaires quand celles-ci sont saisies la première fois.

Il est mis à jour par le programme de maintenance du fonds documentaire.

Il n'est jamais trié.

Ce fichier peut comporter des trous, en effet quand un enregistrement de ce fichier sera effacé, on récupèrera l'adresse de l'enregistrement vacant dans le fichier "PILTEX".

Les seuls contrôles effectués sur certains champs des enregistrements de "TEXTE" sont faits à la saisie ou lors de la modification d'une fiche documentaire.

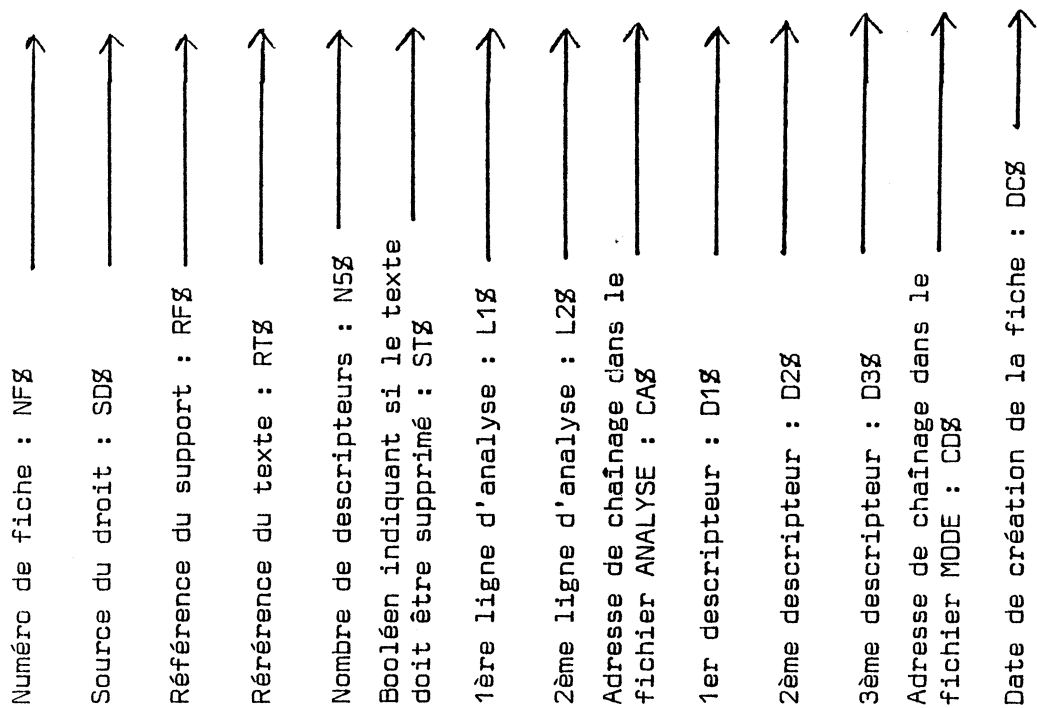
Dessins d'enregistrement du fichier "TEXTE"

- La variable NF\$ sert de clé d'accès, elle correspond à l'adresse relative de l'enregistrement.
- On accède à l'enregistrement correspondant du fichier "MODES" par l'adresse CD\$
- On accède à l'enregistrement correspondant du fichier "ANALYSE" par l'adresse CA\$

L'adresse du dernier enregistrement du fichier TEXTE est : PR% (20)

Nombre d'octets :

4 1 17 14 2 1 38 38 5 38 38 38 5 6



Voir liste des variables en annexe.

2.1.2. 2ème fichier de données du fonds documentaires : "ANALYSE"

A) Caractéristiques du fichier

- fichier permanent,
 - fichier organisé en accès direct,
 - son format d'enregistrement est fixe (voir schéma d'enregistrement)
- longueur d'un enregistrement : 237 octets
 un seul enregistrement par secteur
 volume correspondant à 5000 fiches documentaires : 10 000 secteurs, en effet une fiche documentaire a toujours 2 enregistrements dans ce fichier. (Pour 12 lignes d'analyse, il fallait 2 enregistrements).

B) Ce fichier est créé en même temps que le fichier "TEXTE".

Il est mis à jour par le programme de maintenance du fonds documentaire.

Il n'est jamais trié, ni tassé, il peut contenir des trous : quand un enregistrement de ce fichier est effacé, on récupère l'adresse libre dans le fichier table : "PILTEX", de manière à pouvoir réutiliser l'enregistrement vacant lors d'une prochaine insertion. (Pas fait à ce jour).

L'adresse du dernier enregistrement du fichier ANALYSE est PR% (21).

C) Dessins d'enregistrement dans le fichier "ANALYSE"

F1\$ a la même valeur que la variable NF\$ du fichier "TEXTE"
 CE\$ permet de chaîner le premier enregistrement au second
 dans le second enregistrement CE\$ = " "

Nombre d'octets :

	4	38	38	38	38	38	38	5
								↑
↑		↑	↑	↑	↑	↑	↑	
Numéro de fiche F1\$		Ligne d'analyse L3\$	Ligne d'analyse L4\$	Ligne d'analyse L5\$	Ligne d'analyse L6\$	Ligne d'analyse L7\$	Ligne d'analyse L8\$	Adresse de chaînage à l'autre enregistrement de ce fichier et concernant la même fiche documentaire CE\$

F1\$ a la même valeur que la variable NF\$ du fichier TEXTE.
 CE\$ permet de chaîner le premier enregistrement au second dans le second enregistrement CES = " ".

Liste des variables en annexe.

2.1.3. 3ème fichier de données concernant le fonds documentaire : "MODES"

A) Caractéristiques du fichier

- fichier permanent,
- fichier organisé en accès direct,
- son format d'enregistrement est fixe (voir schéma d'enregistrement)

longueur d'un enregistrement : 237 octets

un seul enregistrement par secteur

volume pour 5000 fiches documentaires est variable mais $< 10\ 000$ secteurs. En effet, si la fiche documentaire a moins de 4 descripteurs, elle n'a pas d'enregistrement dans ce fichier.

Si le nombre de descripteurs est ≥ 4 et ≤ 9 , elle a un enregistrement dans ce fichier.

Si le nombre de descripteurs est >9 , elle a deux enregistrements dans ce fichier.

B) Ce fichier est créé en même temps que les fichiers "TEXTE" et "ANALYSE".

Il est mis à jour par le programme de maintenance du fonds documentaire.

Comme les fichiers "TEXTE" et "ANALYSE" il n'est jamais trié, ni tassé, lors de l'effacement d'un enregistrement de ce fichier, l'adresse libérée est stockée provisoirement dans le fichier "PILTEX" afin de pouvoir être récupérée lors d'une prochaine insertion.

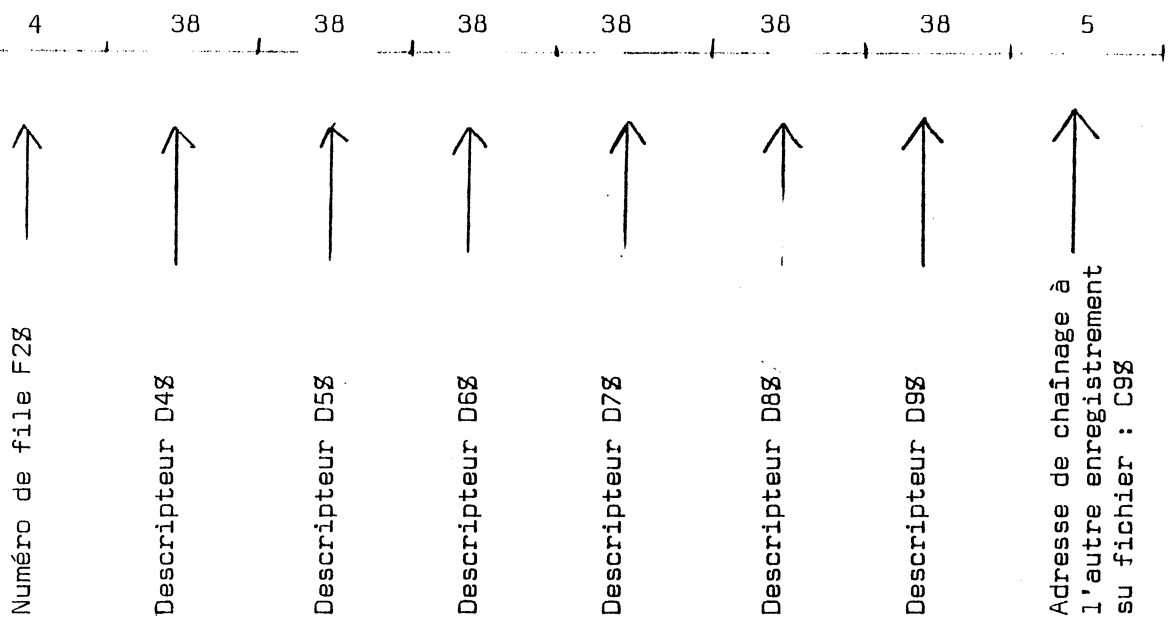
C) Dessin d'enregistrement du fichier "MODES"

La variable F2\$ a la même valeur que les variables F1\$ et NF\$

C9\$ permet de chaîner le 1er enregistrement au second (quand il existe)
s'il n'y a pas de second enregistrement alors C9\$ = " "
dans le second enregistrement C9\$ = " "

L'adresse du dernier enregistrement du fichier MODES est PR% (22).

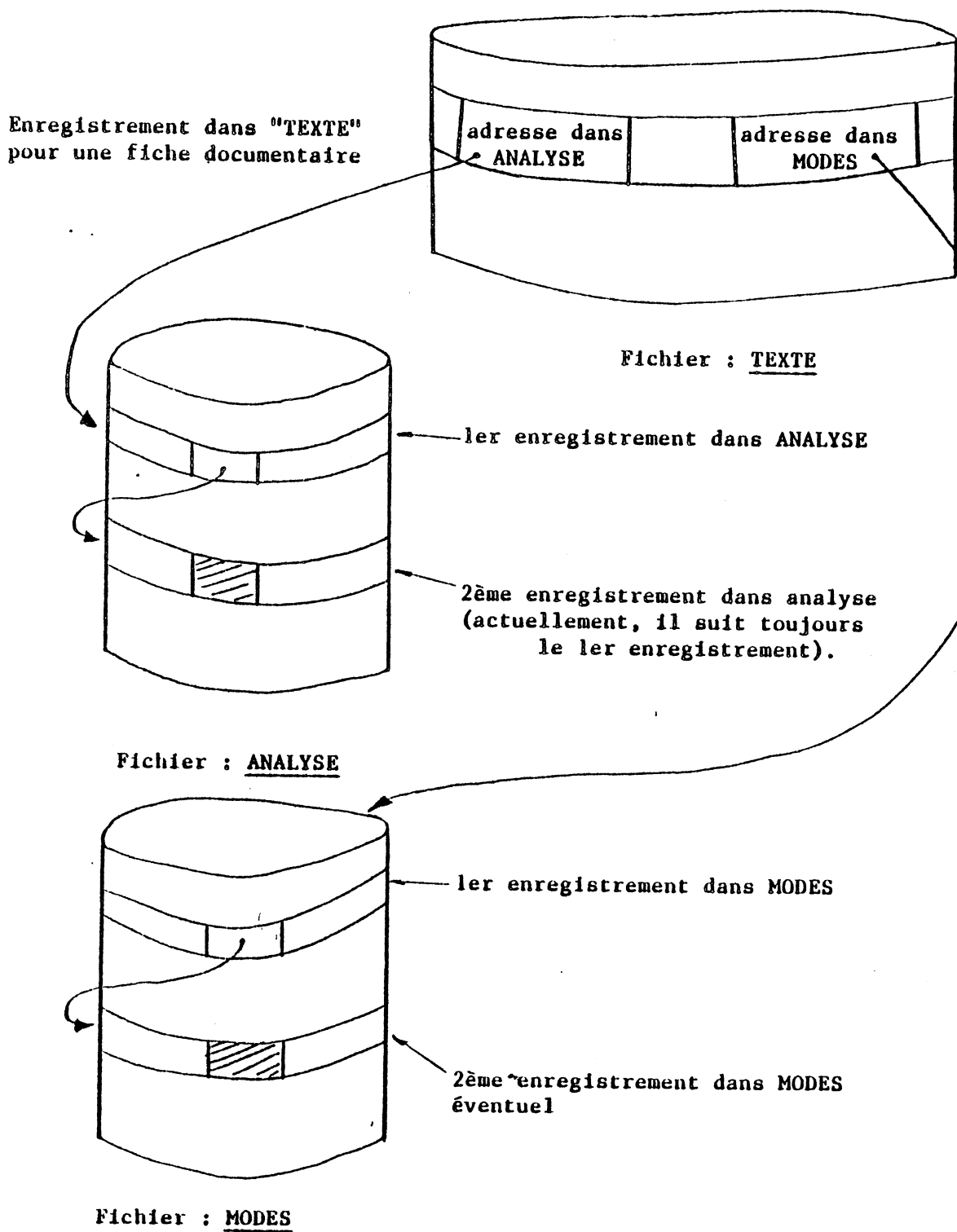
Nombre d'octets :



Voir liste des variables en annexe.

2.1.4. Différents chaînages entre les fichiers

de données du fonds documentaire



2.2. Les fichiers de travail concernant le fonds documentaire :

Ils sont au nombre de 2.

- Le fichier des paramètres : "PARAM" dont nous avons déjà parlé
propos du thésaurus.
- Le fichier contenant la liste des adresses libres des fichiers TEXTE
ANALYSE, MODES, INVERSE, c'est le fichier table : "PILTEX"

Ces 2 fichiers sont organisés en tableaux virtuels.

Pour "PARAM" se reporter à la description faite dans les fichiers de travail du thésaurus.

PILTEX : Fichier table des adresses libres des fichiers)TEXTES
)ANALYSE
)MODES
)INVERSE

Nous avons vu que lors de l'effacement d'un enregistrement dans un des fichiers (TEXTE, ANALYSE, MODES, INVERSE), son adresse est stockée dans le fichier "PILTEX".

"PILTEX" est un fichier permanent, organisé en tableau virtuel désigné par P1% dans les programmes (P1% (4,62))

- La 1ère colonne contient les adresses libres de "TEXTE" (P1%(1,62))
- La 2ème colonne contient les adresses libres de "ANALYSE"(P1%(2,62))
- La 3ème colonne contient les adresses libres de "MODES" (P1%(3,62))
- La 4ème colonne contient les adresses libres de "INVERSE"(P1%(4,62))

Le nombre d'adresses libres dans TEXTE, ANALYSE, MODES, INVERSE sont respectivement les paramètres de PARAM : PR%(16), PR%(17), PR%(18) PR%(19)

3. FICHER INVERSE

- C'est un fichier permanent organisé en tableau virtuel (AZ\$(62))
- Un mot-clé descripteur a 1 ou plusieurs enregistrements dans ce fichier. Ces enregistrements sont chaînés entre eux (le nombre d'enregistrements par mot-clé n'est pas limité).
- Le volume occupé par ce fichier est variable et dépend :
 - . du nombre de descripteurs
 - . du nombre de textes indexés par chaque descripteur (certains descripteurs peuvent indexer une centaine de textes).

- Ce fichier est créé par le programme de création des fiches documentaires.

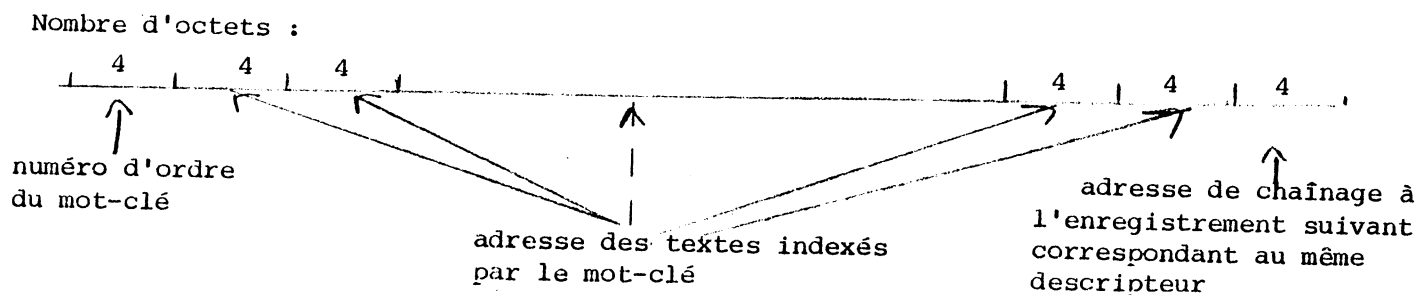
Il est mis à jour dans le programme de maintenance du fonds documentaire :

- . au moment de l'ajout d'un texte,
- . ou de l'ajout ou de la suppression d'un descripteur dans une fiche documentaire.

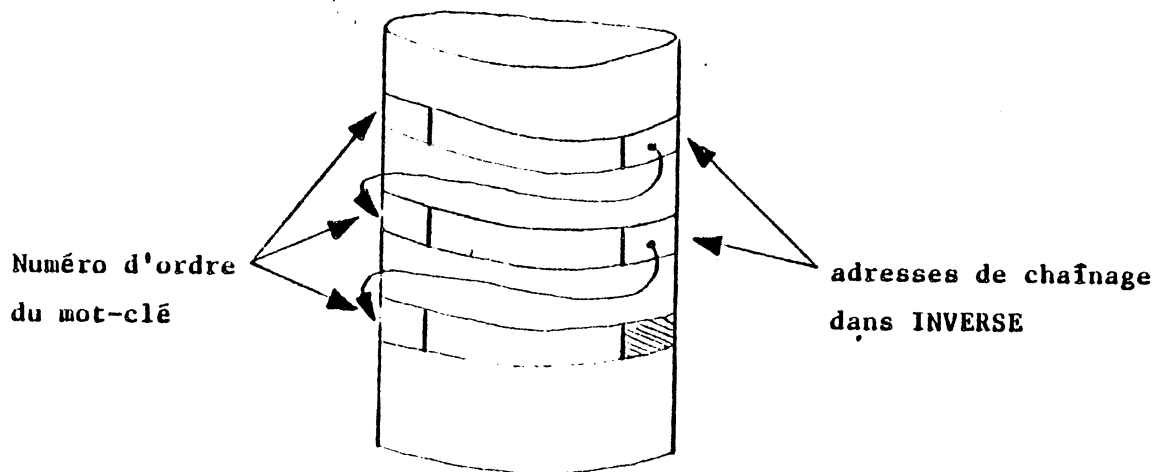
Dessin d'enregistrement : Chaque enregistrement peut contenir jusqu'à 61 adresses de textes.

Comme on le voit sur le dessin de l'enregistrement

- AZ\$ (Ø) : contient le numéro d'ordre du mot-clé
- AZ\$ (I) pour I = 1 à 61 : adresses des textes indexés (adresses dans le fichier TEXTE)
- AZ\$(62) : adresse de chaînage à l'enregistrement suivant (s'il existe)
S'il n'existe pas alors VAL(AZ\$(62)) = Ø



Chainage entre les différents enregistrements
concernant un même descripteur



Les adresses des textes dans les différents enregistrements concernant un même descripteur sont insérées et supprimées, en conservant l'ordre croissant et en les tassant de manière à permettre une sélection plus rapide des textes lors d'une recherche documentaire.

L'adresse du dossier enregistrement du fichier INVERSE est le paramètre : PR% (23)

4. VOLUME OCCUPÉ PAR LES FICHIERS PENDANT LA PHASE EXPÉRIMENTAL

310 mots-clés et 843 textes.

Fichiers de données concernant le thesaurus : 459 secteurs

Fichier	Nombre de secteurs	Nombre d'octets
DICO	83	20 916
THESUS 1	313	78 876
THESUS 2	63	15 876
Total :	459	115 668

Fichiers de données concernant les textes : 2 689 secteurs

Fichier	Nombre de secteurs	Nombre d'octets
TEXTE	848	213 696
ANALYSE	1 688	425 376
MODES	153	38 556
Total :	2 689	677 628

Fichiers de travail : 298 secteurs, avec ANOMALIE : 371 secteurs

Fichier	Nombre de secteurs	Nombre d'octets
PARAM	3	756
PILE	3	756
PILTEX	3	756
TABLEAU	8	2 016
DICOTRIE	48	12 096
INVERSE	233	58 716
ANOMALIE	73	18 396
Total :	371	93 492

TOTAL GENERAL : 3 519 secteurs, soit 886 788 octets

Conclusion

Pour 843 textes et 310 mots-clés, il faut donc compter 3 519 secteurs, soit 886 788 octets.

Il faudrait doubler ces chiffres, si l'on tient compte des sauvegardes de ces fichiers faites sur le même disque.

Volume moyen par mot-clé

Pour 310 mots-clés il a fallu : 459 secteurs + fichiers de données
135 secteurs + fichiers de travail
 594 secteurs

soit 149 688 octets pour 310 mots-clés,
soit en moyenne 483 octets par mot-clé,
 soit environ 1/2 Ko par mot-clé (le double en comptant les sauvegardes).

Volume moyen par texte

Pour 843 textes, il a fallu : 2 689 secteurs + fichiers de données
239 secteurs + fichiers de travail
 2 928 secteurs

soit 737 856 octets pour 843 textes,
soit en moyenne 875 octets par texte,
 soit 1 Ko, et le double si l'on compte les sauvegardes.

Même en comptant les sauvegardes, on arrive au maximum à 2 millions d'octets pour 843 textes et 310 mots-clés.

avec 1 500 mots-clés ↔ 750 Ko ↔ 0,75 millions d'octets	}	6,75 millions d'octets sans les sauvegardes ni les programmes
et 6 000 textes ↔ 6 000 Ko ↔ 6 millions d'octets		



CHAPITRE 5

LES FONCTIONS DU LOGICIEL

Sommaire du chapitre 5

1. Remarques générales sur les programmes
 - 1.1. Les sous-programmes communs
 - 1.2. Les commandes
 - 1.3. Les écrans
 - 1.3.1. Les écrans de saisie
 - 1.3.2. Les écrans de réponse et de diagnostic
 - 1.3.3. Les écrans spécifiques
 - 1.3.4. Les écrans correspondant aux menus
2. Etude des principaux programmes
 - 2.1. Les programmes de création de la base documentaire
 - 2.2. Contrôle et réorganisation du thesaurus
 - 2.3. Gestion du thesaurus : GESTDICO
 - 2.3.1. Remarques générales sur ce programme
 - 2.3.2. Partie Création
 - 2.3.3. Partie Suppression
 - 2.3.4. Partie Modification
 - 2.4. Gestion du fonds documentaire : INTEX, GESTEX1, GESTEX2
 - 2.4.1. Programme création d'une fiche documentaire (INTEX)
 - 2.4.2. Programme modification d'une fiche documentaire (GESTEX1, GESTEX2)
 - 2.4.3. Partie suppression
 - 2.5. Consultation de la base
 - 2.5.1. Consultation du thesaurus : CONSUL 51
 - 2.5.2. Consultation des textes ou recherche : CONSUL 7
 - 2.5.3. Système de recherche (avec filtres) (source ICARE)

1. REMARQUES GÉNÉRALES SUR LES PROGRAMMES

1.1. Les sous-programmes communs

Signalons deux programmes importants qu'on retrouve dans tous les programmes :

1.1.1. Sous-programme de recherche d'un mot-clé dans le fichier DICO

Ce sous-programme est identique dans tous les programmes du projet. La méthode d'accès séquentielle indexée est décrite dans le chapitre "description technique de la base de donnée". Cette méthode utilise le fichier TABLEAU écrit en mémoire au début de chaque programme, de manière à permettre un accès plus rapide.

Pour plus de détails sur ce programme, se reporter au paragraphe qui lui est consacré en annexe.

Remarque : seul le programme de création du thesaurus ne l'utilise pas.

1.1.2. Gestionnaire de l'écran

Pour saisir, consulter, modifier les fiches thesaurus ou les fiches documentaires, on utilise un gestionnaire d'écran GESECRAN. Ce programme permet de modifier les différentes rubriques des fiches, par simple déplacement du curseur ; il effectue aussi un certain nombre de contrôles (type d'information, contrôle de fourchette, appartenance à une liste de valeurs, nombres maximal et minimal de caractères à saisir, etc ..) ; ces facilités offertes à l'utilisateur le rapprochent d'un logiciel de traitement de texte. Cependant, la place mémoire qu'il occupe (6 K) ne nous a pas permis de l'utiliser dans le programme de consultation du fonds documentaire. Il a aussi un autre inconvénient : il ne permet pas des contrôles supplémentaires zone par zone. Il faut que la saisie (modification) globale de la fiche soit terminée pour que les contrôles propres à l'application puissent s'effectuer.

Pour plus de détails sur ce programme, se reporter au paragraphe qui lui est consacré en annexe.

1.2. Les commandes

Nous avons essayé dans tous les programmes d'utiliser un certain nombre de commandes en leur donnant à chaque fois la même signification ; c'est ainsi que nous avons les commandes suivantes :

MENU :

qui permet le retour au MENU et l'abandon de la fonction en cours. Cette commande peut être tapée en réponse à une question posée par le système ; elle peut également être tapée à la place du mot-clé de référence, ou du numéro de fiche pour permettre l'abandon à n'importe quel moment du travail en cours.

RETOUR :

renvoie au travail précédent ou à la page écran précédente.

- . C'est ainsi que dans les programmes de gestion de la base, RETOUR permet la correction de la fiche courante ;
- . et dans le programme de consultation, RETOUR permet la visualisation de la fiche précédente.

SUITE :

signifie suite du travail ou passage à la fiche suivante.

- . C'est ainsi que dans les programmes de gestion,
 - SUITE est employée pour valider la fiche à l'écran et permettre l'insertion dans les fichiers,
 - SUITE est employée aussi pour faire une nouvelle création (ou modification, ou suppression),
 dans ces programmes, SUITE signifie "suite du travail" ;
- . dans les programmes de consultation, SUITE permet la visualisation de la fiche suivante.

GUIDE :

consultation du thesaurus.

RECHERCHE :

consultation du fonds documentaire, recherche à partir des mots-clés.

ANALYSE :

pour consulter ou modifier le bordereau n° 1 d'une fiche documentaire.

DESCRIPTEUR :

pour consulter ou modifier le bordereau n° 2 d'une fiche documentaire.

SAUF :

uniquement employée dans le programme de consultation pour diminuer la cible en utilisant cet opérateur associé à un mot-clé.

CHAMPS :

uniquement employée dans le programme de consultation, pour préciser la cible à l'aide des critères suivants : date des textes, date des supports, nature juridique.

TEXTE :

dans le programme de consultation pour visualiser les textes sélectionnés.

ACCUEIL :

retour à la page d'accueil de CATHI, c'est-à-dire au menu général du logiciel utilisé en période habituelle (deuxième module).

Certaines commandes peuvent figurer dans un menu ; dans ce cas, il faut taper un numéro correspondant à la commande, mais en général l'utilisateur doit taper les trois premières lettres de la commande.

Exemple : MEN pour MENU.

1.3. Les écrans

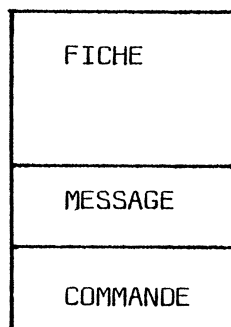
Le dialogue entre l'utilisateur et la machine se fait à l'aide de "pages écran" de 24 lignes de 40 caractères, de manière à correspondre à l'écran de visualisation du terminal Minitel qui sera utilisé dans la version vidéotex.

Afin de ne pas dérouter l'utilisateur par des dessins d'écran trop différents, nous avons tenu à ce qu'ils s'approchent de quelques formes standard en nombre limité.

C'est ainsi que nous pouvons distinguer :

1.3.1. Les écrans de saisie

Les écrans de saisie (consultation, modification) des fiches thesaurus ou documentaires se présentent de la façon suivante :



La partie FICHE contient la fiche thesaurus ou les bordereaux 1 ou 2 d'une fiche documentaire.

La partie MESSAGE contient un diagnostic d'erreur ou une information transmise à l'utilisateur.

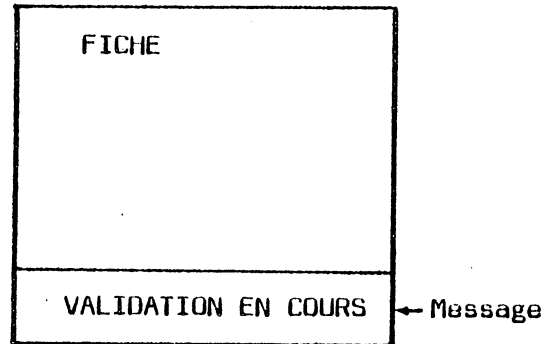
La partie COMMANDE contient les différentes commandes proposées à l'utilisateur et une zone réponse où il doit taper celle qu'il a choisie.

Il arrive que l'une des deux parties MESSAGE ou COMMANDE soit vide.
En voici deux exemples :

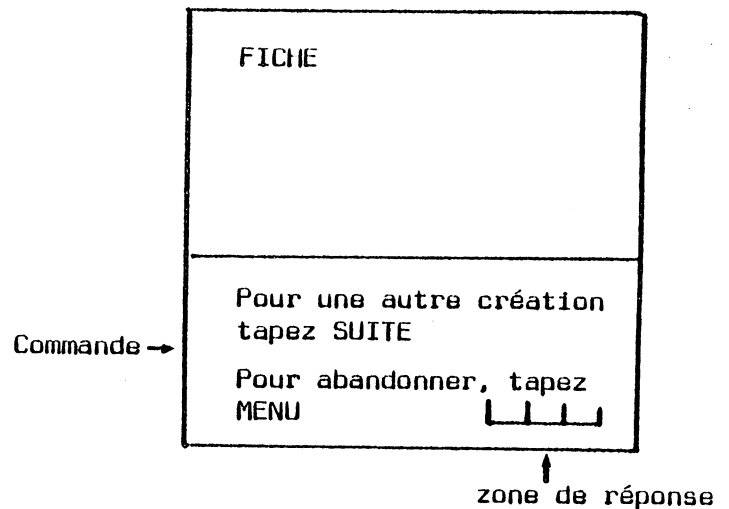
Partie commande absente :

Une fois la fiche à l'écran contrôlée et validée par l'utilisateur, cet écran apparaît pendant l'insertion dans les fichiers qui dure quelques secondes.

L'écran qui succède est le suivant :



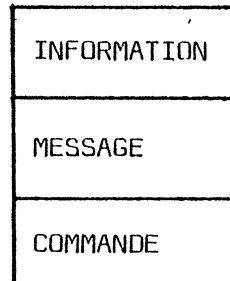
Partie message absente :



Ces écrans de saisie (modification, consultation) des fiches thesaurus ou documentaires se retrouvent dans les programmes de création et de gestion de la base, mais également dans les programmes de consultation. Voir par exemple les écrans correspondant à la partie TEXTE dans le programme CONSUL7.

1.3.2. Les écrans de réponse ou de diagnostic (dans le programme de consultation de la base CONSUL7).

Ce sont les écrans que l'utilisateur reçoit après le calcul de la cible en réponse à sa question. Ces écrans se présentent ainsi (cf. partie "système de recherche" dans CONSUL7) :



Dans la partie INFORMATION sont transmis à l'utilisateur un certain nombre de renseignements sur la base.

Exemple :

1. STAGIAIRE
MOT INCONNU
2. VOIRIE
INDEXE 5 TEXTES

ou

1. STAGIAIRE
INDEXE 1 TEXTE
2. VOIRIE
INDEXE 5 TEXTES

ou encore

8 TEXTES ont été obtenus avant
le filtre : AUXILIAIRE
qui indexe 5 textes.

Dans la partie MESSAGE, est transmis le résultat de la sélection, c'est en quelque sorte la réponse du système.

Exemple :

- RECHERCHE STOPPEE car inexistence du mot-clé n° 1 dans le thesaurus
- ou AUCUN TEXTE N A ETE SELECTIONNE. Vous pouvez obtenir des textes en donnant moins de mots-clés.
- ou 5 TEXTES ONT ETE SELECTIONNES à partir de ce(s) critère(s).

La partie COMMANDE contient les différentes commandes proposées et une zone réponse où l'utilisateur doit taper ce qu'il choisit.

Par rapport aux "écrans de saisie des fiches", ces "écrans de diagnostic" ne sont pas fondamentalement différents ; en effet, les fiches thesaurus et documentaires fournissent bien des informations sur la base.

1.3.3. Ecrans spécifiques

Les écrans de saisie des mots-clés lors d'une recherche, ou des choix de critères (nat. jur., date, SAUF) sont des écrans bien spécifiques. On en distingue 3 :

- 1) Saisie des mots-clés de la question (cf. Système de Recherche CONSUL7) :

```

POSER VOTRE QUESTION
à l' de de vos mots-clés
(taper un mot-clé et envoyez le)
? STAGIAIRE
ET
? VOIRIE
?
  
```

- 2) Saisie du mot-clé servant de filtre (cf. Système de Recherche CONSUL7) :

```

Vous avez sélectionné 8 textes avec
STAGIAIRE
Si vous voulez en ELIMINER
TAPEZ un mot-clé
sauf ? AUXILIAIRE
  
```

- 3) Saisie des critères de choix (cf. Système de Recherche CONSUL7) :

```

Préciser
(voir système de recherche, § 2.5.3.
filtre CHAMP)
  
```

1.3.4. Les écrans correspondant aux menus

```

*****
* TITRE DE LA PARTIE *
* *****
1. 1ère fonction
2. 2ème fonction
3. 3ème fonction
.
.
.
-----
TAPER OPTION CHOISIE : 

```

Exemple :

```

*****
* MISE A JOUR DU THESAURUS *
* *****
1. création d'un nouveau MOT-CLE
2. MODIFICATION d'un MOT-CLE
3. SUPPRESSION d'un MOT-CLE
4. ACCUEIL
-----
TAPER OPTION CHOISIE : 

```

Tous les écrans représentant un MENU se présentent sous cette forme.

2. ÉTUDE DES PRINCIPAUX PROGRAMMES

Nous nous bornerons à étudier les programmes les plus importants.

En ce qui concerne les *programmes de création de la base*, nous ne donnons qu'une représentation schématique, en effet, les fonctions qu'ils assurent, saisie et consultation des fiches thesaurus et documentaires, seront étudiées en détail dans les programmes de maintenance (gestion du thesaurus et du fonds documentaire).

Nous étudierons plus en détail les programmes suivants :

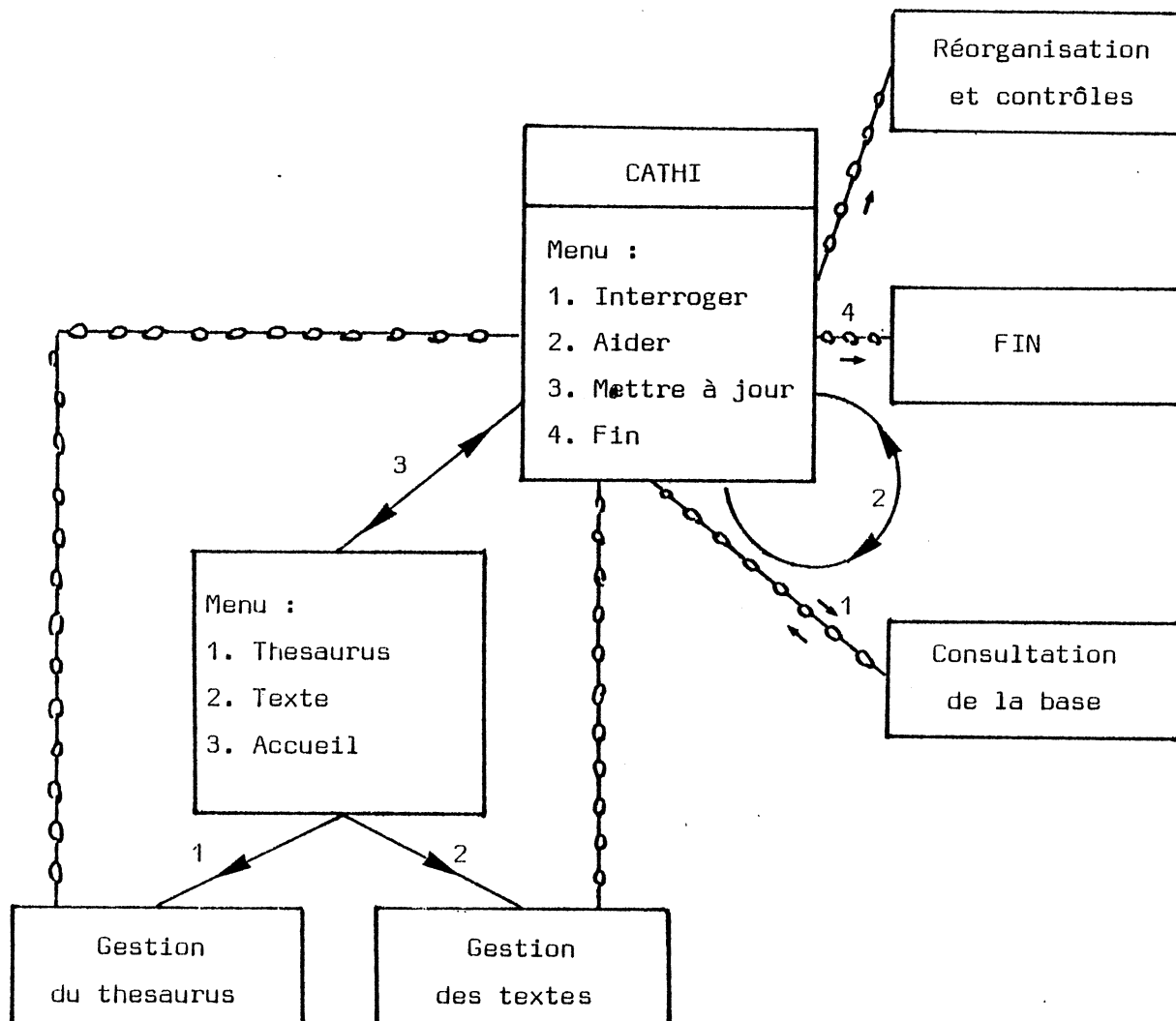
- . *contrôles et réorganisation du thesaurus,*
- . *gestion du thesaurus,*
- . *gestion des textes (fonds documentaire),*
- . *consultation de la base (thesaurus et fonds documentaire).*

On accède à chacune de ces quatre parties à partir de la page d'accueil du logiciel utilisé en période habituelle. Une représentation d'ensemble de ce logiciel est donnée ci-après.

Remarques :

- . Les programmes de réorganisation et de contrôle sont accessibles à partir de cette page d'accueil, bien que ne figurant pas sur le menu.
- . Un mot de passe est demandé avant d'entrer dans les programmes de gestion et de réorganisation de la base.
- . La partie FIN assure la sauvegarde automatique des fichiers de données.
- . La partie AIDE n'a pas été réalisée ; elle présentera une page écran contenant les explications sur le fonctionnement du système documentaire.

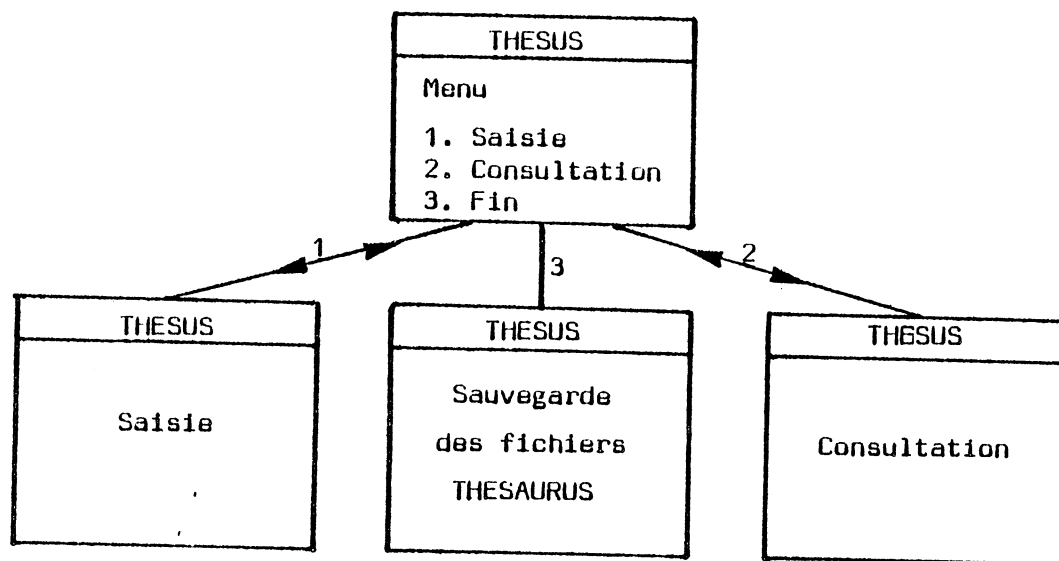
Organigramme général du 2ème module



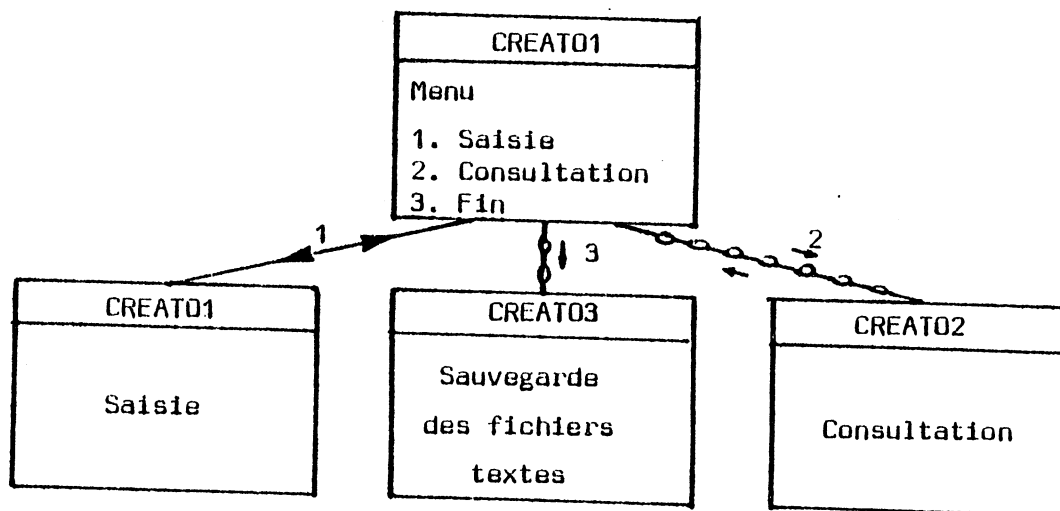
—○— chainage des programmes

2.1. Les programmes de création de la base documentaire

2.1.1. Création du thesaurus : THESUS



2.1.2. Création de textes : CREAT01, CREAT02, CREAT03



On lance par CREAT01 et on sort par CREAT03.
On peut faire des aller-retours CREAT01 ↔ CREAT02.

2.2. Contrôle et réorganisation du thesaurus

Cette partie a trois fonctions essentielles :

- . *Permettre un accès rapide aux enregistrements des mots-clés*
par : le tri du fichier DICO
la création du fichier TABLEAU.
- . *Mettre à jour les fichiers de données du thesaurus*
en éliminant les mots-clés qui ont été mis hors service, à condition que ceux-ci n'indexent aucun texte et ne figurent pas comme mots-clés liés dans une fiche thesaurus,
et en recalculant :
 - les numéros d'ordre dont les valeurs ont pu être modifiées par le tri,
 - les adresses des mots-clés principaux dans le cas des synonymes, en effet, la place des mots-clés principaux, donc leur adresse dans DICO, ont pu être modifiées par le tri ;
 - le nombre de liaisons *vers* chaque mot-clé de référence.
- . *Effectuer un certain nombre de contrôles de cohérence interne au thesaurus*
on vérifie :
 - que tous les mots-clés liés ont bien été créés,
 - qu'aucun mot-clé synonyme ne figure comme mot-clé lié sur une autre fiche thesaurus,
 - qu'un mot-clé descripteur a bien un type = 0,
 - qu'un mot-clé mis hors service n'indexe aucun texte et ne figure pas comme mot-clé lié dans une autre fiche thesaurus.

Toutes les incohérences donnent lieu à des messages d'erreurs, stockés dans la fiche ANOMALIE éditée à la fin du programme.

Par ailleurs, on édite la liste complète des mots-clés triés par ordre alphabétique avec leurs nouveaux numéros d'ordre.

Cette partie se compose de deux programmes s'enchaînant :

- . le programme TRI (voir explications et organigramme en annexe),
- . le programme ORGADICO (voir explications et organigramme en annexe)

Voir également en annexe des exemples d'anomalies.

L'édition du lexique des mots-clés existe sous deux formes :

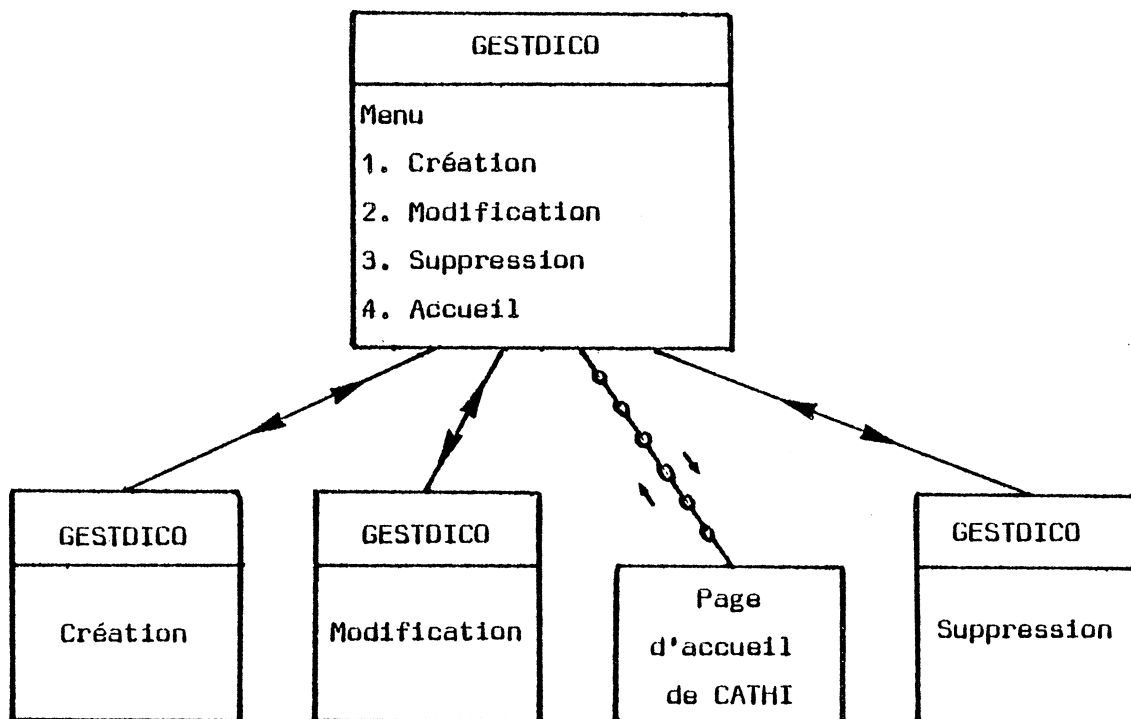
1) *forme simple*

mots-clés + numéros d'ordre ;

2) *forme complète*

chaque mot-clé est édité avec tous les renseignements figurant sur sa fiche thesaurus, ainsi qu'avec des renseignements figurant dans les fichiers. Exemple : nombre de mots-clés liés, nombre de liaisons, nombre de textes indexés par le mot-clé, etc ... On réalise ainsi l'édition complète des fichiers concernant le thesaurus (DICO, THESUS 1, THESUS 2).

2.3. Gestion du thesaurus : GESTDICO



2.3.1. Remarques générales sur GESTDICO

Les trois parties (création, suppression, modification) ont en commun de nombreux sous-programmes, ce qui a permis de les faire résider en mémoire en même temps, et le retour à MENU permet de passer de l'une à l'autre des trois parties.

Volume : 58 secteurs en version compilée, soit 14,5 K.

Sous-programmes communs :

. Recherche d'un mot-clé dans DICO

commun à tous les programmes du projet, il utilise le fichier TABLEAU écrit en mémoire au début de GESTDICO (voir annexe).

. Contrôles de la fiche thesaurus

ces contrôles sont communs aux parties CREATION et MODIFICATION.

On en distingue cinq : les quatre premiers font l'objet d'un message en cas d'erreur. L'utilisateur doit corriger son erreur (en tapant RETOUR) ou abandonner (en tapant MENU). En ce qui concerne l'existence des mots-clés liés, nous avons vu que le rejet de la fiche en cas d'inexistence d'un mot-clé lié ne pouvait pas être imposé (en particulier à cause des relations hiérarchiques), dans ce cas et dans celui de la présence d'un mot-clé lié mis hors service, un point d'interrogation s'affiche à côté du code liaison signalant l'anomalie.

Liste des contrôles :

- 1) présence obligatoire du type,
- 2) type = 1 pour un mot-clé synonyme,
- 3) un mot-clé descripteur de texte ne peut devenir un synonyme,
- 4) un seul mot-clé lié permis pour un mot-clé synonyme,
- 5) existence et non "mise hors service" des mots-clés liés.

Si la fiche est correcte, l'utilisateur a le choix entre :

- la validation de la fiche (SUITE)
- la correction de la fiche (RETOUR)
- l'abandon du travail (MENU).

Dans ce sous-programme de contrôle de la fiche thesaurus,

- . on calcule le nombre de mots-clés liés en les tassant si besoin : une liaison est acceptée s'il y a présence du code liaison et du mot-clé lié, sinon elle est annulée ;
- . l'adresse du mot-clé principal est mise à jour.

. *Gestionnaire d'écran* (voir annexe)

Commun à tous les programmes du projet, il permet la saisie contrôlée des différentes zones de la page écran. Le masque affiché (dans GESTDICO) est celui d'une fiche thesaurus vierge.

Le fichier descripteur des différentes zones est le fichier DESDICO (voir annexe).

Les contrôles faits par le gestionnaire d'écran et rassemblés dans DESDICO sont :

- présence obligatoire du mot-clé de référence,
- longueur maximale de chaque mot-clé = 38 octets,
- type = 0 ou 1,
- longueur maximale de chaque ligne de la note d'application = 38 octets,
- appartenance de chaque code liaison à la liste des valeurs {1, 2, 3, 4}

En cas d'erreur, un "bip bip" sonore est émis jusqu'à ce que l'utilisateur corrige.

(1)

2.3.2. Partie CREATION de GESTDICO

Trois cas peuvent se présenter :

- . l'utilisateur veut créer un mot-clé dont la fiche n'a jamais été saisie, dans ce cas c'est une véritable création ;
- . le mot-clé de référence existe déjà (fiche saisie), mais il a été mis hors service, dans ce cas l'utilisateur peut le recréer ;
- . le mot-clé de référence existe déjà et il n'a pas été mis hors service, l'utilisateur est prévenu par un message que le mot-clé existe déjà et la fiche est rejetée.

(1) Les fichiers susceptibles d'être modifiés dans GESTDICO sont : DICO, THESUS1, THESUS2, DARAM, PILE. Lors de la création de nouveaux enregistrements dans THESUS1 et THESUS2, on utilise d'abord les enregistrements vacants dans ces fichiers (cf. Annexe).

La fiche saisie est à la fois contrôlée par le gestionnaire d'écran et par le sous-programme de contrôle.

Le langage interactif utilise 3 commandes :

- . la commande MENU permet l'abandon de la saisie et le retour au menu ; cette commande peut être tapée à la place du mot-clé de référence pour permettre l'abandon à n'importe quel moment ;
- . la commande SUITE permet
 - la validation, après saisie et contrôle, de la fiche,
 - une nouvelle création, après l'insertion dans les fichiers et l'édition de la fiche ;
- . la commande RETOUR permet la correction de la fiche.

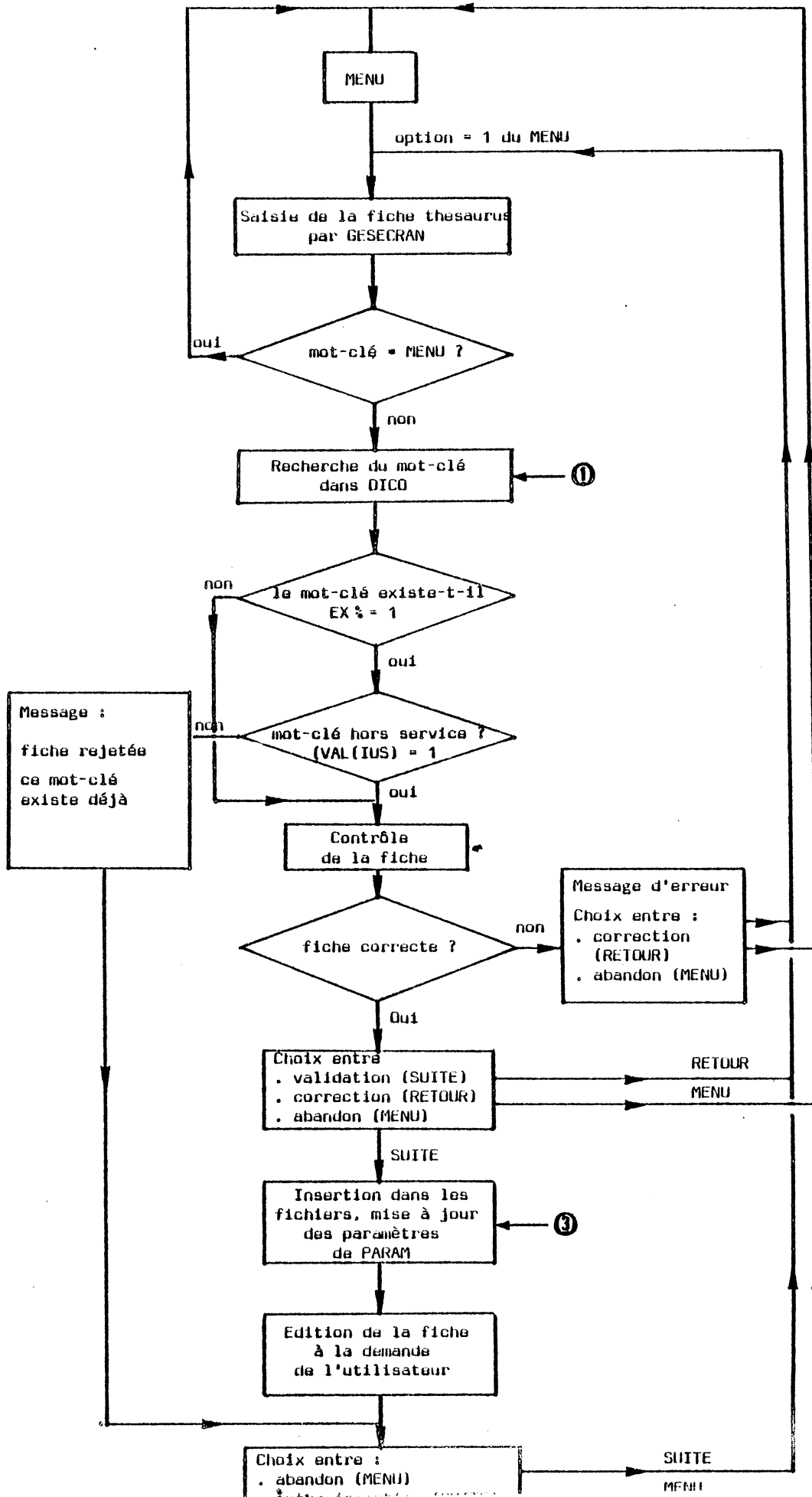
Dans le cas d'une recreation d'un mot-clé mis hors service, on récupère un certain nombre de variables :

- . le numéro d'ordre,
- . le nombre de textes indexés,
- . l'adresse dans le fichier inverse,
- . le nombre de liaisons vers le mot-clé recréé,
- . l'adresse de chaînage dans THESUS 1,
- . le booléen indiquant si le mot-clé de référence est un mot-clé principal ou non ; le booléen "mis hors service" est réinitialisé, les autres champs sont modifiés en fonction de la nouvelle saisie.

(Voir annexe).

① Voir le sous-programme et son organigramme : RECHERCHE d'un mot-clé dans DICO.

② Voir détail dans partie CREATION



2.3.3. Partie SUPPRESSION

Quand l'utilisateur désire supprimer un mot-clé, le programme indique le nombre de textes indexés par ce mot-clé et le nombre de mots-clés renvoyant sur lui (c'est-à-dire le nombre de fiches thesaurus où le mot-clé apparaît comme mot-clé lié). Le mot-clé est mis hors service, après confirmation de la volonté de suppression par l'utilisateur ; les enregistrements dans les fichiers ne sont pas effacés, la suppression effective du mot-clé n'aura lieu que lors de la réorganisation des fichiers sous certaines conditions (nombre de textes indexés = 0, nombre de liaisons vers le mot-clé = 0).

Le langage interactif utilise les deux commandes suivantes :

MENU

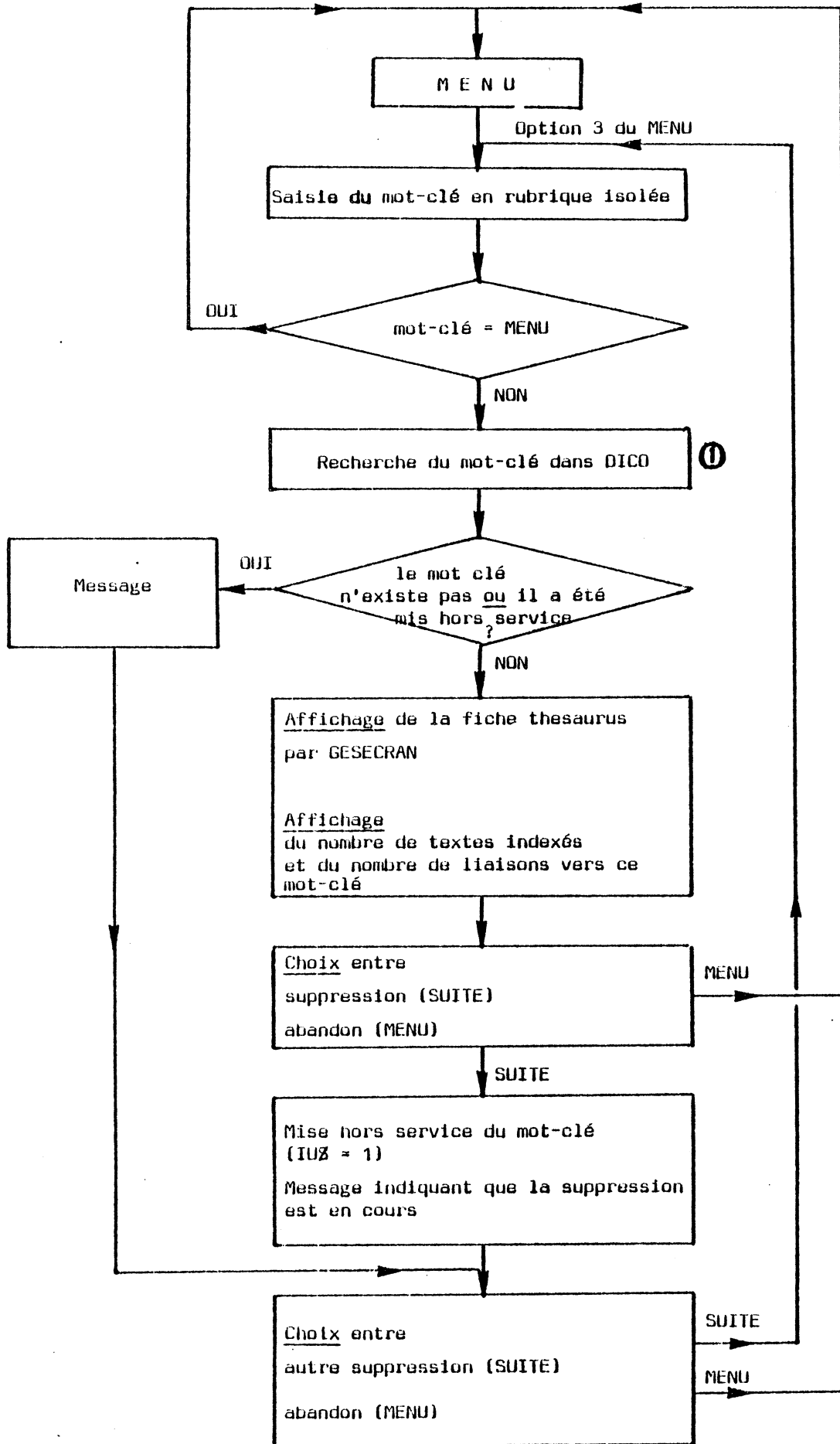
qui permet l'abandon de la fonction suppression et le retour au menu ; cette commande peut être tapée à la place du mot-clé pour permettre l'abandon à n'importe quel moment.

SUITE

qui permet la confirmation de la suppression, ou une autre suppression.

Voir détails en annexe.

Organigramme de la partie SUPPRESSION



① Voir le sous-programme RECHERCHE d'un mot-clé dans DICO

2.3.4. Partie MODIFICATION de GESTDICO

L'utilisateur ne peut modifier un mot-clé mis hors service.

Toutes les zones de la fiche thesaurus sont modifiables à l'exception du mot-clé de référence et du numéro d'ordre. L'utilisateur ne peut positionner le curseur sur l'une de ces deux zones.

Les contrôles effectués sont les mêmes que dans la partie CREATION.

Le langage interactif utilise les trois commandes :

MENU

pour abandon de la modification en cours et retour au menu ; cette commande peut être tapée à la place du mot-clé de référence, de manière à permettre l'abandon à n'importe quel moment.

SUITE

pour valider la fiche thesaurus modifiée, ou faire une autre modification.

RETOUR

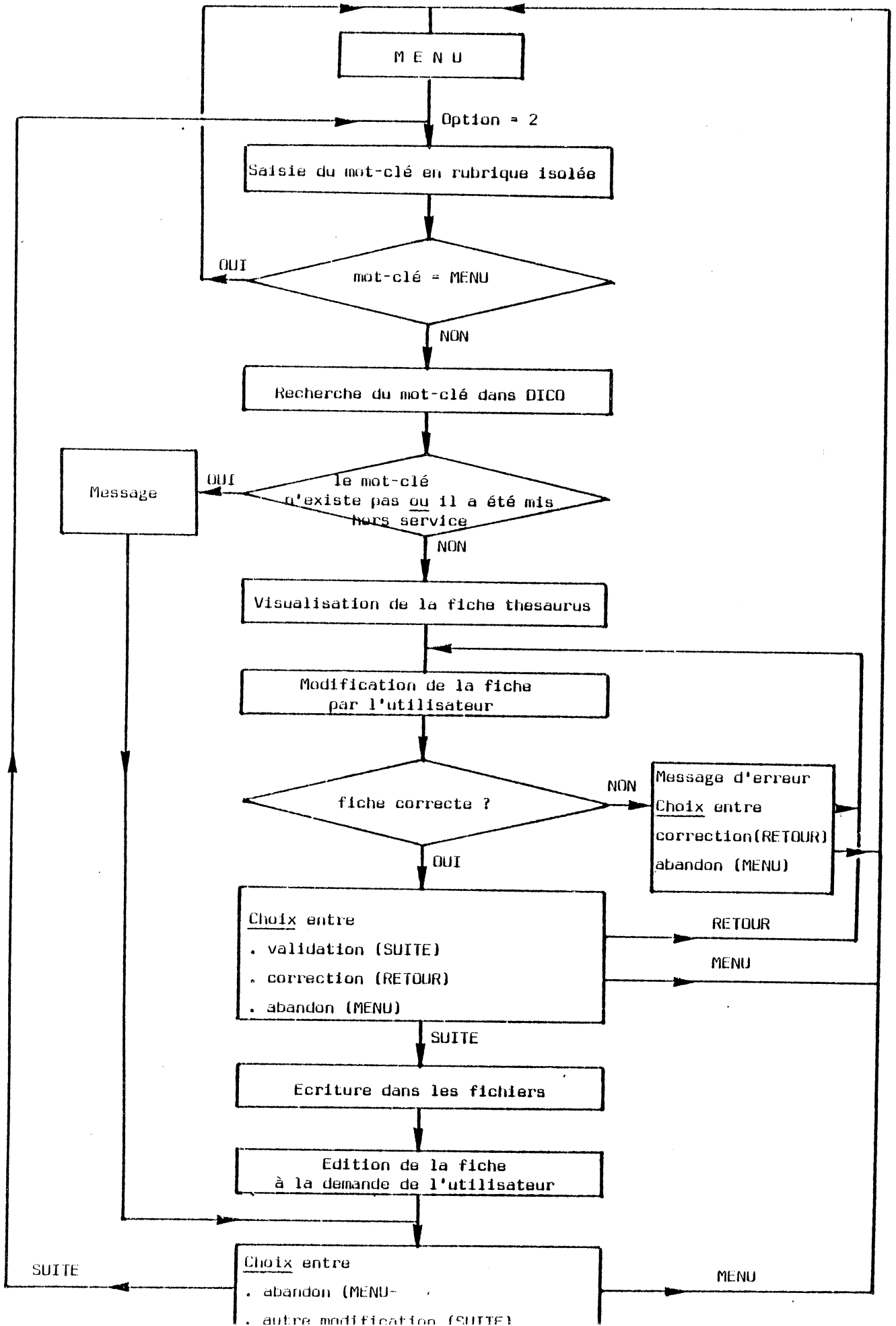
pour corriger la fiche.

Certaines zones dans les fichiers ne sont pas modifiées, il s'agit :

- . des numéros d'ordre
- . du mot-clé de référence
- . de l'adresse de chaînage dans THESUS1
- . de l'adresse de chaînage dans le fichier INVERSE
- . du nombre de textes indexés
- . du nombre de liaisons vers le mot-clé de référence
- . du booléen indiquant si le mot-clé de référence est un mot-clé principal ou non
- . du booléen indiquant si le mot-clé a été mis hors service.

Voir détails en annexe.

Organigramme de la partie MODIFICATION



2.4. Gestion du fonds documentaire : INTEX, GESTEX1, GESTEX2

La gestion du fonds documentaire assure les trois fonctions suivantes :

- . création d'un nouveau texte,
- . modification d'un texte,
- . suppression d'un texte.

A la sortie du programme de maintenance du fonds documentaire, on revient à la page d'accueil de CATHI.

La *partie création* est assurée par le programme INTEX ;

la *partie modification* est assurée par :

- . le programme GESTEX1 pour la modification du bordereau n° 1 d'une fiche documentaire,
 - . le programme GESTEX2 pour la modification du bordereau n° 2 d'une fiche documentaire
- GESTEX2 est chaîné à GESTEX1.

La *partie suppression* a été simplement étudiée, mais non programmée.

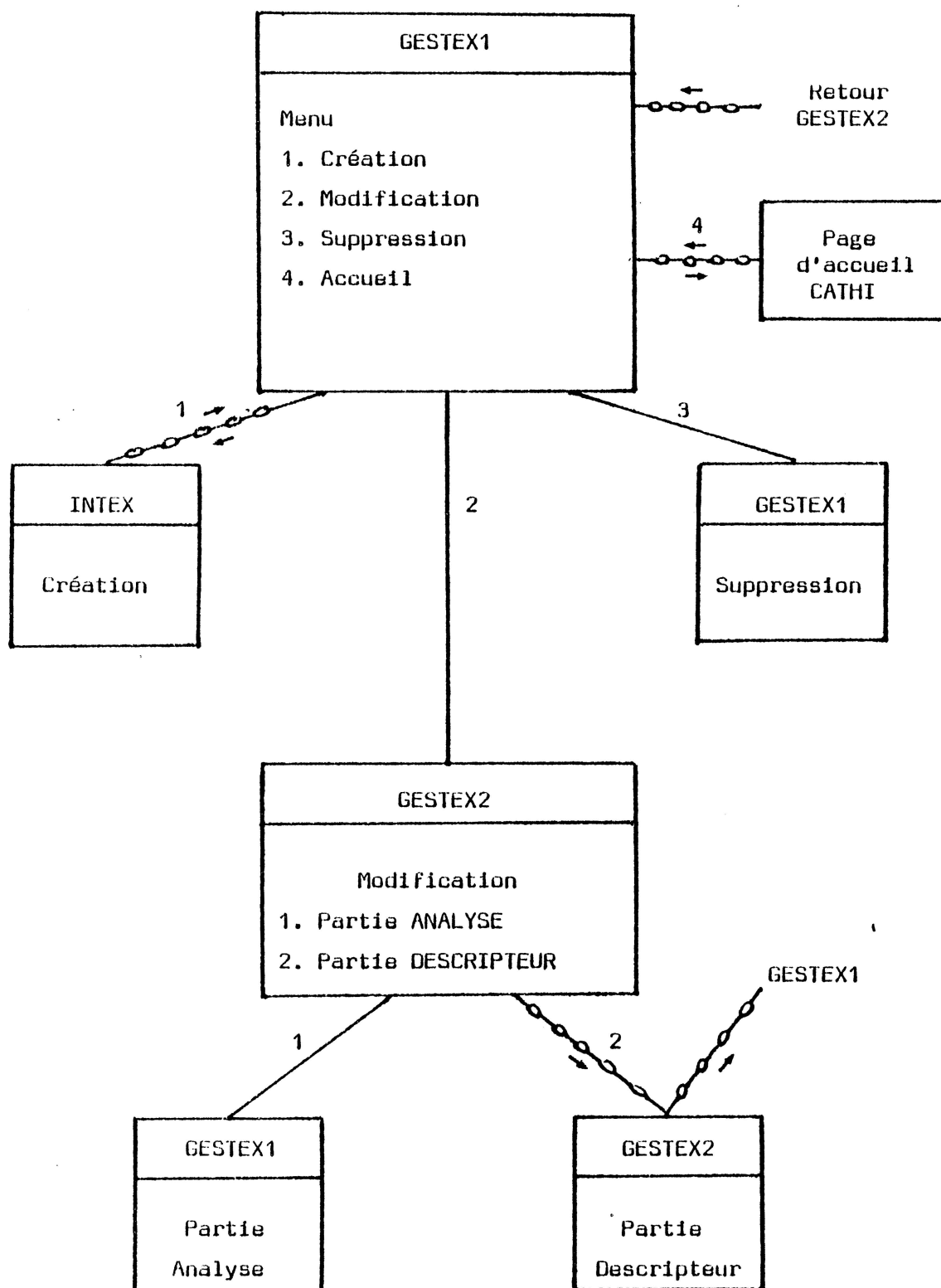
On passe d'une partie à l'autre par l'intermédiaire du MENU qui sert de plaque tournante.

Le *programme de recherche d'un mot-clé dans DICO* est le même que celui utilisé dans tous les autres programmes du projet (voir annexe).

Le *gestionnaire d'écran* est le programme GESECRAN commun à tous les programmes du projet (voir annexe). Le fichier descripteur des deux "pages écran" correspondant aux deux bordereaux de la fiche documentaire, est le fichier ECRAN qui comporte deux pages.

La mise à jour des fichiers se fait en temps réel, sauf en ce qui concerne la suppression effective d'une fiche documentaire (voir partie SUPPRESSION).

Gestion de textes : INTEX, GESTEX1, GESTEX2



2.4.1. Programme CREATION d'une fiche documentaire (INTEX)

Volume : 64 secteurs compilés, soit 16 Ko.

Chaque fiche documentaire est composée de deux bordereaux. *La saisie se fait en deux temps* : Saisie du bordereau 1, puis après la validation de cette première page, s'effectue la saisie du bordereau n° 2.

La fiche documentaire est contrôlée :

- . *par le gestionnaire d'écran* qui assure les contrôles classiques : présence obligatoire des zones : titre, date, numéro du support, appellation, source du droit, appartenance à une plage de valeurs pour la date et la source du droit (1, 2, 3), nombre minimal et maximal de caractères à saisir, etc ..
 - . *par le programme lui-même* qui vérifie pour chaque descripteur,

<ul style="list-style-type: none"> - son existence dans le lexique - le fait qu'il n'a pas été mis hors service - qu'il n'est pas synonyme 	}	contrôles de la deuxième page
---	---	-------------------------------
- à chaque erreur, un message et un signal sonore préviennent l'utilisateur qui doit corriger ou abandonner.

Le numéro de fiche des deux bordereaux est attribué automatiquement (ce numéro correspond au numéro d'ordre de l'enregistrement dans le fichier TEXTE).

Une fois faite la validation de la deuxième page, l'insertion dans les fichiers se fait dans l'ordre suivant :

- . fichier MODES si nécessaire (1 ou 2 enregistrements)
- . fichier ANALYSE
- . fichier TEXTE
- . fichier INVERSE
- . fichier DICO.

Les commandes utilisées sont :

MENU

pour l'abandon de la saisie et le retour au menu général du programme de gestion de fonds documentaire. MENU peut être tapé à la place du numéro de fiche pour permettre l'abandon à n'importe quel moment.

RETOUR

pour corriger la fiche documentaire en cours de saisie.

SUITE

pour valider la page saisie (1er ou 2ème bordereau),
pour faire une nouvelle création, après l'insertion dans les fichiers.

Voir en annexe pour plus de détails.

2.4.2. Programmes de MODIFICATION d'une fiche documentaire (GESTEX1, GESTEX2)

Première partie : modification du bordereau n° 1 (GESTEX1)

Volume : 47 secteurs compilés, soit 11,75 Ko.

L'accès au texte à modifier se fait en tapant le numéro de fiche. Ce numéro de fiche correspond d'ailleurs à l'adresse relative de l'enregistrement correspondant dans le fichier TEXTE.

Si le texte a été trouvé et s'il n'a pas été virtuellement supprimé (booléen de suppression = 0), alors le bordereau n° 1 s'affiche. L'utilisateur a alors le choix entre :

- . corriger le premier bordereau en tapant ANALYSE (ANA),
- . corriger le deuxième bordereau en tapant DESCRIPTEUR (DES), dans ce cas le programme GESTEX2 est chargé en mémoire pour être exécuté,
- . abandonner en tapant MENU (MEN).

Les contrôles effectués sont les mêmes que dans la partie création (voir INTEX).

La mise à jour dans les fichiers est faite immédiatement.

Les commandes sont :

MENU

pour abandonner la partie modification et revenir au menu général de gestion du fonds documentaire ; cette commande peut être tapée à la place du numéro de fiche pour permettre l'abandon à n'importe quel moment.

ANALYSE

pour modifier le bordereau n° 1 d'une fiche documentaire.

DESCRIPTEUR

pour modifier le bordereau n° 2 d'une fiche documentaire (chainage avec GESTEX2).

RETOUR

pour corriger le bordereau n° 1 de la fiche documentaire en cours de modification.

SUITE

pour valider ce bordereau.

Voir détails en annexe.

Deuxième partie : modification du bordereau n° 2 d'une fiche documentaire (GESTEX2)

Volume : 61 secteurs, soit un peu plus de 15 Ko.

On accède à cette partie en tapant la commande DESCRIPTEUR (DES) dans la première partie (GESTEX1) quand l'utilisateur veut modifier le deuxième bordereau d'un texte précis (qui a été trouvé).

Puisque ce deuxième bordereau ne contient que la liste des descripteurs, les contrôles ne concernent que ceux-ci.

L'utilisateur peut modifier, supprimer, tous les descripteurs précédents ; il peut également en rajouter d'autres.

Contrôles effectués sur les descripteurs :

- 1) Elimination des descripteurs en double (et tassement des descripteurs),
- 2) Rejet des descripteurs inexistants dans le lexique, ou qui ont été mis hors service, ou qui sont synonymes.

Chaque erreur donne lieu à un message et à un signal sonore.

Les fichiers sont mis à jour immédiatement.

Les commandes utilisées sont :

MENU

voir GESTEX1 ou INTEX.

RETOUR

pour corriger le bordereau n° 2 de la fiche documentaire en cours de modification.

SUITE

pour valider ce bordereau.

Après la modification effectuée, on revient directement au menu général du programme du fonds documentaire.

Voir détails en annexe.

2.4.3. Partie SUPPRESSION d'une fiche documentaire

Cette partie n'a pas été programmée ; en effet, elle semblait moins urgente pour l'application, dans la mesure où un texte de loi ne peut être supprimé, il peut simplement être abrogé ou rectifié par un autre. Cependant, il faut pouvoir disposer de cette fonction, ne serait-ce que pour récupérer de la place en supprimant les textes très anciens ou abrogés depuis longtemps.

La fonction "suppression d'une fiche documentaire" se fera en deux temps, comme c'est le cas pour la suppression d'un mot-clé du thesaurus :

- 1) en temps réel : aucune suppression n'a effectivement lieu, seul le booléen ST% prévu à cet effet prendra la valeur 1 ;
- 2) dans un deuxième temps, au cours d'une phase de réorganisation des fichiers, aura lieu la suppression effective, qui implique les suppressions suivantes :
 - a) l'enregistrement correspondant dans TEXTE
 les enregistrements correspondants dans ANALYSE
 le ou les enregistrements correspondants dans MODES (s'ils existent),
 avec récupération des adresses vacantes dans le fichier PILTEX
 - b) pour chaque descripteur de la fiche documentaire supprimée :
 - . dans le fichier INVERSE, effacer le numéro de la fiche documentaire,
 - . dans DICO :
 - décrémenter la variable NT% qui indique le nombre de textes indexés par le descripteur (NT% figure dans DICO)
 - si cette variable NT% a pour valeur 0, effacer l'adresse de chaînage dans le fichier INVERSE (AI%) (AI% figure dans DICO), dans ce cas, effacer l'enregistrement dans le fichier INVERSE et récupérer l'adresse vacante dans PILTEX
 - . tasser les adresses des textes dans le fichier INVERSE avec récupération des enregistrements vacants.

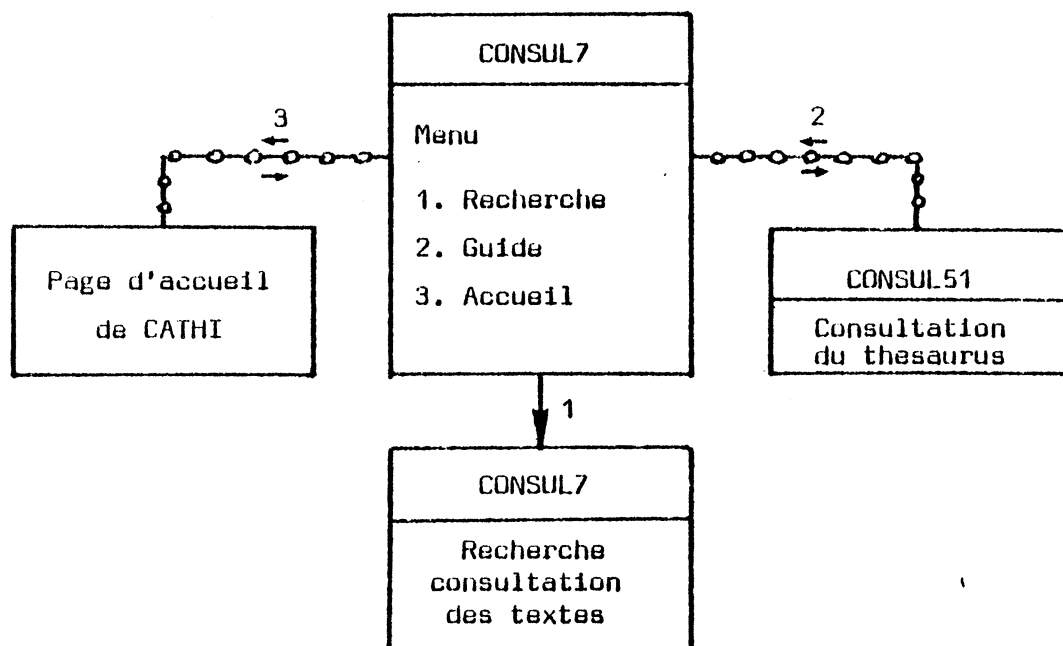
2.5. Consultation de la base

Il existe deux versions du programme de consultation :

- . dans la première version, la recherche de documents se fait uniquement à partir de mots-clés reliés par l'opérateur ET,
- . dans la deuxième version, cette recherche est complétée :
 - en utilisant l'opérateur SAUF sur un autre mot-clé,
 - ou/et en précisant la cible par une sélection sur la date et/ou sur la nature juridique.

Nous n'étudions ici que la deuxième version (CONSUL7) car elle englobe la première.

La présentation générale du logiciel CONSUL7 montre qu'à partir du menu, l'utilisateur a accès au programme de consultation du thesaurus CONSUL51. En cas de réponse non satisfaisante à partir d'un mot-clé, l'utilisateur devrait pouvoir reformuler sa question en utilisant le thesaurus. Si pour les documentalistes, la consultation du thesaurus est un "réflexe", ce n'est peut-être pas le cas pour les utilisateurs "non spécialistes". L'expérience nous montrera ce qu'il en est. Il n'est pas évident que les notions de synonymie, d'analogie et de hiérarchie, soient familières à la majorité des gens.



2.5.1. Consultation du thesaurus : CONSUL51

Volume : 17 secteurs compilés, soit 4,25 Ko.

La fonction de ce programme est la visualisation de la fiche thesaurus correspondant au mot-clé tapé par l'utilisateur.

Comme nous le montre l'organigramme général qui suit, il n'y a rien d'original dans ce programme :

- . le petit sous-programme de recherche du mot-clé dans DICO est le même que dans les autres programmes du projet,
- . par contre, la visualisation des fiches thesaurus est faite sans le gestionnaire d'écran GESECRAN, inutile dans ce programme puisque l'utilisateur n'a aucun travail à faire sur les fiches.

Les deux commandes sont :

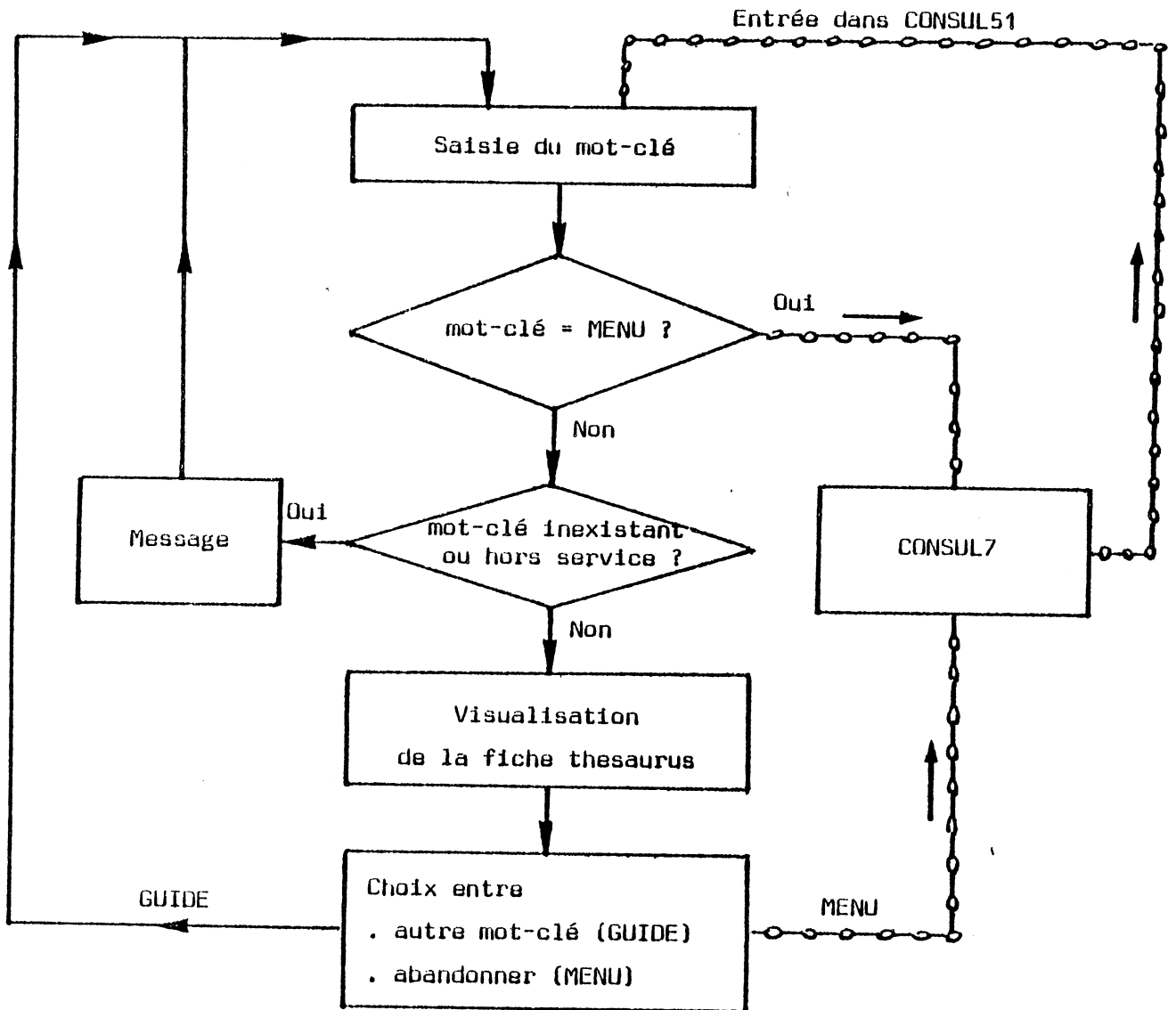
MENU

pour abandonner la consultation du thesaurus ; cette commande provoque le retour au programme plus général de consultation de la base CONSUL7.

GUIDE

pour visualiser la fiche thesaurus d'un autre mot-clé.

Organigramme de CONSUL51



2.5.2. Consultation des textes ou recherche : CONSUL7

Volume : 83 secteurs compilés, soit 21 Ko.

L'algorithme général de recherche de document se présente de la façon suivante :

recherche à partir d'une équation composée d'un ou plusieurs mots-clés reliés par l'opérateur ET.

Les résultats sont communiqués à l'utilisateur sur un écran diagnostic.

Deux cas se présentent :

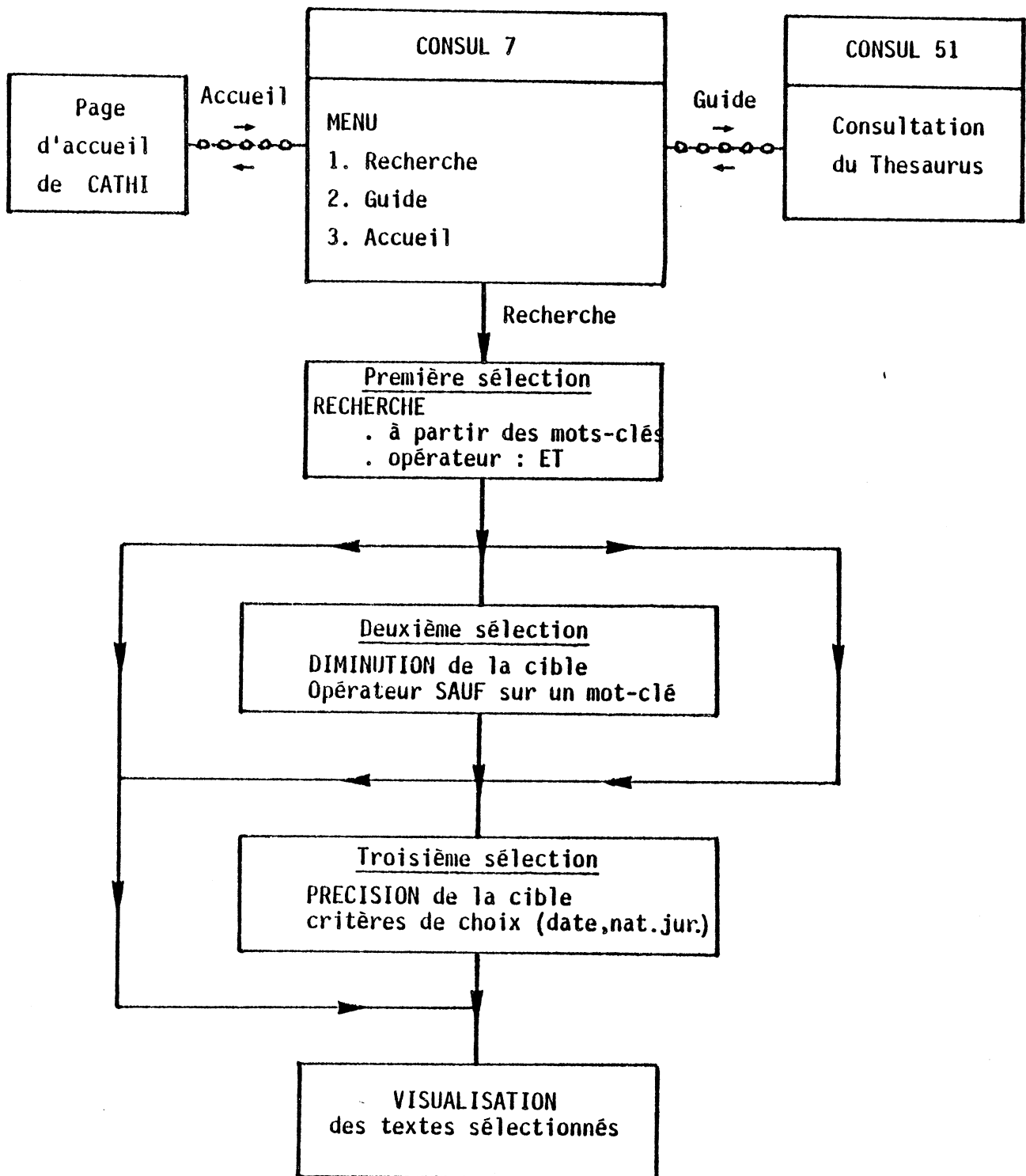
- . la recherche a été stopée pour différentes raisons, dans ce cas le retour automatique au MENU permettra éventuellement à l'utilisateur de recommencer avec d'autres mots-clés ;
- . la recherche a abouti, dans ce cas l'utilisateur a le choix entre :
 - voir les textes sélectionnés (commande TEXTE),
 - préciser la cible par une sélection sur la nature juridique et/ou la date (commande CHAMP),
 - diminuer la cible en utilisant l'opérateur SAUF sur un mot-clé (commande SAUF).

Remarque : l'utilisateur peut à tout moment abandonner la recherche en cours en tapant MENU.

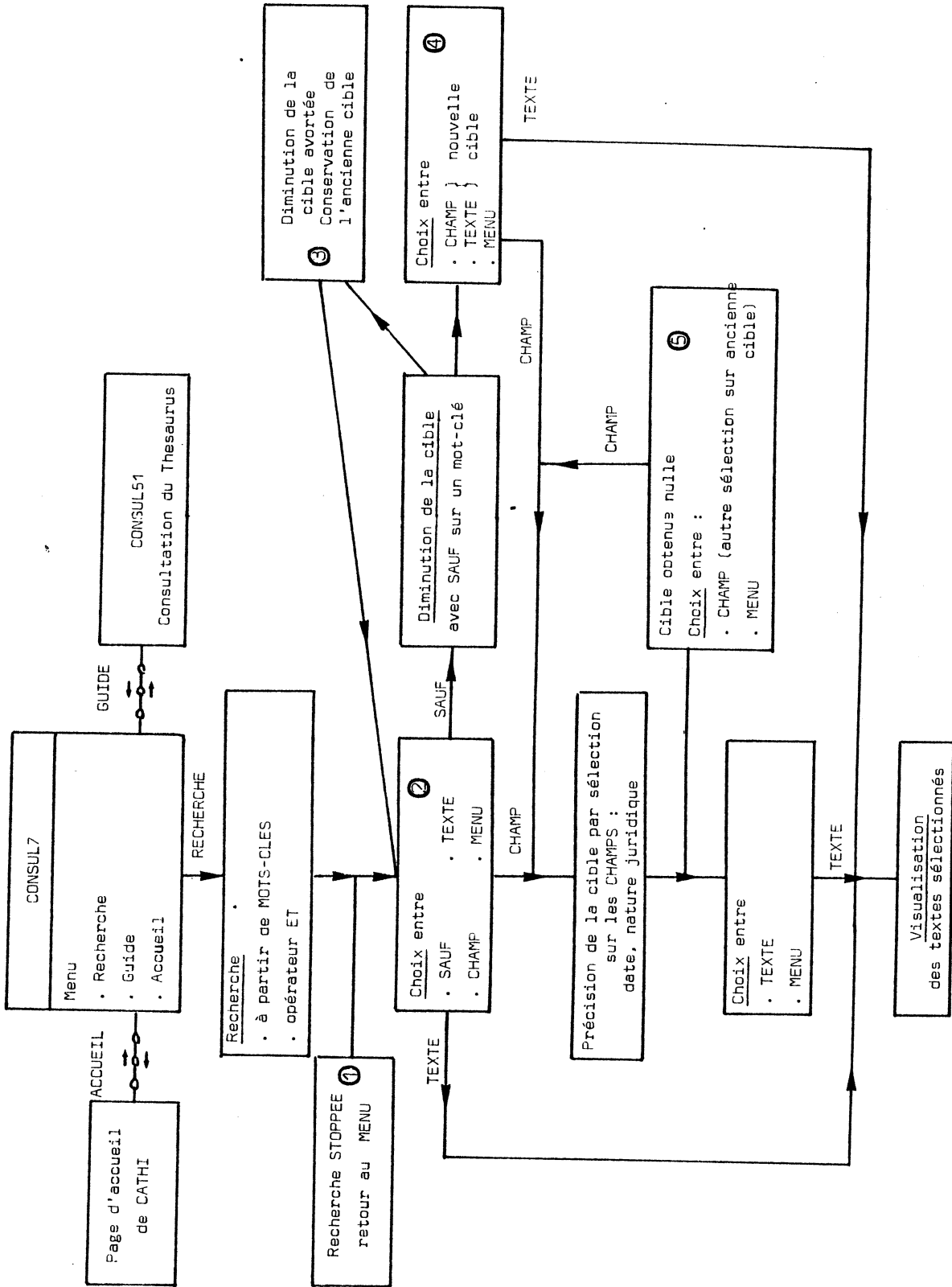
Pour le moment, le passage direct aux filtres n'est pas prévu. Cela pourrait se faire à condition d'introduire des fichiers INVERSE sur la date et la nature juridique. La création de ces fichiers ne poserait pas de problèmes

Par contre, il faudrait modifier en conséquence les programmes de création et de gestion des textes, de telle sorte que la mise à jour des fichiers INVERSE se fasse en temps réel.

ORGANIGRAMME DE CONSULTATION DU FONDS DOCUMENTAIRE



Organigramme du programme de CONSULTATION (CONSUL7)



① ② ③ ④ ⑤ Les numéros renvoient aux explications qui suivent
Les retours à MENU n'ont pas été tracés pour ne pas alourdir le schéma.

Commentaires sur l'organigramme de CONSUL7

- ① La recherche à partir des mots-clés a pu être stoppée pour différentes raisons :
- . aucun mot-clé n'a été rentré dans la question,
 - . au moins un des mots-clés de la question est inexistant dans le thesaurus ou a été mis hors service,
 - . un des mots-clés est un synonyme, mais la liaison avec son mot de référence (mot-clé principal) n'a pas été établie. Ce qui peut se produire dans deux cas :
 - le mot de référence n'a pas été rentré (fiche thesaurus non saisie)
 - ou bien le synonyme a été rentré avant le mot de référence, mais le programme de réorganisation n'a pas été lancé depuis ;
 - . un des mots-clés n'indexe aucun texte, ou est le synonyme d'un mot-clé qui n'indexe aucun texte. En effet, dès que la liaison synonyme → mot de référence a été établie, tous les textes indexés par un mot de référence sont supposés être indexés également par ses synonymes, mais ceci est transparent pour l'utilisateur ;
 - . aucun des cas précédents ne se présentent, mais la "cible résultat" est nulle.

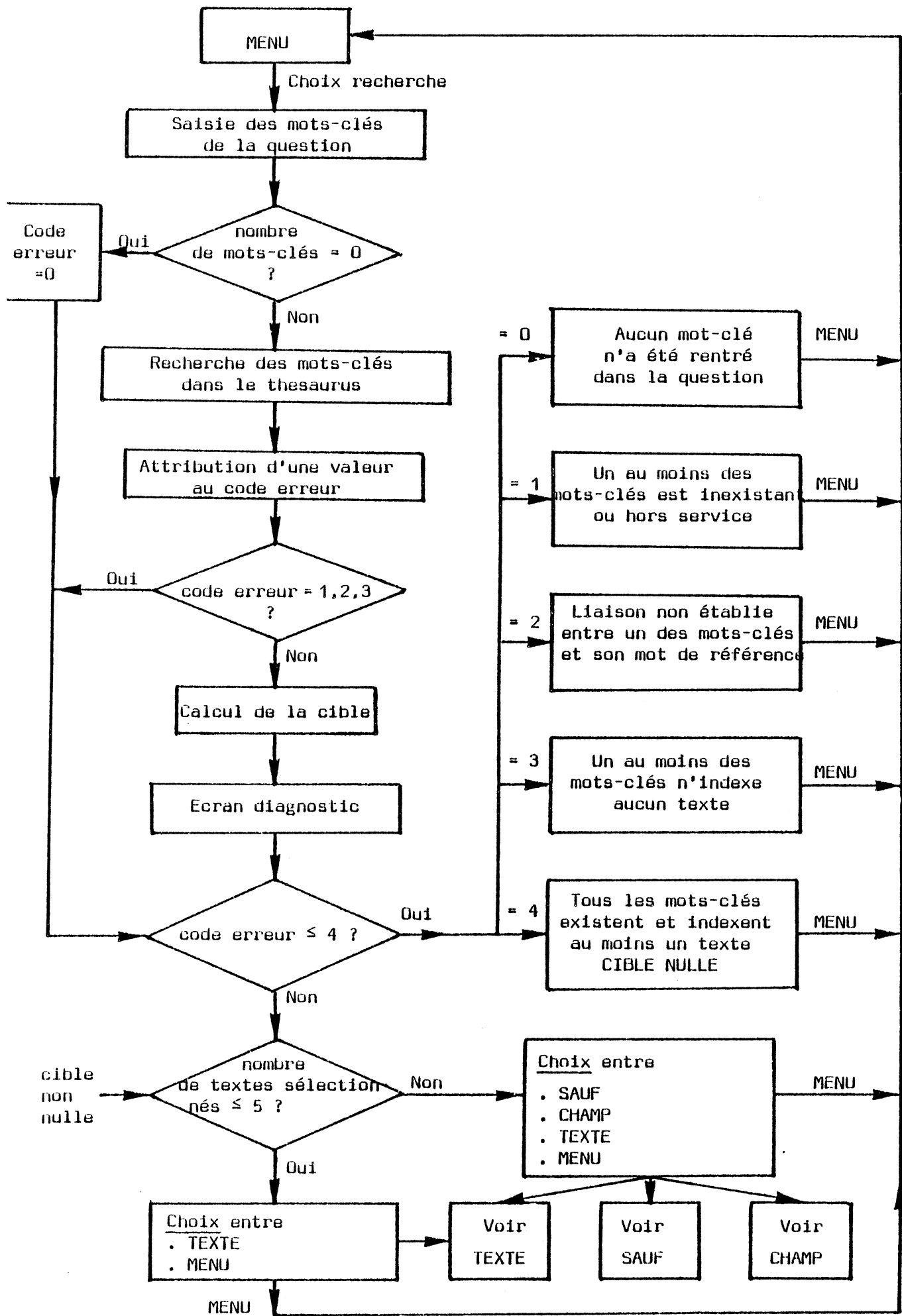
Dans tous ces cas, aucun texte n'a pu être sélectionné, c'est pourquoi le retour au menu est automatique. L'utilisateur pourra recommencer une nouvelle recherche.

- ② La recherche a abouti, la cible est non nulle :
- . si le nombre de textes sélectionnés est inférieur ou égal à 5, on ne propose pas à l'utilisateur de diminuer sa cible (avec SAUF ou les filtres) il peut simplement les visualiser (TEXTE) ou abandonner (MENU) ;
 - . dans le cas contraire (nombre de textes sélectionnés supérieur à 5), l'utilisateur a le choix entre les quatre commandes (SAUF, CHAMP, TEXTE, MENU)
 - . Si le nombre de textes sélectionnés est supérieur à 64, on ne retient que les 64 premiers pour la suite du travail.

- ③ La diminution de la cible avec l'opérateur SAUF a pu être stoppée ou ne pas être satisfaisante pour les raisons suivantes :
- . mot-clé inexistant dans le thesaurus ou hors service,
 - . mot-clé synonyme mais liaison avec le mot de référence non établie,
 - . la sélection n'a pas été modifiée par le filtre,
 - . ou bien tous les textes sélectionnés ont été éliminés.

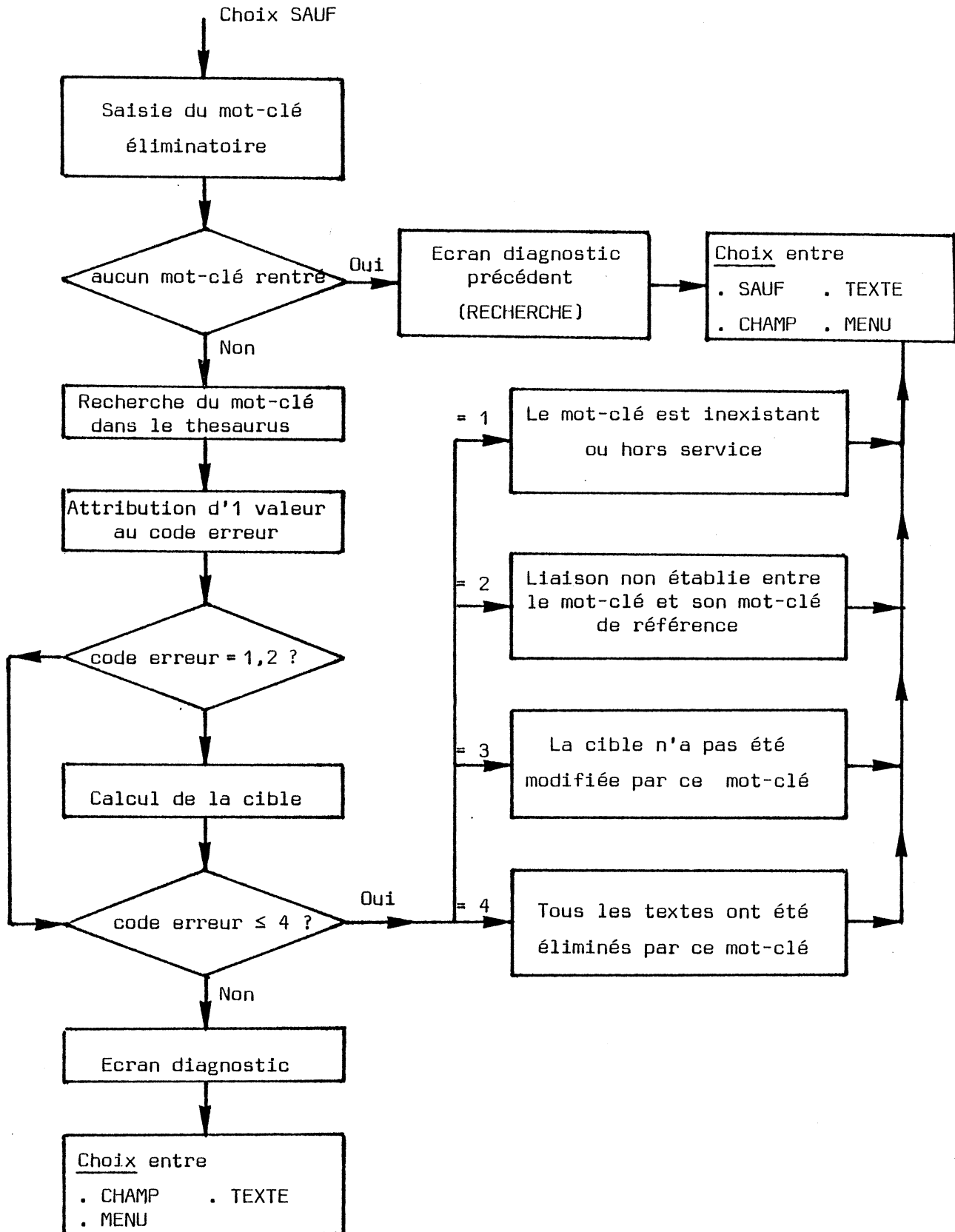
Dans tous les cas, on revient au niveau précédent la diminution avec SAUF : choix entre SAUF, CHAMP, TEXTE, MENU, sur l'ancienne cible (résultante de la recherche par mots-clés).

- ④ Si la diminution de la cible avec SAUF a été positive, l'utilisateur peut encore diminuer la nouvelle cible par une sélection sur les champs (date, nature juridique).
- ⑤ Enfin, on propose à l'utilisateur de recommencer la sélection sur les champs, quand celle-ci a abouti à une cible nulle.



Organigramme de la partie RECHERCHE à partir des mots-clés

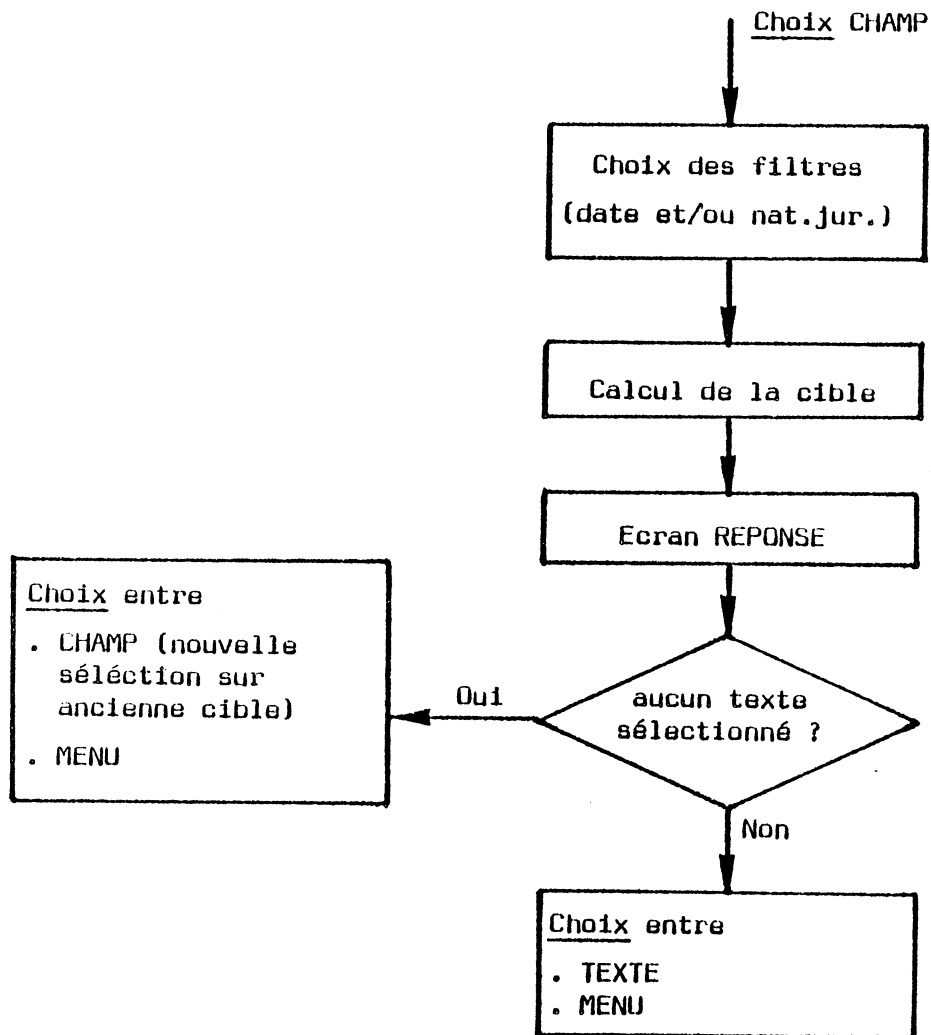
Organigramme de la partie SAUF



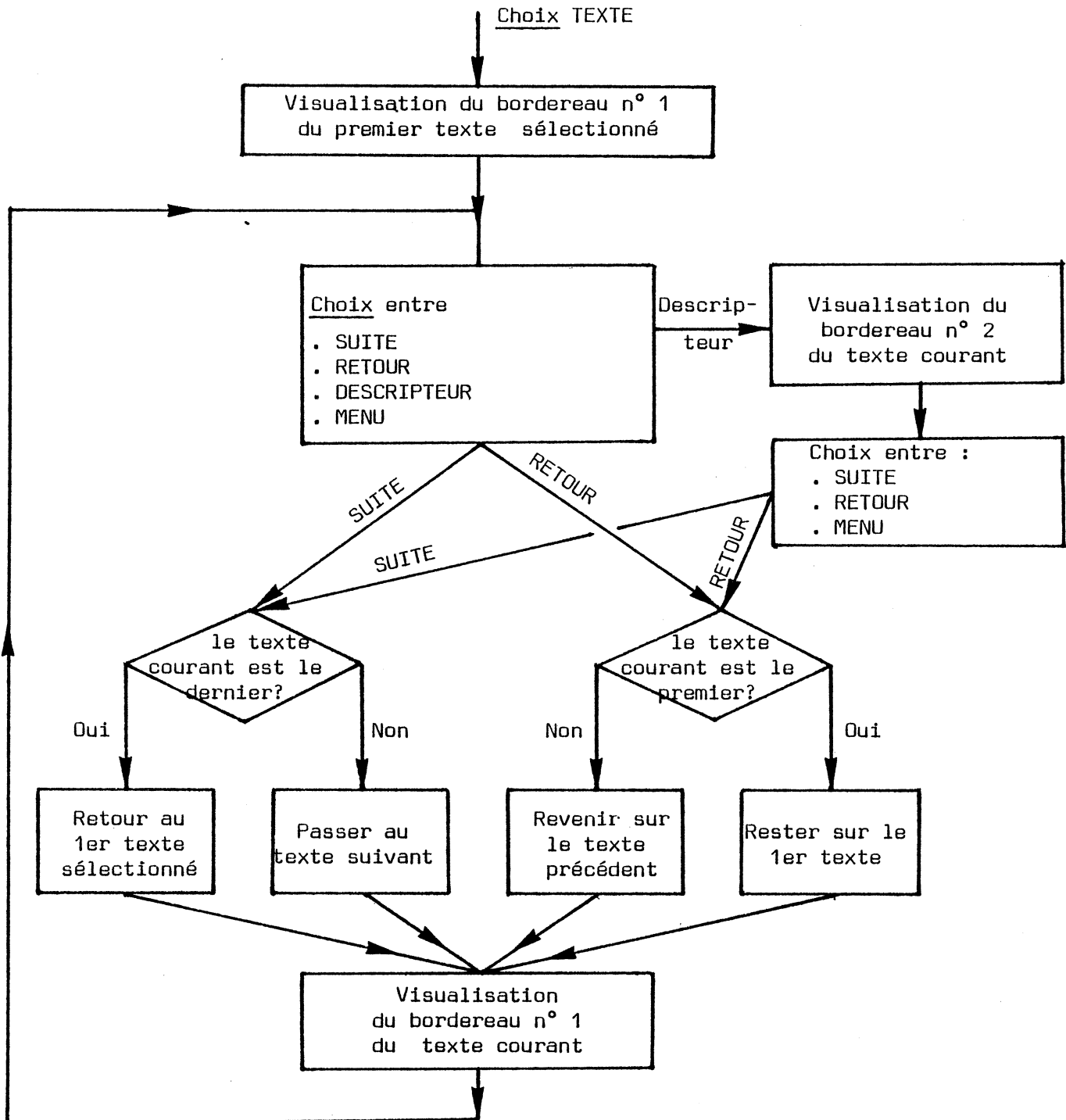
Organigramme de la partie CHAMP

Remarques

On peut, après avoir utilisé le filtre SAUF, préciser la cible par une sélection sur la nature juridique ou les dates. Par contre, une fois la cible précisée par les critères (nature juridique, dates), on ne peut utiliser le filtre SAUF. Ceci ne résulte pas d'une difficulté de programmation, mais d'un choix dans la conception du système de recherche.



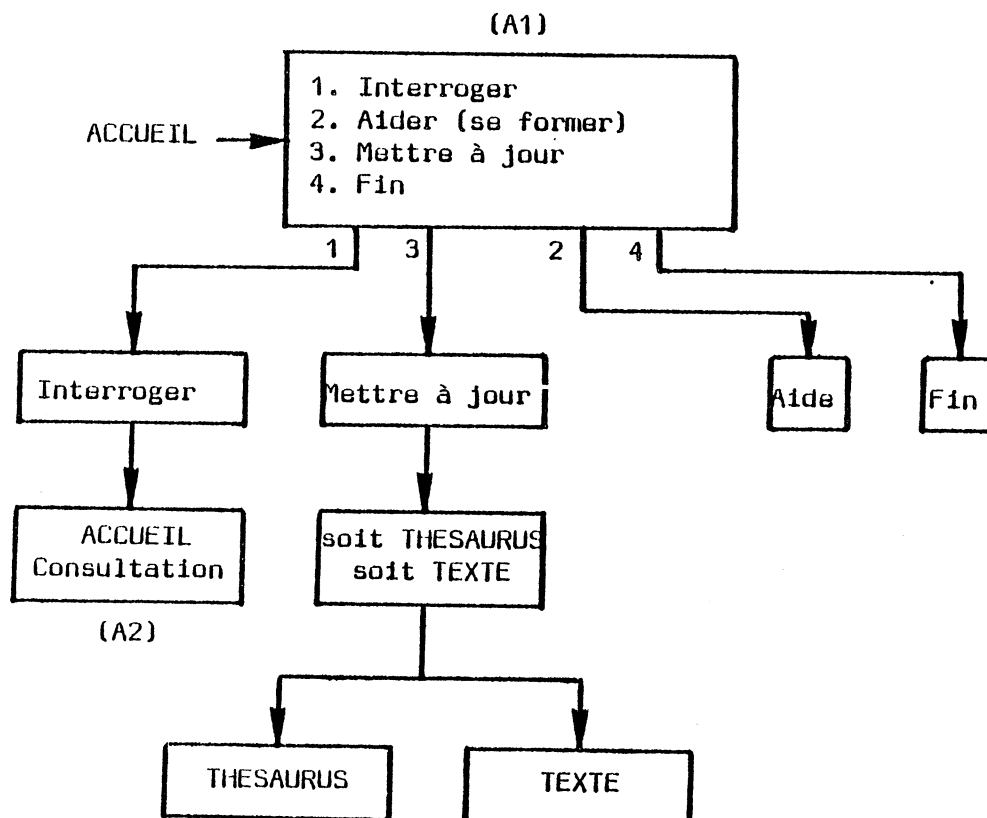
Organigramme de la partie TEXTE

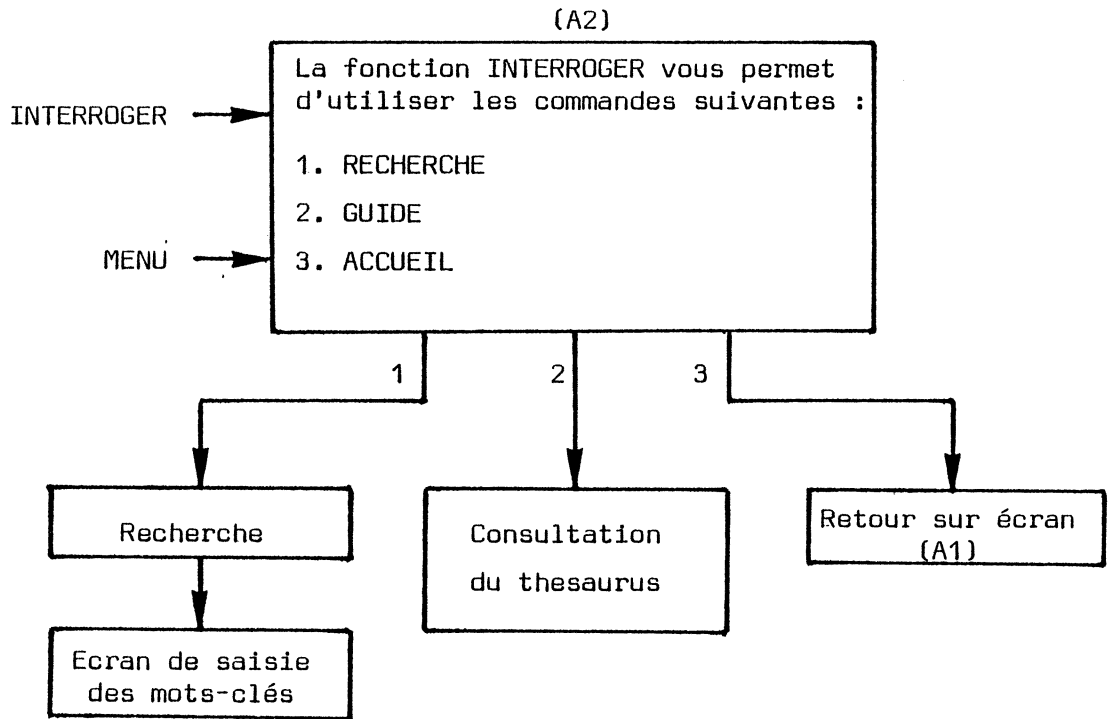


2.5.3. Le système de recherche (avec filtres) *

Principe :

sur la page accueil, choisir une fonction :



InterrogerRecherchea) Ecran de saisie des mots-clés

POSER VOTRE QUESTION
à l'aide de vos mots-clés
(taper un mot-clé et envoyer le)
?

(Cas où l'on veut
passer directement
par filtre
... plus tard !)

POSER VOTRE QUESTION
à l'aide de vos mots-clés
(taper un mot-clé et envoyer le)
? STAGIAIRE
et
? VOIRIE
et
?

b) Ecrans de diagnostic

VOUS N'AVEZ PAS DONNE DE MOT CLE
 ... LA RECHERCHE PAR MOTS CLES
 N'EST PAS EFFECTUEE

 POUR ABANDONNER TAPER MENU : _____

1 _____ STAGIAIRE
 MOT INCONNU
 2 _____ VOIRIE
 INDEXE 5 TEXTES

 RECHERCHE STOPPEE CAR INEXISTENCE
 DU MOT-CLE N° : 1 DANS LE THESAURUS

 POUR ABANDONNER TAPER MENU : _____

1 _____ STAGIAIRE
 INDEXE 4 TEXTES
 2 _____ VOIRIE
 N'INDEXE PAS DE TEXTE

 AUCUN TEXTE N'A ETE SELECTIONNE CAR LE
 MOT-CLE N° 2 N'INDEXE PAS DE TEXTE.
 UTILISER UN MOT DESCRIPTEUR QUI A UN
 LIEN AVEC LUI DANS LE THESAURUS

 POUR ABANDONNER TAPER MENU : _____

1 _____ STAGIAIRE
 INDEXE 4 TEXTES
 2 _____ VOIRIE
 PAS DE LIAISON AVEC MOT DE REFERENCE

 RECHERCHE STOPPEE CAR LA LIAISON DU MOT
 N° 2 AVEC SON MOT DE REFERENCE N'EXISTE
 PAS DANS LE THESAURUS. A METTRE A JOUR.

 POUR ABANDONNER TAPER MENU : _____

1 ____ STAGIAIRE
 INDEXE 4 TEXTES

2 ____ VOIRIE
 INDEXE 1 TEXTE

 AUCUN TEXTE N'A ETE SELECTIONNE A PARTIR
 DE CES MOTS-CLES. VOUS POUVEZ OBTENIR
 DES TEXTES EN DONNANT MOINS DE MOTS-CLES.

 POUR ABANDONNER TAPER MENU : ____

1 ____ STAGIAIRE
 INDEXE 10 TEXTES

2 ____ VOIRIE
 INDEXE 15 TEXTES

 8 TEXTES ONT ETE SELECTIONNES.

 POUR VOIR LES TEXTES, TAPER TEXTE
 POUR PRECISER LA CIBLE, TAPER CHAMP
 POUR ELIMINER DES TEXTES, TAPER SAUF : ____
 POUR ABANDONNER, TAPER MENU

1 ____ STAGIAIRE
 INDEXE 8 TEXTES

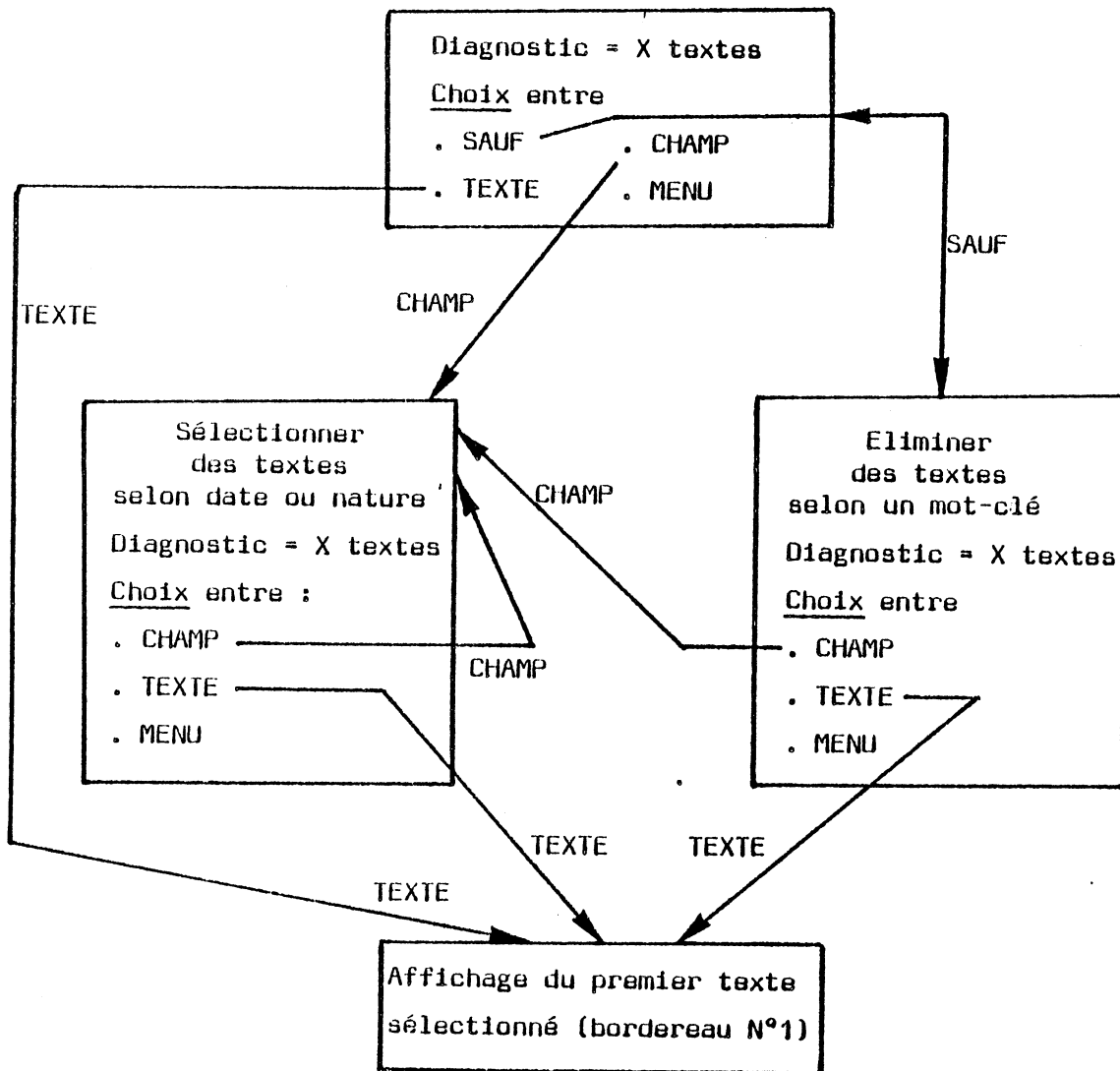
2 ____ VOIRIE
 INDEXE 5 TEXTES

 2 TEXTES ONT ETE SELECTIONNES

 POUR VOIR LES TEXTES, TAPER TEXTE : ____
 POUR ABANDONNER, TAPER MENU

(si moins de 5 textes sélectionnés, on ne propose par le filtre SAUF, ni la sélection sur CHAMPS).

Filtre et texte : principe



Filtre SAUFa) Ecran de choix et saisie :

VOUS AVEZ SELECTIONNE 8 TEXTES AVEC :
 STAGIAIRE
 VOIRIE
 SI VOUS VOULEZ ELIMINER DES TEXTES
 TAPEZ UN MOT-CLE (EN MAJUSCULES)
 SINON APPUYEZ SUR ENVOI
SAUF AUXILIAIRE

b) Ecrans de diagnostic :

8 TEXTES ONT ETE OBTENUS AVANT LE
 FILTRE :
 AUXILIAIRE

 ELIMINATION STOPPEE CAR LA LIAISON DU
 MOT-CLE AVEC SON MOT DE REFERENCE
 N'EXISTE PAS DANS LE THESAURUS.
 A METTRE A JOUR.
 VOUS POUVEZ REDONNER UN AUTRE FILTRE
 OU CONTINUER AVEC LES 8 TEXTES.

 POUR UN AUTRE FILTRE, TAPER SAUF
 POUR VOIR LES TEXTES, TAPER TEXTE
 POUR PRECISER LA CIBLE, TAPER CHAMP : ____
 POUR ABANDONNER, TAPER MENU

8 TEXTES ONT ETE OBTENUS AVANT LE FILTRE :
 AUXILIAIRE

 ELIMINATION STOPPEE CAR INEXISTENCE DU
 MOT-CLE DEMANDE DANS LE THESAURUS.
 VOUS POUVEZ REDONNER UN AUTRE FILTRE OU
 CONTINUER AVEC LES 8 TEXTES.

 POUR UN AUTRE FILTRE, TAPER SAUF
 POUR VOIR LES TEXTES, TAPER TEXTE
 POUR PRECISER LA CIBLE, TAPER CHAMP : ____
 POUR ABANDONNER, TAPER MENU

8 TEXTES ONT ETE OBTENUS AVANT LE FILTRE :
AUXILIAIRE QUI INDEXE 15 TEXTES

TOUS LES TEXTES SONT ELIMINES PAR CE
MOT-CLE.

VOUS POUVEZ RECOMMENCER AVEC UN AUTRE
FILTRE OU CONTINUER AVEC LES 8 TEXTES.

POUR UN AUTRE FILTRE, TAPER SAUF

POUR VOIR LES TEXTES, TAPER TEXTE

POUR PRECISER LA CIBLE, TAPER CHAMP : ____

POUR ABANDONNER, TAPER MENU

8 TEXTES ONT ETE OBTENUS AVANT LE FILTRE :
AUXILIAIRE QUI N'INDEXE PAS DE TEXTE

8 TEXTES ONT ETE SELECTIONNES
LA SELECTION N'A PAS ETE MODIFIEE PAR CE
MOT-CLE, MAIS VOUS POUVEZ L'AFFINER EN
REDONNANT UN AUTRE FILTRE.

POUR UN AUTRE FILTRE (Id. ci-dessus)

....

8 TEXTES ONT ETE OBTENUS AVANT LE FILTRE :
AUXILIAIRE QUI INDEXE 5 TEXTES

4 TEXTES ONT ETE SELECTIONNES.

POUR VOIR LES TEXTES, TAPER TEXTE

POUR PRECISER LA CIBLE, TAPER CHAMP

POUR ABANDONNER, TAPER MENU : ____

Filtre CHAMPa) Ecran de choix et saisie

PRECISER LES CHAMPS DEVANT SERVIR DE CRITERE DE SELECTION			
	Année	Mois	Jour
DATE DES TEXTES :
DATE SUPPORTS :
Par rapport à cette date, les textes à sélectionner doivent être (0 = OK) :			
POSTERIEURS :	0		
DE LA MEME DATE :	.		
ANTERIEURS :	.		
Quelle est la nature des textes à sélectionner ?			
LEGISLATION :	0		
JURISPRUDENCE :	.		
COMMENTAIRE :	.		
APPUYER SUR LA TOUCHE ENVOI POUR SAUTER UNE RUBRIQUE			

b) Ecrans de réponse

8 TEXTES ONT ETE OBTENUS AVANT LE FILTRE :
DATE : APRES 15.01.81
SOURCE DROIT : LEGISLATION, JURISPRUDENCE

5 TEXTES ONT ETE SELECTIONNES
A PARTIR DE CES CRITERES

POUR VOIR LE PREMIER TEXTE, TAPER <u>TEXTE</u>
POUR ABANDONNER, TAPER <u>MENU</u> : ___

8 TEXTES ONT ETE OBTENUS AVANT LE FILTRE :
 SOURCE DROIT = LEGISLATION

 AUCUN TEXTE N'A ETE SELECTIONNE
 A PARTIR DE CE(S) CRITERE(S)

 POUR UNE AUTRE SELECTION, TAPER CHAMP
 POUR ABANDONNER, TAPER MENU : ____

Commande TEXTE :

N° DE FICHE :
 SOURCE DU DROIT :
 REFERENCE DU SUPPORT
 REFERENCE DU TEXTE
 ANALYSE DU TEXTE

SUITE (TEXTE SUIVANT),
RETOUR (PRECEDENT),
DESCRIPTEUR (MOTS-CLES),
MENU : ____

On affiche le bordereau n° 1 du texte suivant dans la sélection.

Il faudrait toujours préparer l'affichage du bordereau n° 2 de ce texte pour diminuer le temps de visualisation, si l'utilisateur tape DESCRIPTEUR : remplir les tables GESECRAN ...

Si on tape TEXTE pour la première fois (voir ECRAN TEXTE), c'est le premier texte qui s'affiche.

Si on tape SUITE alors qu'on est sur le dernier texte, on reboucle sur le premier texte.

RETOUR :

N° DE FICHE : SOURCE DU DROIT : REFERENCE DU SUPPORT REFERENCE DU TEXTE ANALYSE DU TEXTE <u>SUITE</u> (TEXTE SUIVANT) <u>RETOUR</u> (PRECEDENT) <u>DESCRIPTEUR</u> (MOTS-CLES), <u>MENU</u> : _____

On affiche le bordereau 1 du texte précédent dans la sélection.

Il faudrait, comme dans la commande TEXTE, préparer l'affichage du bordereau 2.

Si on tape RETOUR alors qu'on est sur la première fiche texte, on boucle sur la première fiche.

DESCRIPTEUR :

N° DE FICHE : MOTS-CLES : <u>SUITE</u> (TEXTE SUIVANT) <u>RETOUR</u> (TEXTE PRECEDENT) : _____ <u>MENU</u> (ON RECOMMENCE !)
--

On affiche le bordereau 2 de la fiche texte courante.

CHAPITRE 6

ÉTUDE COMPARATIVE D'AUTRES LOGICIELS DU MÊME TYPE

Les réalisations du type "micro-informatique documentaire" sont encore balbutiantes en France ; il en existe cependant comme le prouvent les quelques exemples suivants :

Nous excluons les Systèmes de Gestion de Bases de Données trop éloignés du projet CATHI, ainsi que les systèmes documentaires nécessitant des moyens informatiques relativement puissants tels que les projets :

FRERES, SYSIDORE, fonctionnant sur IRIS 80 de CII HB

PRODOC sur MULTICS de CII HB

UNIBASE sur 66 DPS de CII HB

PIAF et ses dérivés qui ont demandé plusieurs années de recherche en informatique et en linguistique. La signification de PIAF "Programmes Interactifs d'Analyse des Langues Naturelles" prouve que ce projet va bien au-delà des simples systèmes documentaires avec thesaurus que nous étudions.

En nous limitant donc aux simples systèmes documentaires fonctionnant sur micro-ordinateur, nous pouvons citer les produits suivants : ADDOC, BIBLIO 8000, DESCARTES, DICO, GESTDOC, MANDOLINE, SIGMINI, TE 80, VALSE et un prototype mis au point au Laboratoire de Photosynthèse de Gif sur Yvette.

Sur la plupart, nous n'avons que des renseignements partiels, nous nous contenterons donc de les présenter ; par contre, en ce qui concerne BIBLIO 8000 et GESTDOC, une documentation plus fournie nous permettra d'établir un tableau comparatif de leurs principales caractéristiques, ainsi que de celles de CATHI et d'un autre logiciel, MILOR, que nous avons rajouté bien que fonctionnant sur un mini-ordinateur, car il semblait difficile de passer sous silence un logiciel présentant autant de possibilités (la frontière entre minis et micros devient d'ailleurs de plus en plus floue).

ADDOC

Logiciel conversationnel en PASCAL et exploitable sur APPLE II.

Matériel : mémoire 48 Ko, 2 unités de disquettes.

Définition : système documentaire et SGBD développé sur micro-ordinateur, afin de créer une base d'informations personnelles, de manipuler et de rechercher ces informations (notamment recherche de documents en fonction de leur contenu). La recherche est effectuée par affinages successifs des réponses ; les questions sont posées sous forme de mots-clés reliés entre eux par des opérateurs booléens.

Organisme : CERISS, Université de Toulouse 1.

DESCARTES

Développé par AV21.

Configuration nécessaire : un GOUPIL 2 (64 Ko)

- . avec un disque 8 pouces DF, DD, peut gérer 2500 à 4000 fiches,
- . avec un disque dur, 10 fois plus de fiches.

Paramétrable, DESCARTES est un gestionnaire de fichier auquel s'ajoutent les modules de gestion de prêt, d'abonnement, d'adhérents et de thesaurus.

DICO

Mis à l'étude en 1979, ce logiciel est surtout adapté aux bibliothèques.

Il propose les fonctions suivantes :

- . saisie, modification, suppression de documents
 - différentes rubriques
 - descripteurs (autant que l'on désire jusqu'à concurrence de 135 caractères)
 - résumé en langage naturel de longueur fixe (256 caractères)
- . indexation
- . listage du lexique,

. recherche documentaire

- équation de recherche composée de un ou plusieurs mots-clés reliés par les opérateurs booléens ET, OU, SAUF
- les résultats de 10 étapes peuvent être mémorisés et entrecroisés soit entre eux, soit avec une nouvelle recherche
- recherche sur nom d'auteur
- recherche sur nom d'auteur et mot-clé

MANDOLINE

Définition : MANDOLINE permet de gérer et de manipuler en conversationnel des fichiers d'objets décrits comme une suite de couples d'attributs-valeur en format variable.

Configuration : TEKTRONIX 4051, mémoire 32 Ko.

Organisme : CRA (Centre de Recherches Archéologiques) de Valbonne.

MANDOLINE s'apparente plus à un SGBD qu'à un système documentaire.

SIGMINI

Définition : système d'informations apparenté aux SGBD et aux systèmes documentaires, avec un dictionnaire de "mots" et leurs relations de synonymie.

Traitements effectués : mise à jour en conversationnel du dictionnaire, des informations, des fiches de la base, interrogation interactive, éditions diverses.

Configuration : SIGMINI est développé sur UNIVAC V72 avec 64 Ko minimum, il est prévu pour être portable sur MINI 6.

Langage : 10 000 instructions FORTRAN.

Organisme : ENSMP - CAI (Ecole Nationale Supérieure des Mines de Paris).

TE 80

Définition : système de gestion bibliographique, mis au point par sélection informatique et diffusé depuis deux ans.

Le temps d'accès aux enregistrements des différents fichiers d'articles documentaires et de mots-clés, est réduit en moyenne à 200 millisecondes.

Matériel : MICRAL 80 21C.

VALSE

Gestion de références bibliographiques développé par Marcel Gibert (informaticien) et Hélène Desvals (documentaliste) au Centre de Thermodynamique et et Microcalorimétrie de Marseille.

VALSE fonctionne sur Horizon de Northstar System.

Prototype du Laboratoire de Photosynthèse de Gif sur Yvette

Composé de deux programmes intitulés :

- . entrée et gestion de données,
- . recherche bibliographique.

D'après les premiers tests, la recherche sur deux critères prendra 30 secondes sur une base de 1 500 articles.

Un des buts fixés : rendre les chercheurs et documentalistes propriétaires de cet outil : entrée des données faite par différentes personnes, choix des mots-clés et édition des fichiers par les chercheurs.

Matériel : micro-ordinateur TRS 80 mod II (TANDY), 4 lecteurs de disquettes.

Nom du Logiciel	Société ou organisme concepteur	Type de matériel
BIBLIO 8000	Laboratoire de Biochimie Physique, INRA, Université de Paris Sud, Centre d'Orsay, 91405 ORSAY CEDEX Réalisation : Roland SALESSE	Système COMMODORE 8001. Pour son utilisation, le programme nécessite une unité centrale CBM 8096 (32 K + 64 K extension), une unité de disquette 77 secteurs (2 x 512 K) CBM 8050 ou CBM 8250 ou 2 x 1,1 Mo.
GESTDOC	Société DATAFRAME Micro Informatique 45 rue de Jourmel, 75015 PARIS 577 13 12 + 577 13 04	Fonctionne sur des micro-ordinateurs utilisant CP/M ou MP/M. Exemple de configuration pour gérer 20 000 documents de 200 caractères : 1 micro ITT 3030, 64 K de mémoire, 1 disquette de sauvegarde et 1 disque dur de 5 Mo.
MILOR	? GIXI ?	Disponible sur plusieurs gammes de mini-ordinateurs HP 1000 et HP 3000 de Hewlett Packard, MINI 6 de CII HB, PDP de Digital Equipment. Transportable sur les ordinateurs de la gamme 16 bits disposant de ces outils de base.
CATHI	ICARE, Mairie de Grenoble, Boulevard Jean Pain, 38000 GRENOBLE, 42 81 42. Ont participé à la réalisation : CEDIS, 28 Cours Lafayette, 69003 Lyon et le GIS de Mini et Micro Informatique CNRS/INPG, b.p. 68, 38402 St Martin d'Hères Cedex 44 55 96	Conçu sur GOUPIL 2 avec 64 K de mémoire, avec une unité de disquette (5" et 8"), ou une unité de disque dur CII HB de 2 fois 10 millions d'octets dont 10 amovibles.

Nom du Logiciel	Définition	Langage	Description fonctionnelle
BIBLIO 8000	Gestion d'un fichier bibliographique comptant jusqu'à 7 000 et 14 000 fiches selon la version. Le logiciel est utilisable par des personnes non initiées à l'informatique.	BASIC mais les programmes de tri et de recherche sont écrits en ASSEMBLEUR 6502.	Ce logiciel est découpé en 12 modules effectuant les fonctions suivantes : <ul style="list-style-type: none"> . inscription, édition, modification, suppression de fiches . recherche multi-critères, . modification de mots-clés et des noms de journaux, . impression : - fichier <ul style="list-style-type: none"> - listes des mots-clés, des journaux, des références par journal et des références bibliographiques . mises à jour des tables de tri et de recherche.
CATHI	Système documentaire avec thesaurus sur une base de données communes	BASIC, le tri et quelques instructions ont été écrits en ASSEMBLEUR.	CATHI se compose de deux modules : <ul style="list-style-type: none"> . le premier comprend les programmes de création, de réorganisation et de contrôle de la base, . le second comprend les programmes de mise à jour (thesaurus, fonds documentaire), de consultation (thesaurus, fonds documentaire) et un programme d'aide à l'utilisateur. Ces programmes s'enchaînent par l'intermédiaire d'une page d'accueil où se trouve le menu général de ce second module.
GESTDOC	La Société DATATRAM a construit GESTDOC autour du SGBD relationnel DBASE II. Elle y a rajouté les algorithmes complémentaires propres à la recherche documentaire. GESTDOC peut être utilisé pour gérer des fichiers clients.	D BASE II	Fonctions effectuées : définition de la structure des documents ; recherche par mots clés et index et étendue à tous les champs ; création et mise à jour des documents ; gestion du lexique des mots-clés ; gestion des lecteurs, gestion des prêts ; listes et statistiques diverses ; utilitaires de sauvegarde/restauration.
ILOR	Ce logiciel permet d'organiser et de retrouver des informations à caractère analytique. Il est conçu pour gérer des fichiers privés sur des mini-ordinateurs disposant d'un système de gestion de terminaux, d'un système de gestion de fichiers permettant l'accès direct, et d'un compilateur FORTRAN IV.	FORTRAN IV	Fonctions effectuées : maintenance du langage documentaire ; maintenance des signalements ; recherche rétrospective ; utilitaires d'édition, de statistiques, etc ... Le logiciel a été pourvu d'un système de MENU. Deux types de MENUS : un MENU pour l'administrateur de la base qui donne accès à l'ensemble des tâches et en particulier à celles de la mise à jour, et un MENU pour l'utilisateur, qui permet l'accès aux tâches d'interrogation et de consultation, voire de saisie.

Nom du logiciel	Listes, éditions statistiques	
BIBLIO 8000	<p>Edition des fiches avec possibilité de constituer un fichier en utilisant certains types d'imprimantes et de supports (papier accordéon, cartes, fiches, ..) ;</p> <p>Edition des noms de journaux avec abréviation d'usage par ordre alphabétique ;</p> <p>Edition de la liste des mots clés avec leurs abréviations courantes par ordre alphabétique ;</p> <p>Listage des références et numéros de fiches issues d'un même journal ;</p> <p>Listes bibliographiques pour un article ou une publication, avec possibilité de les modifier avec un logiciel de traitement de texte.</p>	
GESTDOC	<p>Liste des documents sélectionnés lors d'une recherche.</p> <p>Liste inverse (mots-clés - document).</p> <p>Liste pour chaque lecteur intéressé, des nouveaux documents par pôle d'intérêt.</p>	<p>Statistiques diverses :</p> <p>Nombre de documents dans la base</p> <p>Type de documents les plus empruntés.</p>
MILOR	<p>Listes diverses (nous possédons peu d'information à ce sujet).</p> <p>Editions statistiques.</p>	
CATHI	<p>Edition du lexique des mots-clés avec ou sans leurs relations.</p> <p>Edition des fiches thesaurus et des fiches documentaires.</p>	

Nom du Logiciel	Utilisations, champs d'application	Extensions télématiques
BIBLIO 8000	Testé sur cinq sites (autres que le Labo). Utilisable par des personnes non initiées à l'informatique.	Ne semblent pas prévues pour le moment.
CATHI	Mise en place d'une maquette d'expérimentation à la Mairie de Grenoble. Améliorations diverses du logiciel et extensions télématiques.	Extension vidéotex multiconsultation pour permettre la communication entre différents services de documentation de différentes villes de la région. A long terme, pouvoir interroger d'autres bases locales, voire nationales.
GESTDOC	Peut être paramétré pour des utilisations très diverses : <ul style="list-style-type: none"> . gestion de documents, de livres, de photos, films, microfilms, plans, etc .. . gestion de devis, de propositions, . aide au diagnostic médical, . gestion de stock, de chantiers, etc .. . gestion des lecteurs, des prêts ... 	En utilisant des logiciels de communication (exemple : ASCOM), l'utilisateur peut relier son micro-ordinateur à un serveur et acquérir des fichiers. DBASE II permet d'intégrer ces fichiers dans la base de données locales.
MILOR	Premières applications : gestion de bibliothèques et de centres de documentation. Outre les applications bibliographiques, le logiciel MILOR a été utilisé pour gérer des données dans les domaines suivants : <ul style="list-style-type: none"> . gestion d'archives en bureautique (suivi de dossier), . aide à la décision dans un service après vente, . assistance technique aux utilisateurs : système de type "SVP" pour répondre à des questions techniques . gestion de dossiers. 	Connexion à un réseau de transmission réalisée. Possibilité pour plusieurs utilisateurs d'interroger simultanément le fonds documentaire pendant que d'autres postes connectés à l'unité centrale sont actifs en saisie.

Nom du Logiciel	Thesaurus	Fonds documentaire	Recherche
BIBLIO 8000	<p>Nombre total de mots-clés utilisés : 250 ou 1000 (2000 en version 8250). Nombre maximum d'auteurs et de mots-clés par fiche : 25. Exemple : si 15 mots-clés, alors 10 noms d'auteurs. Améliorations en cours : introduction des relations de synonymie et de hiérarchie à deux niveaux entre les mots-clés.</p>	<p>7 000 ou 14 000 fiches selon la version Création, modification, suppression, édition des fiches et contrôle en temps réel. Structure des fiches définies par l'utilisateur. Améliorations en cours : utilisation d'un générateur de grille de saisie. Temps de saisie d'une fiche : 2 mn en moyenne et 3 mn avec 5 300 fiches.</p>	<p>Recherche multi-critères rapide sur toute combinaison de références et/ou mots-clés et/ou noms d'auteurs avec combinatoire et/ou/sauf sur les mots-clés et sortie ordonnée selon l'ordre alphabétique du 1er auteur, ordre chronologique ou ordre du fichier. Lors de la recherche par mots-clés, ceux-ci peuvent être abrégés ou tronqués. Visualisation sur écran et/ou impression des fiches sélectionnées. Indication du nombre de fiches sélectionnées et possibilités de cumuler des recherches successives. La recherche s'effectue sur des tables transférées en mémoire. Temps de recherche multi-critères : 1 mn 20 s sur 2 700 fiches et 2 mn 30s sur 5 300 fiches.</p>
ATHI	<p>Le lexique peut contenir 1500 mots-clés de 38 caractères maximum chacun. Type de relation : synonymie, hiérarchie, analogie. Saisie, mise à jour et contrôle du thesaurus en conversationnel et en temps réel. 2 types de mots clés : . les descripteurs indexant les textes, reliés par les relations de hiérarchie et d'analogie, . les synonymes associés aux descripteurs. Le choix des mots-clés, des relations entre mots-clés et des descripteurs est fait par les documentalistes. A terme : 1 500 mots-clés.</p>	<p>Fonds documentaire : ensemble de fiches documentaires. Chaque fiche se compose d'une partie "référence" et d'une partie "descripteur" : mots-clés indexant un texte. Il n'y a pas d'indexation automatique. La saisie, la mise à jour et le contrôle des fiches documentaires se font en conversationnel et en temps réel à l'aide d'un gestionnaire d'écran. A terme : 5 000 ou 6 000 fiches documentaires.</p>	<p>Les programmes de consultation permettent la consultation du thesaurus et du fonds documentaire. La recherche sur le fonds documentaire se fait en utilisant les mots-clés (descripteurs et synonymes), chaque question comportant un ou plusieurs mots-clés (6 au maximum) reliés uniquement par l'opérateur ET. L'utilisateur peut diminuer la cible obtenue en utilisant l'opérateur SAUF sur un mot-clé, ou en utilisant les filtres sur la date et/ou la nature juridique. L'utilisateur devrait utiliser le thesaurus (grâce aux relations entre les mots-clés) pour affiner sa question ou la reformuler. Temps de réponse ≤ 2 s.</p>

<p>ESTDOC</p>	<p>Existence d'un lexique avec possibilité de création, mise à jour, suppression de mots-clés. Nous disposons de peu d'informations dans ce domaine.</p>	<p>32 champs par document de 1 à 255 caractères. L'utilisateur peut définir la structure des différents types de documents. 16 champs sur 32 peuvent servir d'index (recherche rapide). Nombre variable de mots-clés par document. Vérification de l'existence des mots-clés descripteurs dans le lexique. La saisie et la correction des documents se font à l'aide d'une grille par déplacement du curseur.</p>	<p>Recherche rapide sur les index et les mots-clés, opérateurs ET, OU, SAUF. Recherche étendue à tous les champs, plus lente que la précédente. Opérateurs ET, OU, SAUF, >*, <*, § et C. Affiche le nombre de références trouvées et permet de poursuivre les recherches sur les résultats acquis ou antérieurs, ou de lister les documents sélectionnés.</p>
<p>ELOR</p>	<p>Thesaurus avec relations de hiérarchie et de synonymie. Les descripteurs peuvent appartenir à trois langues différentes. La maintenance du langage documentaire est faite en conversationnel.</p>	<p>Définition par l'utilisateur, du format de la notice bibliographique et du mode de dissémination de l'information (bande magnétique, connexion à un réseau ..). Gestion habituelle d'un fonds documentaire avec saisie, correction, mise à jour des bordereaux.</p>	<p>Choix de la langue pour la recherche rétrospective (en conversationnel). Possibilité pour plusieurs utilisateurs d'interroger simultanément le fonds documentaire pendant que d'autres postes connectés à l'unité centrale sont actifs en saisie. Choix du dialogue de consultation, soit en mode didactique : jeu de questions et réponses entre l'utilisateur et la machine, soit en mode commande : le dialogue se fait grâce à un langage de commandes développé à partir des recommandations du "Common Command Set" d'EURONET et de la commission AFNOR "Documentation Automatisée."</p>

Conclusion

Il est difficile de faire le point et d'établir des comparaisons. En effet

- . la plupart des logiciels existants ont été écrits en vue d'applications spécifiques (c'est en particulier le cas de CATHI) et les traitements ou fonctions qu'ils proposent sont très différents d'un logiciel à l'autre. Ils existent d'ailleurs plus à l'état expérimental qu'à un stade véritablement opérationnel ;
- . l'absence de coordination ou de centralisation dans ce domaine, rend longue et difficile la récolte d'informations, si bien que les renseignements dont nous disposons sont partiels et varient d'un logiciel à l'autre

On peut cependant faire un certain nombre de remarques :

1. Il est regrettable :

- que les logiciels soient difficilement portables d'un matériel à un autre, mais ceci est surtout du ressort des constructeurs ;
- qu'ils soient également difficilement adaptables à d'autres bases de données en vue d'applications diverses ; il faudrait cependant une étude plus approfondie pour être catégorique à ce sujet.

Certains logiciels documentaires par leur aspect "SGBD" ont des champs d'applications plus vastes, mais ils sont en général dépourvus de thesaurus, tout au plus ont-ils un lexique de mots-clés, exception faite de MILOR qui possède un thesaurus et dont la structure paramétrée le rend adaptable à différentes bases.

En ce qui concerne CATHI, il faut regretter que le logiciel ne puisse s'adapter directement à d'autres fonds documentaires, du moins que cet aspect n'ait pas été envisagé dès le départ (voir critiques au niveau du logiciel CATHI).

2. Autre aspect négatif : les extensions télématiques ne sont pas toujours prévues, or à terme les systèmes informatisés dialogueront entre eux et la recherche d'informations se fera à travers des réseaux.

3. Sur les performances des logiciels et les volumes de stockage, on ne peut comparer deux logiciels, dont l'un fonctionne sur un micro-ordinateur disposant de 96 K et d'une unité de disque dur, avec un autre fonctionnant sur un micro-ordinateur de 32 K et d'une unité de disquette 5".
A l'avenir, les problèmes de performance de logiciels (temps de réponse par exemple) et volume de stockage, se poseront moins car des micro-ordinateurs possédant 128 K ou plus, ainsi que des mémoires de masse (10 à 20 millions d'octets), font leur apparition sur le marché et on peut espérer que leur coût diminuera. En attendant ils ne sont pas accessibles à tout le monde.

Faut-il rejeter pour autant les systèmes documentaires fonctionnant sur des matériels disposent de peu de ressources ; le Professeur ARSAC répond : "l'inconvénient d'un temps de réponse trop long, l'utilisateur le compense par une indépendance plus grande", ou encore "les facilités d'emploi et d'échange devraient séduire un grand nombre d'utilisateurs" (Bulletin du Centre des Hautes Etudes Internationales d'Informatique Documentaire, Editorial 2ème trimestre 1982).

4. Le développement des bases de données sur des micro-ordinateurs ne se fera que si dès le départ le travail est conçu en pensant aux futurs utilisateurs ; il faut spécialement soigner les "images écran", veiller à la simplification des modes d'interrogation et bien sûr à tous les problèmes de normalisation.



CHAPITRE 7

CRITIQUES ET CONCLUSIONS

Nous énumérons dans ce chapitre les critiques les plus importantes que l'on peut faire sur le plan technique, en suggérant un certain nombre d'améliorations possibles.

Nous nous plaçons uniquement sur le plan technique, sans discuter du bien-fondé de telle ou telle décision résultant de la philosophie même qui a guidé le projet et qui n'est pas de notre ressort.

Nous distinguons les critiques suivantes :

- 1) Champ d'application du logiciel trop restreint,
- 2) taille trop importante donnée à certains programmes,
- 3) absence de recherche par mots-clés tronqués, et par thème, absence d'indexation des documents par des champs autres que les descripteurs, ces deux critiques ne sont d'ailleurs pas du domaine technique,
- 4) Erreurs dans la programmation.

1. Champ d'application du logiciel

Le logiciel a un champ d'application trop restreint.

On peut envisager de faire certaines modifications de manière à utiliser CATHI pour d'autres bases de données, à condition toutefois que celles-ci aient la même structure que celle de CATHI : indexation des documents par des descripteurs appartenant à un thesaurus, celui-ci étant composé de mots-clés reliés par des relations de synonymie, de hiérarchie et d'analogie.

Pour s'en tenir à des modifications raisonnables, c'est-à-dire qui n'entraîneraient pas des corrections trop importantes dans les programmes, on peut proposer que :

- . la gestion du thesaurus reste la même, ainsi que le volume d'informations associées à chaque mot-clé qui devraient pouvoir être représentées sur une page écran de 20 x 40 caractères (correspondant à l'écran du terminal Minitel). Par contre, on paramètrerait la longueur maximale des mots-clés et le nombre de mots-clés liés. L'utilisateur définirait lui-même la présentation d'une fiche thesaurus. Les deux rubriques type et numéro d'ordre seraient conservées. La note d'application serait toujours une zone libre entièrement gérée par l'utilisateur, mais sa longueur pourrait varier ainsi que son nom.
- . En ce qui concerne le fonds documentaire, chaque fiche documentaire serait toujours composée de deux bordereaux correspondant chacun à une page écran de 20 x 40 caractères :
 - sur le premier figureraient toujours les descripteurs ; le nombre maximal par fiche serait un paramètre ;
 - les différentes rubriques du premier bordereau seraient entièrement définies par l'utilisateur.

On se limiterait donc à un paramétrage dans la description des documents ou des fiches thesaurus. Ceci n'est d'ailleurs possible et envisageable que parce que les masques des fiches thesaurus et documentaires sont définis dans des fichiers spéciaux. *Il y a indépendance entre les programmes et la description de ces fiches.*

Le même programme, GESECRAN, peut saisir, modifier, consulter des fiches complètement différentes. Un seul point négatif : dans les programmes de consultation, l'affichage des bordereaux ne se fait pas en utilisant GESECRAN. En effet, la taille trop importante de ce dernier n'a pas permis de l'intégrer dans les programmes de consultation. Cependant, on devrait trouver une solution en utilisant le nouveau système d'exploitation.

2. Taille des programmes

La taille trop importante des programmes entraîne des enchaînements de programmes trop nombreux, en particulier en ce qui concerne la gestion du fonds documentaire.

- . Améliorations à apporter :
 - supprimer les initialisations inutiles,
 - empêcher l'utilisateur de modifier le numéro de fiche (ce qui supprimerait les tests sur ce numéro),
 - revoir les références du support et de la date.

- . Certaines parties sont identiques et peuvent être mises en commun :
 - recherche d'un mot-clé dans le lexique,
 - contrôle des descripteurs,
 - gestion des commandes.

- . Le tri et le tassement des adresses dans le fichier INVERSE peuvent se faire en dehors des programmes de gestion du fonds documentaire,
 - soit à la sortie de ces programmes avant de revenir sur la page d'accueil de CATHI,
 - soit dans le programme de réorganisation.

- . On peut également diminuer la taille du tableau des mots-clés résidant en mémoire, mais cela risque d'augmenter le temps de réponse (voir chapitre description technique de la base de données, explication sur l'organisation des fichiers).

- . L'utilisation d'un nouveau système d'exploitation permettra de gagner de la place en mémoire.

- . La sortie sur le marché du GOUPIL 3 disposant de 228 K ou même de 512 K octets de mémoire résoudra bien sûr beaucoup de problèmes.

3. Au niveau de la consultation de la base

Au niveau de la consultation de la base, il faudrait :

- . pouvoir indexer les documents par certains champs (comme ils le sont par les descripteurs), ce qui rendrait possible une recherche multi-critères rapide, qui serait indépendante de celle des mots-clés, ou combinée à cette dernière, selon le choix de l'utilisateur ;
- . pouvoir interroger à partir de mots-clés tronqués ou mal orthographiés. Ceci semble indispensable, surtout quand on songe à la longueur de certains mots-clés !
- . Introduire une recherche par thème, de type arborescent.

4. D'un point de vue technique

- . Ecriture des nombres dans les fichiers : le choix d'une fonction de conversion plutôt qu'une autre pour l'écriture des nombres dans les fichiers a fait perdre de la place sur le disque (voir explications détaillées en annexe) ;
 en adoptant l'autre fonction de conversion, nous aurions pu mettre :
 - 5 enregistrements par secteur dans le fichier DICO, au lieu de 4,
 - 123 adresses de textes dans le fichier INVERSE au lieu de 61.
 Nous aurions également gagné en temps de réponse.
- . L'absence de chaînage avant dans le fichier INVERSE peut être considérée comme une erreur assez importante puisque le tri des adresses dans ce fichier est rendu beaucoup plus difficile.

5. Conclusion

Un certain nombre d'erreurs ont été commises, soit par manque d'information sur le matériel, soit par manque d'expérience dans le domaine des systèmes documentaires, soit par notre précipitation à "programmer". Il faut toutefois signaler qu'au départ, il s'agissait d'un "prototype".

Il faut également souligner des points positifs tels que :

- . des temps de réponse inférieurs à 2 secondes,
- . une certaine facilité d'emploi pour l'utilisateur, en particulier en ce qui concerne la saisie des fiches ou des questions lors d'une recherche.

B I B L I O G R A P H I E

- [1] Bulletin du Centre des Hautes Etudes Internationales d'Informatique Documentaire, 2ème, 3ème et 4ème Trimestres 1982.
- [2] RIGAL Jacqueline, *Le thesaurus en documentation automatique*, Thèse, Université Paris IX Dauphine.
- [3] LE GUENNEC Nicole, *Les bases de données publiques, documentation*, Temps Réel, n° 9.
- [4] MULLER Andrée, *CATHI, Outil de télématique communale*, Interview de Monsieur S. LAVALLEE, Temps Réel, n° 42.
- [5] *Bases et banques de données, Systèmes d'information automatisés*, COM'7, n° 4.
- [6] PORTAL Dominique et ROSSIGNOL Philippe, *BASIS et l'automatisation des bibliothèques*, O1 Informatique, n° 149.
- [7] ESCULIER Christian, *Bases de données, une architecture pour y voir plus clair*, O1 Informatique, n° 154.
- [8] Revue "Documentaliste", vol 19 n° 2 et n° 1 (1982), vol 18 n° 4 et n° 1 (1981)
- [9] DELOBEL Claude, *Les Bases de Données*, Cours ENSIMAG.
- [10] *Initiation à la documentation automatisée*, Université de Picardie à Amiens.
- [11] CHAUFFAUT A. et RIVOIRE R., *Présentation de la méthodologie employée pour l'adaptation de l'application ADM aux spécifications VIDEOTEX*, IRISA, Rennes, Juillet 1981.
- [12] *Accès VIDEOTEX à une banque de données médicales*, Rapport Technique INRIA, Centre de Rennes.

- [13] CHAUFFAUT A. et DRAGONE M., *Note technique de l'interface VIDEOTEX*, IRISA, Centre Interuniversitaire de Calcul de Bretagne, 1982.
- [14] PUJOLLE Guy, *La télématique, réseaux et applications ; chapitre applications télématiques : le VIDEOTEX ; le VIDEOTEX interactif* (Chapitre 3-2), Editions EYROLLES.
- [15] CHAUVEINC Marc, *Le réseau bibliographique et l'accès aux documents*, Collection Systèmes d'Information et de Documentation.
- [16] Bulletins de la DICA, vol 3, n° 1, 2, 9, 10 et vol 2 n° 5.
- [17] Catalogues 1980 et 1982 de prototypes de recherche en logiciel, CNRS.
- [18] *Spécifications VIDEOTEX de visualisation et de codage*, Document diffusé par la DGT.



A N N E X E S

- Annexe 1 - Liste des supports.
- Annexe 2 - Liste des appellations des textes.
- Annexe 3 - Liste des variables de DICO.
- Annexe 4 - Liste des variables de THESUS 1.
- Annexe 5 - Liste des variables de THESUS 2.
- Annexe 6 - Liste des paramètres PARAM.
- Annexe 7 - Liste des variables du fichier ANOMALIE.
- Annexe 8 - Liste des variables des fichiers TEXTE, ANALYSE, MODES.
- Annexe 9 - Sous-programme recherche d'un mot-clé dans le thesaurus.
- Annexe 10 - Gestionnaire d'écran utilisé dans le projet : GESECRAN.
- Annexe 11 - Programme TRI (détails du programme et organigramme).
- Annexe 12 - Programme ORGADICO (détails et organigramme).
- Annexe 13 - Exemples d'anomalies du type 1, 2, 3.
- Annexe 14 - Exemples d'anomalies du type 4, 5.
- Annexe 15 - Modification des fichiers THESUS 1, THESUS 2, PARAM, PILE, dans le programme GESDICO.
- Annexe 16 - Fichier descripteur de la page écran correspondant à une fiche thesaurus dans le programme GESTDICO.
- Annexe 17 - Détails sur la partie CREATION de GESTDICO.
- Annexe 18 - Détails sur la partie SUPPRESSION de GESTDICO.
- Annexe 19 - Détails sur la partie MODIFICATION de GESTDICO.
- Annexe 20 - Détails sur la partie CREATION de la gestion du fonds documentaire (INTEX).
- Annexe 21 - Détails sur la partie MODIFICATION d'une fiche documentaire
Programme GESTEX1
Programme GESTEX2
- Annexe 22 - Critiques techniques du logiciel.

ANNEXE 1 - LISTE DES SUPPORTS

Convention

Si la revue est en un seul mot : *les deux premières lettres du mot servent de code.*

Exemple : les Affiches = AF

Si la revue est en deux mots : *les initiales des deux mots servent de code.*

Exemple : Gazette des Communes = G.C.

Si la revue est en trois mots ou plus : *les initiales des deux derniers mots servent de code.*

Codes et titres de publications

JO	Journal Officiel	On ne compte pas :
NC	JO Numéro complémentaire	. les articles
AN	JO Assemblée Nationale	. bulletin officiel
SE	JO Sénat	
ES	JO Conseil Economique et Social	
AM	Action Municipale	
GC	Gazette des Communes	
PI	AJ PI	
DA	AJ DA	
AF	Les Affiches	
CM	Correspondance Municipale	
FL	Feuillets Lefebvre	
JC	Journal des Communes	
JM	Journal des Maires	
TP	Le Moniteur des Travaux Publics	
RJ	Répertoire de Jurisprudence	
RA	Revue Administrative	
CD	Revue Communale et Départementale	
RC	Revue des Communes	
TT	Technique et Trésor	
GI	BO Direction Générale des Impôts	
CC	BO Concurrence et Consonnation	
LT	BO environnement, urbanisme, logement et transports	
EN	BO Education Nationale	
ID	BO Intérieur et Décentralisation	
IS	BO Isère	
TR	BO Travail	
SS	BO Santé, Sécurité Sociale	
RP	Répertoire Administratif	
VC	Vie Communale	
RF	Revue Fiduciaire	
PM	BO Premier Ministre	
DC	Départements et Communes	
FC	Revue des finances communales.	



ANNEXE 2 - LISTE DES APPELLATIONS DE TEXTES

Convention

Pour les appellations les plus courantes : code alphabétique.

Pour les appellations peu usitées : code chiffré.

LEGISLATION

OR Ordonnance
LO Loi
DE Décrêt
AR Arrêté
CI Circulaire (+ lettre circulaire, note, communiqué, compte-rendu)
IN Instruction
+11 Autres textes officiels.

JURISPRUDENCE

CE Conseil d'Etat
TA Tribunal Administratif
CC Cour de cassation
+21 Jurisprudence autre

DOCTRINE

QE Question écrite
+31 Autre doctrine



ANNEXE 3 - DÉFINITION DES CHAMPS ET LISTE DES VARIABLES DE DICO

Liste des variables de DICO avec leurs caractéristiques

Numéro d'ordre

Longueur maximale : 4 octets.

Désignation dans les fichiers : NO%

Ce numéro indique le numéro d'ordre du mot-clé dans le thesaurus trié.

Lors de la saisie d'une fiche dans le programme de maintenance du thesaurus, une valeur est attribuée automatiquement au numéro d'ordre.

Ce numéro d'ordre est inscrit avec la même valeur dans tous les enregistrements concernant un même mot-clé dans les différents fichiers de la base.

Cette zone n'est jamais à blanc dans DICO, c'est une zone numérique.

Elle n'est modifiée que par programme, dans le programme de réorganisation.

Mot-clé de référence

Maximum 38 octets.

Désignation dans les fichiers : MC%.

Ce mot-clé est la clé d'accès pour toute fiche thesaurus dans tous les programmes

- . zone saisie,
- . obligatoire à la saisie,
- . alphanumérique,
- . jamais à blanc dans le fichier,
- . jamais modifiée ni par programme, ni par transaction écran-clavier,
- . quand on saisit une fiche thesaurus, le programme vérifie que le mot-clé est inexistant ou qu'il a été mis hors service.

Nombre de textes indexés

Maximum 4 octets.

Désignation dans les fichiers : NT%

Indique le nombre de fiches documentaires où le mot-clé est descripteur.

Zone non saisie mais calculée par programme lors de la saisie ou lors de la modification des fiches documentaires.

Zone numérique.

Non obligatoirement remplie si le nombre de textes indexés est égal à 0.
La zone est initialisée à blanc.

N'est modifiée que par programme lors de la saisie ou de la modification d'un texte.

Cette zone permet plusieurs contrôles :

- 1) si VAL (NT%) ≠ 0 alors TY% doit être = 0
- 2) permet de vérifier qu'un mot-clé mis hors service n'indexe pas de texte
 - . ce contrôle est fait dans le programme de gestion du thesaurus au moment où l'utilisateur veut supprimer une fiche ;
 - . il est refait dans le programme général de contrôle ; un message d'erreur est édité à ce moment-là.

Cette zone a donc un double intérêt :

- . informatique, en permettant un certain nombre de contrôles,
- . documentaire : importance du mot-clé en tant que descripteur de textes.

Type de mot-clé

Maximum 1 octet.

Désignation dans les fichiers : TY%

Indique si le mot-clé est susceptible d'être descripteur (0 = descripteur, 1 = non descripteur).

Zone saisie par transaction écran-clavier.

Zone obligatoire à la saisie, jamais à blanc dans le fichier.

Zone numérique, plage de valeurs : 0 ou 1.

N'est modifiable que par transaction écran-clavier lors de la modification d'une fiche thesaurus.

Cette zone fait l'objet de plusieurs contrôles :

- 1) lors de la saisie ou de la modification d'une fiche thesaurus, le type doit être = 1 si le mot-clé est un synonyme, sinon la fiche est rejetée ;
- 2) dans le programme général de contrôle, si un mot-clé indexe au moins un texte et si son type = 1, un message d'erreur est édité.

Adresse de chaînage dans thesus 1

Maximum 5 octets.

Désignation dans les fichiers : ATØ

Permet de chaîner l'enregistrement de DICO à l'enregistrement correspondant de thesus 1.

Zone non saisie, calculée et attribuée par programme lors de la saisie d'une fiche thesaurus.

Zone jamais à blanc, car toute fiche thesaurus a un enregistrement dans thesus 1.

Zone numérique. Jamais modifiée. Ne donne lieu à aucun contrôle. Son rôle est purement informatique.

Adresse dictionnaire du mot-clé principal

Maximum 4 octets.

Désignation dans les fichiers : ADØ.

Sert à aller directement sur l'enregistrement du mot-clé principal quand on a affaire à un mot-clé synonyme (adresse de secteur).

Zone non saisie.

ADØ = " " si le mot-clé n'est pas synonyme. Sinon ADØ contient l'adresse du mot-clé principal si celui-ci existe, sinon ADØ = ZZ (mot-clé principal non trouvé).

Cette zone peut donc être considérée comme alphanumérique, puisqu'elle peut avoir trois sortes de valeurs : " ", ZZ, ou un nombre représentant une adresse.

Zone modifiable uniquement par programme.

Zone permettant de vérifier, lors du programme général de contrôle, qu'un mot-clé synonyme (ADØ ≠ " ") n'est pas un mot-clé lié dans une autre fiche.

Son rôle principal : lors de la recherche, si l'utilisateur tape un mot-clé synonyme, il pourra voir toutes les fiches indexées par le mot-clé principal de ce synonyme ;

mot-clé synonyme $\xrightarrow{\quad}$ mot-clé principal \rightarrow fiches indexées par le mot-clé principal
 \uparrow
 par ADØ

Indice d'enregistrement du mot-clé principal dans son secteur

Longueur : 1 octet.

Désignation : ID \mathcal{X} .

L'adresse (AD \mathcal{X}) de secteur du mot-clé principal est insuffisante pour trouver celui-ci, puisqu'il y a quatre enregistrements par secteur ; ID \mathcal{X} indique donc le numéro d'enregistrement du mot-clé principal dans son secteur.

Zone non saisie.

Si le mot-clé de référence n'est pas un synonyme, ID \mathcal{X} = " "

sinon ID \mathcal{X} contient l'indice d'enregistrement du mot-clé principal (dans son secteur), si celui-ci existe. S'il n'existe pas ID \mathcal{X} = " ".

Plage de valeurs : " ", 0, 1, 2, 3.

Modifiable uniquement par programme.

Ne donne lieu à aucun contrôle.

Son rôle : accéder directement à l'enregistrement du mot-clé principal.

Adresse relative dans le fichier inverse

Longueur maximale : 5 octets.

Désignation dans les fichiers : AI \mathcal{X}

Cette zone indique l'adresse relative de l'enregistrement dans le fichier inverse correspondant au mot-clé.

Cette zone n'est pas saisie. Elle est calculée et attribuée par programme.

Elle n'est pas obligatoirement remplie, elle peut être un blanc si le mot-clé n'indexe aucun texte.

C'est une zone numérique.

Elle ne fait l'objet d'aucun contrôle.

L'adresse n'est pas modifiée, mais elle peut être effacée si le mot-clé n'indexe plus de texte.

Son rôle est uniquement informatique : accéder directement aux adresses des textes indexés par le mot-clé.

Booléen indiquant la mise hors service du mot-clé

Longueur maximale : 1 Octet.

Désignation dans les fichiers : IU%

Cette zone est initialisée par " " lors de la saisie et prend la valeur 1 quand l'utilisateur décide de supprimer un mot-clé, elle est réinitialisée par " " si l'utilisateur décide de recréer le mot-clé mis hors service.

Dans le programme de contrôle, on édite tous les mots-clés qui ont été mis hors service et qui indexent des textes, ou qui apparaissent comme mots-clés liés dans une autre fiche thesaurus.

Cette variable a été introduite pour éviter l'effacement dans le lexique d'un mot-clé que l'utilisateur veut supprimer, mais qui est descripteur de textes ou présent dans le thesaurus en tant que mot-clé lié.

Définition de la variable Appellation	Désignation dans les programmes	Longueur maximale en octets dans le fichier	Variable saisie = S Variable gérée par programme = NS	Zone obligatoire en saisie = O Non obligatoire à la saisie = NO	Valeur d'initialisation	Existence d'une plage de valeurs avec différentes valeurs possibles	Zone alphanumérique = A Zone numérique = N	Variable modifiable = M Non modifiable = NM	Modifiable par transection écran-clavier = T ; par programme = P	Programmes où la zone est susceptible d'être modifiée
Mot-clé de référence	MCZ	38	S	O	" "	Non	A	NM	jamais modifié	aucun
Numéro d'ordre	NOZ	4	NS		" "	> 0	N	M	P	ORGADICO
Nombre de textes indexés	NTZ	4	NS		" "	"" si nbre=0 sinon nbre≥1	N	M	P	GESTEXT ou DREATEXT
Type de mot-clé	TYZ	1	S	O	" "	0 ou 1	N	M	T	GESTDICO
Adresse de chaînage dans Thesus 1	ATZ	5	NS		" "	≥ 1	N	NM	jamais modifié	aucun
Adresse relative dans DICO du mot-clé principal correspondant	ADZ	4	NS		" "	"" si non synonyme sinon = "zz" ou adresse	A	M	P	ORGADICO ou GESTDICO
Adresse relative dans le fichier inverse	AIZ	5	NS		" "	"" si indexe aucun texte sinon adresse	N	M	P	GESTEXT
Indice d'enregistrement du mot-clé principal de son secteur	IDZ	1	NS		" "	0 1 2 3	N	M	P	ORGADICO ou GESTDICO
Booléen indiquant la mise hors service du mot-clé	IUZ	1	NS		" "	0 ou 1	N	M	P	GESTDICO

ANNEXE 4 - LISTE DES VARIABLES DE THESUS 1

Liste des variables dans THESUS 1 et leurs caractéristiques

Numéro d'ordre

Se reporter à NOZ dans DICO).

Désignation : NOZ

Longueur maximale : 4 octets.

Même signification et même valeur que NOZ dans DICO.

Cette variable n'est modifiée que par programme et en même temps que NOZ.

C'est une zone numérique toujours remplie.

Note d'application

Désignation : première ligne : N1Z

deuxième ligne : N2Z

troisième ligne: N3Z

quatrième ligne: N4Z

La note d'application n'a pas été désignée par une seule variable, car le gestionnaire d'écran utilisé l'interdisait.

Longueur maximale : 4 x 38 octets.

Les quatre zones de la note d'application sont saisies et modifiées par transaction écran-clavier, jamais directement par programme.

Chacune des quatre zones peut être vide et ne fait l'objet d'aucun contrôle.

Ce sont des zones alphanumériques.

Nombre de mots-clés liés

Désignation : NMZ

Longueur maximale : 2 octets.

Désigne le nombre de mots-clés liés du mot de référence.

C'est une zone numérique entière dont la valeur est ≥ 0 et ≤ 8 ; cette valeur est calculée et modifiée par programme lors de la saisie ou de la modification d'une fiche thesaurus ; elle permet de vérifier qu'un mot-clé synonyme n'a qu'un mot-clé lié.

Premier code liaison

Désignation : C1%

Longueur maximale : 1 octet.

Désigne le code liaison du premier mot-clé lié.

C'est une zone numérique qui peut être vide (dans ce cas, C1% est initialisé par " "), sinon elle prend une des valeurs 1, 2, 3, 4.

C'est une zone saisie et modifiée uniquement par transaction écran-clavier

Dans le cas où le mot-clé de référence est un synonyme, C1% = "1" et le nombre de mots-clés liés (NM%) doit être égal à 1 ; ce contrôle est fait lors de la saisie ou de la modification d'une fiche thesaurus.

Premier mot-clé lié

Désignation : M1%

Longueur maximale : 38 octets.

Désigne le premier mot-clé lié.

Zone alphanumérique qui peut être vide et qui est saisie, modifiée, gérée par l'utilisateur, par transaction écran-clavier.

Lors de la saisie ou de la modification d'une fiche, on vérifie si le premier mot-clé lié existe dans le thesaurus, et s'il n'a pas été mis hors service.

Deuxième code liaison

Désignation : C2%

Longueur maximale : 1 octet.

Désigne le deuxième code liaison du deuxième mot-clé lié.

Zone numérique qui peut être vide (dans ce cas C2% = " "), sinon elle prend une des valeurs 1, 2, 3, 4.

Zone saisie, modifiée, par transaction écran-clavier.

Mise à part l'appartenance à la plage de valeurs, cette zone ne fait pas l'objet d'un contrôle spécifique.

Deuxième mot-clé lié

Désignation : M2X

Longueur maximale : 38 octets.

Désigne le deuxième mot-clé lié.

Zone alphanumérique qui peut être vide et qui est saisie, modifiée, gérée par l'utilisateur, par transaction écran-clavier.

Lors de la saisie ou de la modification d'une fiche, on vérifie si le deuxième mot-clé lié existe dans le thesaurus et s'il n'a pas été mis hors service.

Nombre de liaisons vers le mot-clé de référence

Désignation : NLX

Indique le nombre de mots-clés renvoyant sur le mot-clé de référence.

Longueur maximale : 4 octets.

Zone numérique, non saisie, entièrement gérée par programme.

Cette zone peut être vide, elle est initialisée par " ", sinon elle prend une valeur entière ≥ 1 .

Sa valeur est attribuée et recalculée après chaque tri dans le programme de réorganisation du thesaurus.

Cette zone permet un contrôle indirect : un message prévient l'utilisateur qui désire mettre hors service un mot-clé existant en tant que mot-clé lié dans une autre fiche thesaurus (dans ce cas NLX = " "). Ce contrôle est refait dans le programme général de contrôle ; à ce moment un message d'erreur est édité.

Booléen indiquant si le mot-clé de référence est un mot-clé principal ou non

Désignation : MMX

Longueur maximale : 1 octet.

Indique si le mot-clé de référence possède au moins un synonyme.

Zone numérique qui peut être vide ; dans ce cas MMX = " ", sinon la seule valeur possible est 1.

Cette zone est entièrement gérée par programme, et n'est modifiée que dans le programme de réorganisation.

Son rôle est surtout informatique, elle facilite un certain nombre de contrôles.

Adresse relative de chaînage dans thesus 2

Désignation : CHØ

Longueur maximale : 5 octets.

Cette adresse permet de chaîner l'enregistrement dans thesus 1 avec celui de thesus 2 quand il existe.

C'est une zone numérique qui peut être vide ; dans ce cas elle est initialisée par " " s'il n'y a pas d'enregistrement dans thesus 2, c'est-à-dire si le nombre de mots-clés liés est ≤ 2 .

Dans le cas contraire, le programme attribue automatiquement une valeur à cette zone.

Cette zone peut être modifiée, toujours par programme, si le nombre de mots-clés liés passe d'une valeur ≤ 2 à une valeur > 2 , ou inversement.

Cette zone ne fait pas l'objet d'un contrôle spécifique ; son rôle est purement informatique.

Définition de la variable appellation	Désignation dans les programmes	Longueur maximale en octets dans le fichier	Variable saisie : S Gérée par programme * NS	Zone obligatoire en saisie : O Non obligatoire en saisie : NO	Zone obligatoire-ment remplie : OR	Valeur d'initialisation	Existence d'une plage de valeurs avec différentes valeurs possibles	Zone alphanumérique : A Zone numérique : N	Variable modifiable : M Non modifiable : NM	Modifiable par tranche section écran clavier : T Programme : P	Programmes où le variable est susceptible d'être modifiée
Numéro d'ordre	NDZ	4	NS		OR	*	≤ 1	N	M	P	ORGADICO
Note d'application : 1ère ligne	N1Z	38	S	NO		*		A	M	T	GESTDICO
Note d'application : 2ème ligne	N2Z	38	S	NO		*		A	M	T	GESTDICO
Note d'application : 3ème ligne	N3Z	38	S	NO		*		A	M	T	GESTDICO
Note d'application : 4ème ligne	N4Z	38	S	NO		*		A	M	T	GESTDICO
Nombre de mots-clés liés	NMZ	2	NS		OR	*	≥ 0 ≤ 8	N	M	P	GESTDICO
Premier code liaison	C1Z	1	S	NO		*	** 1, 2, 3, 4	N	M	T	GESTDICO
Premier mot-clé lié	M1Z	38	S	NO		*		A	M	T	GESTDICO
Deuxième code liaison	C2Z	1	S	NO		*	** 1, 2, 3, 4	N	M	T	GESTDICO
Deuxième mot-clé lié	M2Z	38	S	NO		*		A	M	T	GESTDICO
Nombre de liaisons vers le mot-clé de référence	NI Z	4	NS			*	** ou nombre ≥ 1	N	M	P	ORGADICO
Booléen indiquant si le mot-clé de référence est un mot-clé principal ou non	MMZ	1	NS			*	** ou 1	N	M	P	ORGADICO
Adresse relative de chaînage dans Thesus 2	CI Z	5	NS			*	** ou un nombre (adresse relative)	N	M	P	GESTDICO



ANNEXE 5 - LISTE DES VARIABLES DE THESUS 2

Liste des variables dans thesus 2 et leurs caractéristiques

Le numéro d'ordre

Désignation : DDØ

Comme pour NDØ, DDØ a en permanence la même valeur et les mêmes caractéristiques que NDØ (numéro d'ordre dans DICO). Se reporter à NDØ dans DICO.

Code liaison n° I

I = 3 à 8.

Désignation : CIØ

Se reporter au deuxième code liaison (voir thesus 1).

Les codes liaisons du deuxième au huitième ont les mêmes caractéristiques.

Mot-clé lié n° I

I = 3 à 8.

Désignation : MIØ

Se reporter au premier mot-clé lié (voir thesus 1).

En effet, les mots-clés liés du premier au huitième ont les mêmes caractéristiques.

Définition de la variable Appellation	Designation dans les programmes	Longueur maximale en octets dans le fichier	Variable saisie = S Gérée par programme = NS	zone obligatoire en saisie = O zone non obligatoire en saisie = NO	Zone obligatoirement remplie = OR	Valeur d'initialisation	Existence d'une plage de valeurs avec différentes valeurs possibles	zone alphanumérique = A zone numérique = N	Variable modifiable = M non modifiable = NM	Modifiable par transaction écran-clavier : T	ORGADICO	Programmes où la variable est susceptible d'être modifiée
Numéro d'ordre	DDZ	4	NS		OR	" "	1	N	M	P	ORGADICO	
3ème code liaison	C3Z	1	S	NO		" "	"", 1, 2, 3, 4	N	M	T	GESTDICO	
3ème mot-clé lié	M3Z	38	S	NO		" "		A	M	T	GESTDICO	
4ème code liaison	C4Z	1	S	NO		" "	"", 1, 2, 3, 4	N	M	T	GESTDICO	
4ème mot-clé lié	M4Z	38	S	NO		" "		A	M	T	GESTDICO	
5ème code liaison	C5Z	1	S	NO		" "	"", 1, 2, 3, 4	N	M	T	GESTDICO	
5ème mot-clé lié	M5Z	38	S	NO		" "		A	M	T	GESTDICO	
6ème code liaison	C6Z	1	S	NO		" "	"", 1, 2, 3, 4	N	M	T	GESTDICO	
6ème mot-clé lié	M6Z	38	S	NO		" "		A	M	T	GESTDICO	
7ème code liaison	C7Z	1	S	NO		" "	"", 1, 2, 3, 4	N	M	T	GESTDICO	
7ème mot-clé lié	M7Z	38	S	NO		" "		A	M	T	GESTDICO	
8ème code liaison	C8Z	1	S	NO		" "	"", 1, 2, 3, 4	N	M	T	GESTDICO	
8ème mot-clé lié	M8Z	38	S	NO		" "		A	M	T	GESTDICO	

voir NDZ dans DICO

pour les codes liaison

voir C2Z (THESUS 1)

pour les mots-clés liés

voir M2Z (THESUS 2)

ANNEXE 6 - LISTE DES PARAMÈTRES DU FICHIER PARAM

PR%(0)

Initialisée à 1 dans le programme de création du thesaurus, cette variable indique que le fichiers de données DICO a bien été créé ; PR%(0) n'est plus jamais modifié par la suite.

PR%(1)

Indice d'enregistrement dans son secteur du dernier mot-clé de DICO. Il peut donc prendre une des quatre valeurs : 0, 1, 2, 3, puisqu'il y a 4 enregistrements par secteur dans DICO. Sa valeur est calculée et modifiée dans les programmes de création et de maintenance du thesaurus, ainsi que dans le programme de réorganisation ORGADICO, puisque certains enregistrements de DICO peuvent ne pas être recopiés (ce qui correspond à des suppressions).

PR%(2)

Adresse relative (correspond au secteur) du dernier mot-clé de DICO. C'est donc un entier ≥ 1 . Il permet un accès direct sur le secteur contenant le dernier mot-clé de DICO. Sa valeur est calculée et modifiée dans les programmes de création et de maintenance du thesaurus, ainsi que dans le programme de réorganisation ORGADICO en même temps que PR%(1).

N.B. Lors d'une suppression dans le programme de maintenance, PR%(1) et PR%(2) ne sont pas modifiées, car la suppression, virtuelle dans ce programme, n'aura effectivement lieu que dans le programme de réorganisation, sous certaines conditions (nombre de textes indexés = 0, nombre de liaisons vers le mot-clé = 0).

PR%(3)

Adresse relative du dernier enregistrement de THESUS 2. C'est un entier ≥ 0 . Sa valeur est calculée et modifiée uniquement dans les programmes de création et de maintenance du thesaurus ; celle-ci ne peut être qu'incrémentée ;

en effet, quand il y a suppression d'un enregistrement dans thesus 2, l'adresse libérée est récupérée dans le tableau PILE pour être réutilisée lors de la création d'un nouvel enregistrement dans ce fichier ; quand il n'y a plus d'adresse à récupérer, on met le nouvel enregistrement à la fin du fichier en incrémentant PR%(3).

PR%(4)

Adresse relative du dernier enregistrement de thesus 1. C'est un entier ≥ 1 . Sa valeur est calculée et modifiée uniquement dans les programmes de création et de maintenance du thesaurus. Pour les mêmes raisons que pour PR%(3), la valeur de PR%(4) ne diminue jamais.

PR%(5)

Nombre d'enregistrements dans DICO. PR%(5) est incrémenté à la création d'un nouvel enregistrement dans DICO, dans les programmes de création et de maintenance du thesaurus (PR%(5) n'est pas incrémenté lors d'une re-création). PR%(5) est recalculé dans le programme de réorganisation, puisque c'est dans ce programme que s'effectuent les suppressions d'enregistrements (comme pour PR%(2), PR%(5) n'est pas décrémenté lors d'une suppression dans le programme de maintenance, puisque cette suppression est virtuelle).

PR%(6)

Dimension de TABLEAU. Nous reviendrons sur ce paramètre en décrivant TABLEAU. La valeur affectée à PR%(6) est attribuée et modifiée uniquement dans le programme de réorganisation et dans aucun autre programme.

PR%(7)

Adresse relative (correspond au secteur) du dernier mot-clé de la partie triée de DICO. En effet, nous avons vu que le fichier DICO n'est pas trié en permanence. Entre deux tris, les nouveaux mots-clés sont mis à la fin du fichier, si bien que DICO comporte une partie triée et une partie non triée.

La valeur affectée à PR%(7) est attribuée et modifiée uniquement dans le programme de réorganisation et dans aucun autre programme. Après un tri, nous avons bien entendu PR%(7) = PR%(2).

PR%(8)

Indice d'enregistrement dans son secteur du dernier mot-clé de la partie triée de DICO. PR%(8) peut prendre une des quatre valeurs 0, 1, 2, 3 .
Même remarque que pour PR%(7), la valeur affectée à PR%(8) est attribuée et modifiée uniquement dans le programme de réorganisation après le TRI. A ce moment PR%(8) = PR%(1).

PR%(9)

Indique le nombre de créations d'enregistrements dans DICO depuis le dernier tri. Ce paramètre a été créé parce qu'on pensait lancer le tri automatiquement dès que PR%(9) atteindrait une certaine valeur, pour éviter que DICO ne comporte une partie non triée trop importante, ce qui pénalise en temps la recherche d'un mot-clé dans le fichier. Après quelques mois d'expérience, il semble que l'utilisateur préfère lancer le TRI lui-même à la fin d'une journée de travail, dans ce cas le paramètre PR%(9) perd de son utilité. PR%(9) n'est pas incrémenté lors d'une re-création.

PR%(10)

Nombre d'enregistrements libres (à récupérer) dans thesus 1. PR%(10) est incrémenté chaque fois qu'un enregistrement de thesus 1 est effacé lors de la suppression d'un mot-clé du thesaurus dans le programme de réorganisation. Par contre, PR%(10) est décrémenté chaque fois qu'un récupère un de ces enregistrements lors de la création d'un nouveau mot-clé dans le programme de maintenance. En dehors des deux programmes de maintenance et de réorganisation du thesaurus, PR%(10) n'est jamais modifié.

PR%(11)

Nombre d'enregistrements libres (à récupérer) dans thesus 2. PR%(11) est incrémenté chaque fois qu'un enregistrement de thesus 2 est effacé, ce qui se produit :

- . soit dans le programme de réorganisation quand certains mots-clés sont effectivement supprimés dans les fichiers,
- . soit dans le programme de maintenance, lors de la modification d'une fiche thesaurus (si le nombre de mots-clés qui était > 2 devient ≤ 2). Inversement PR%(11) est décrémenté chaque fois qu'on récupère un de ces enregistrements, ce qui a lieu uniquement dans le programme de maintenance du thesaurus. Comme PR%(10), en dehors des deux programmes de maintenance et de réorganisation du thesaurus, PR%(11) n'est jamais modifié.

PR%(12)

Le plus grand numéro d'ordre dans le thesaurus. Après chaque tri de DICO, les numéros d'ordre sont recalculés ; ils correspondent aux numéros d'ordre des mots-clés dans le fichier trié. Dans le programme de maintenance, lors de toute nouvelle création de mot-clé, un numéro d'ordre est attribué automatiquement : $PR\%(12) + 1$; en même temps PR%(12) est incrémenté (PR%(12) n'est pas incrémenté lors d'une re-création). On voit que PR%(12) est calculé et modifié uniquement dans les programmes de maintenance et de réorganisation du thesaurus.

PR%(13), PR%(14), PR%(15)

sont inutilisés ; on s'en servira si on a besoin de nouveaux paramètres pour le thesaurus.

PR%(16)

Nombre d'enregistrements vacants dans le fichier TEXTE ; PR%(16) indique donc le nombre d'adresses à récupérer dans le fichier TEXTE, ces adresses sont stockées dans la première colonne du fichier PILTEX (P1%(1,62)).

PR%(17)

Nombre d'enregistrements vacants dans le fichier ANALYSE. PR%(17) indique le nombre d'adresses à récupérer dans le fichier ANALYSE. Ces adresses sont stockées dans la deuxième colonne du fichier PILTEX (P1%(2,62)).

PR%(18)

Nombre d'enregistrements vacants dans le fichier MODE. PR%(18) indique le nombre d'adresses libres à récupérer dans le fichier MODES ; ces adresses sont stockées dans la troisième colonne du fichier PILTEX (P1%(3,62)).

PR%(19)

Nombre d'enregistrements vacants dans le fichier INVERSE. PR%(19) indique le nombre d'adresses libres à récupérer dans le fichier INVERSE ; ces adresses sont stockées dans la quatrième colonne du fichier PILTEX (P1%(4,62)).

PR%(20)

Adresse du dernier enregistrement du fichier TEXTE.

PR%(21)

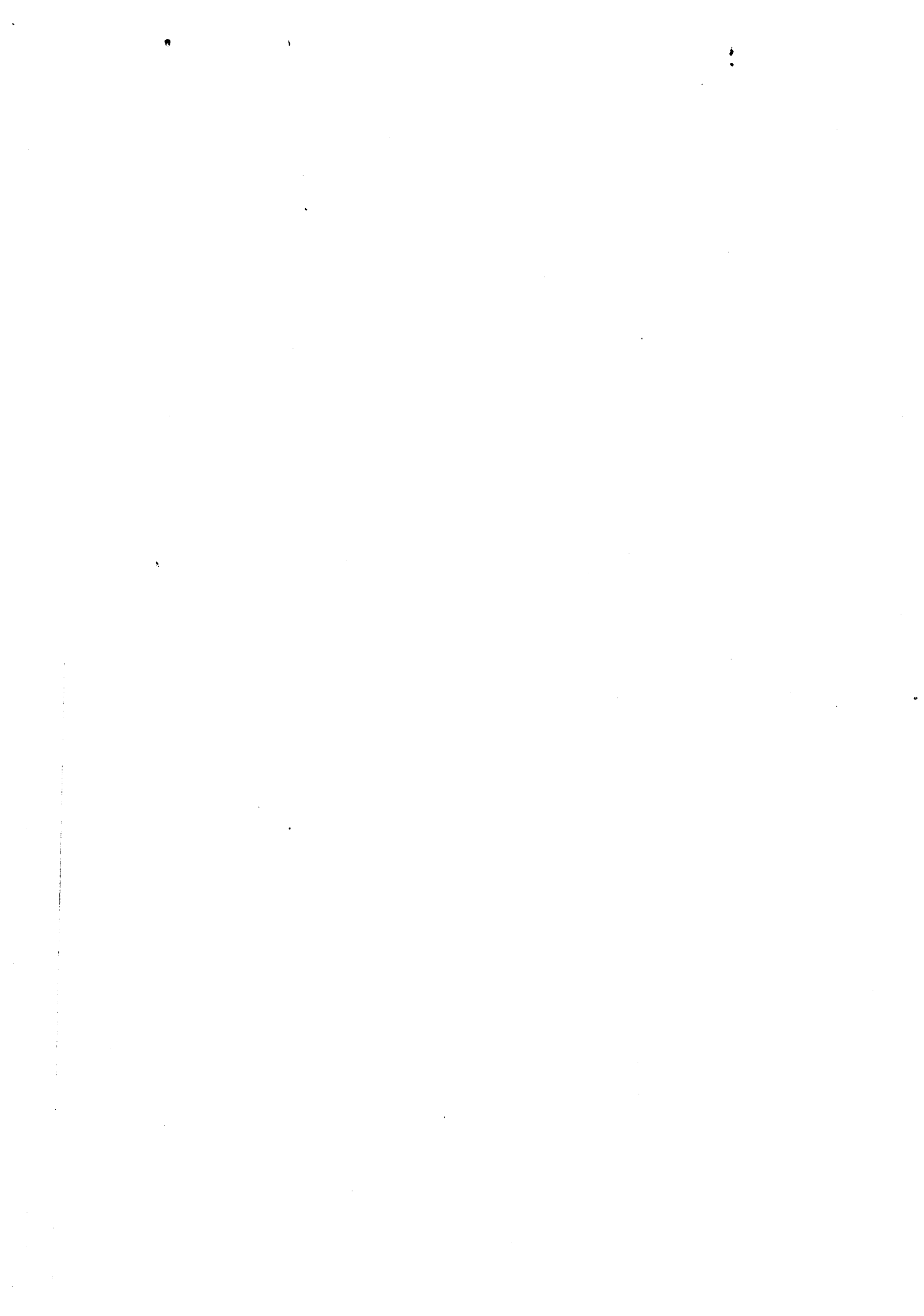
Adresse du dernier enregistrement du fichier ANALYSE.

PR%(22)

Adresse du dernier enregistrement du fichier MODES.

PR%(23)

Adresse du dernier enregistrement du fichier INVERSE.



ANNEXE 7 - LISTE DES VARIABLES DU FICHIER ANOMALIE

Mot-clé

Désignation : MA $\%$.

Longueur maximale : 38 octets.

MA $\%$ est soit le nom du mot-clé de référence, soit le nom d'un mot-clé lié. C'est une zone alphanumérique, qui peut être vide (c'est le cas si plusieurs types d'erreurs existent ; on n'édite qu'une seule fois le mot-clé de référence).

Numéro d'ordre du mot-clé

Désignation : NA $\%$.

Longueur maximale : 4 octets.

NA $\%$ est le numéro d'ordre du mot-clé écrit dans MA $\%$.

C'est une zone numérique, mais qui peut être vide si MA $\%$ contient le nom d'un mot-clé lié, non encore saisi, ou si MA $\%$ est une zone vide.

Code erreur

Désignation : CD $\%$.

Longueur : 1 octet.

CD $\%$ est le code erreur correspondant à l'anomalie écrite dans ME $\%$ (il y a une exception : CD $\%$ prend la valeur 3 alors que ME $\%$ = " " ; c'est une astuce de programmation pour faire imprimer le mot-clé de référence quand il y a une erreur de type 1 ou 2).

C'est une zone numérique entière (≥ 1 , ≤ 5) qui n'est jamais vide.

Message d'erreur

Désignation : ME $\%$.

Longueur maximale : 109 octets.

ME $\%$ contient un des cinq messages d'erreur suivants :

- . (1) : mot lié non encore saisi,
- . (2) : incohérence, mot synonyme de Y,
- . (3) : incohérence entre type et liaison texte,
- . (5) : ce mot-clé est présent dans X fiches thesaurus,
- . (5) : ce mot-clé est présent dans X fiches thesaurus, il a des synonymes,
- . (4) : on ne peut supprimer ce mot-clé qui indexe les textes suivants :

Le chiffre écrit entre parenthèses est la valeur correspondante affectée à CD%. .

Cette zone peut être vide (quand on veut imprimer le mot-clé de référence quand le type d'erreur est 1 ou 2).

Remarque :

Ces variables ne sont jamais modifiées, puisque le fichier ANOMALIE est détruit puis recréé à chaque réorganisation ; elles sont entièrement gérées par programme.

En outre, ce fichier étant destiné à l'édition, un certain nombre d'astuces expliquent pourquoi certaines zones sont vides.

ANNEXE 9 - SOUS-PROGRAMME : RECHERCHE D'UN MOT-CLÉ DANS LE THESAURUS

Ce sous-programme est commun à tous les programmes du projet ; son rôle est de rechercher un mot-clé dans DICO.

Le mot-clé recherché est préalablement écrit dans la variable MR%.
 EX% est le booléen indiquant si le mot-clé a été trouvé ou non :

EX% = 0 si MR% ∈ DICO

EX% = 1 si MR% ∉ DICO

AD% contient l'adresse relative de MR% dans DICO.

IE% contient l'indice d'enregistrement de MR% dans son secteur.

Si le mot-clé a été trouvé, la tête de lecture est positionnée sur son enregistrement dans DICO quand on sort du sous-programme.

Le fichier TABLEAU a été créé pour permettre une recherche plus rapide d'un mot-clé ; ce fichier est écrit en mémoire au début de chaque programme dans la table T%. La dimension de T% est PR%(6).

Méthode de recherche de MR% dans DICO

Entre deux tris, DICO comporte une partie triée et une partie non triée. On commence à rechercher MR% dans la partie triée de DICO, si MR% n'est pas trouvé, on le recherche dans la partie non triée.

Si DICO n'a jamais été trié, ou si MR% > au dernier mot-clé de la partie triée de DICO (MR% > T% (PR%(6))), la recherche se fait directement dans la partie non triée.

a) Recherche dans la partie triée de DICO

On compare MR% à chaque mot de T%. Il existe un plus petit indice : K% tel que T% (K% - 1) < MR% ≤ T% (K%).

- . Si MR% = T% (K%) alors MR% ∈ DICO, on se positionne sur MR% dans DICO, EX% = 0, on calcule AD% et IE% et on sort du sous-programme ;
- . si MR% ≠ T% (K%), on compare MR% à tous les mots-clés de DICO compris entre TS (K% - 1) et TS (K%) ; cette recherche est séquentielle sur cinq secteurs
 - si MR% est trouvé (EX% = 0, on calcule AD% et IE% et on sort du sous-programme,
 - si MR% n'est pas trouvé, on va le rechercher dans la partie non triée de DICO.

b) Recherche dans la partie non triée

C'est une recherche séquentielle.

Il est souhaitable que cette partie soit la plus petite possible, donc que le tri de DICO s'effectue dès qu'un nombre de mots-clés sont rajoutés.

Si MR% n'a pas été trouvé, alors EX% = 1.

Sinon EX% = 0, on calcule AD% et IE% et on sort du sous-programme.

Légende de la figure qui suit :

EX% = 0 si MR% ∈ DICO

EX% = 1 si MR% ∉ DICO

AD% = adresse relative de MR% dans DICO

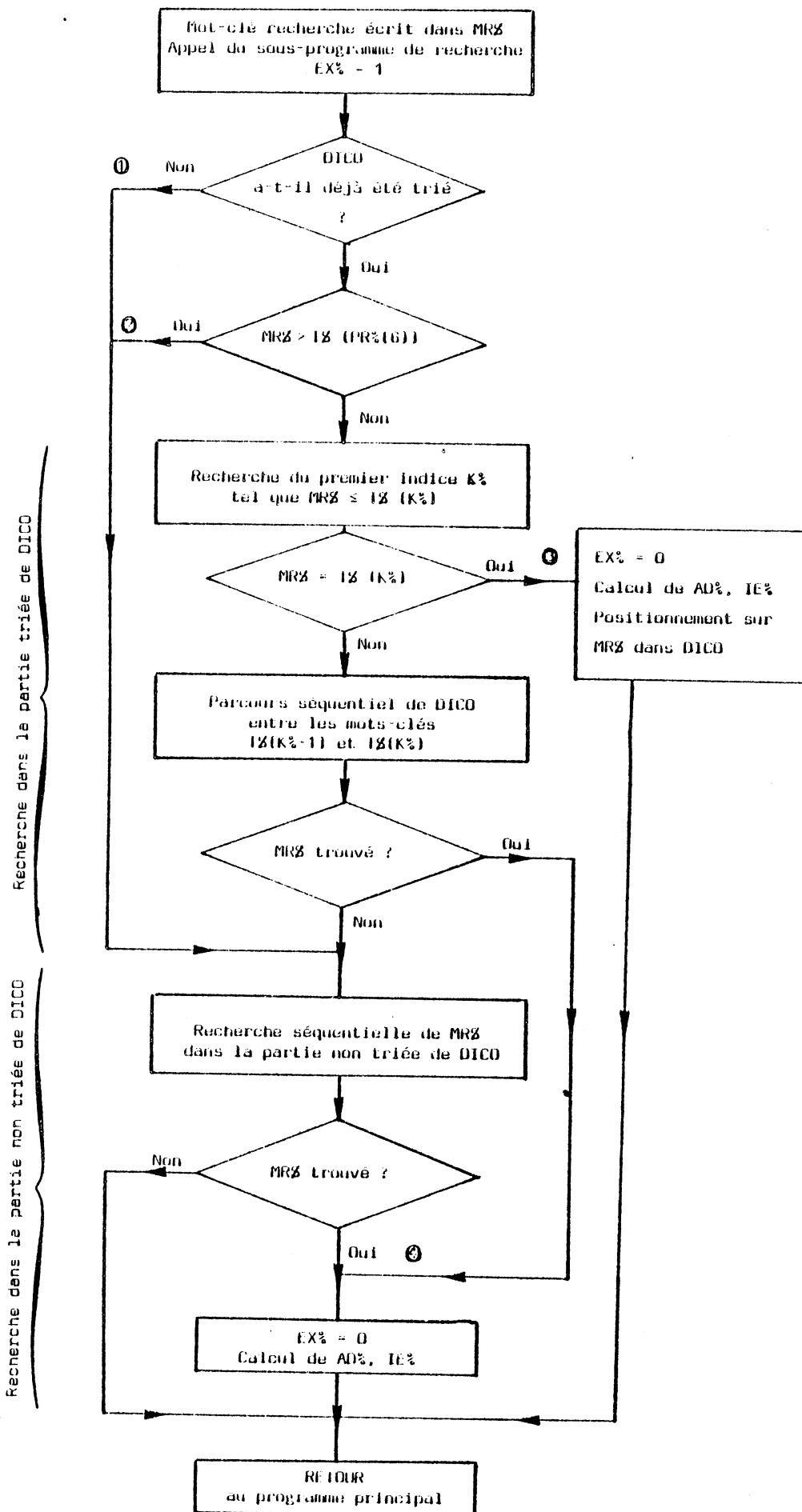
IE% = indice d'enregistrement de MR% dans son secteur.

3 si MR% ∈ T%, il faut se positionner sur MR% dans DICO.

4 si MR% ∈ DICO, mais MR% ∉ T%, on est positionné sur MR%.

1 si DICO n'a jamais été trié, la table T% est vide, il n'y a pas de partie triée ; on fait directement la recherche dans la partie non triée

2 si MR% > TS (PR%(6)) MR% ne se trouve pas dans la partie triée de DICO, on fait directement la recherche dans la partie non triée.



Organigramme du sous-programme :
recherche de MR% dans DICO



ANNEXE 10 - GESTIONNAIRE D'ÉCRAN UTILISÉ DANS LE PROJET : GESECRAN

Dans l'ensemble du projet, pour saisir, consulter, modifier les fiches thesaurus ou les fiches documentaires, on utilise le gestionnaire d'écran GESECRAN, qui est un produit SMT sur GOUPIL. Ce gestionnaire permet de modifier les différentes zones des fiches documentaires ou thesaurus, par simple déplacement du curseur ; il effectue aussi un certain nombre de contrôles (type d'information, contrôle de fourchette, appartenance à une liste de valeurs, nombre minimal et maximal de caractères à saisir, etc ..). Ces facilités offertes à l'utilisateur le rapprochent d'un logiciel de traitement de texte. Cependant, il occupe une place mémoire non négligeable (6 K). Il a aussi le gros inconvénient de ne pas permettre des contrôles supplémentaires zone par zone. Il faut que la saisie (modification) globale de la fiche soit terminée pour que les contrôles indirects propres à l'application puissent s'effectuer. C'est pourquoi lors de la modification ou de la consultation d'un fichier thesaurus, le mot-clé de référence est saisi en rubrique isolée.

Utilisation de GESECRAN

L'utilisateur définit l'image des différentes zones de chaque page écran dans un *fichier descripteur*. Ce fichier est :

DESDICO.DAT pour les fiches thesaurus

ECRAN.DAT pour le bordereau n° 1 des fiches documentaires

ECRAN.DAT pour le bordereau n° 2 des fiches documentaires.

Les fichiers descripteurs sont créés, modifiés et listés par le programme DESCRIPT.BAS.

Chaque rubrique est composée de deux parties :

- . La première partie regroupe les informations obligatoires :
PP, PR, I, NB, NO, TM, TS
- PP : numéro de la page écran (2 caractères)
- PR : numéro de la rubrique dans la page écran (2 caractères)
- I : type d'information à saisir (1 caractère)

NB : nombre de caractères à saisir (2 caractères)
NO : numéro de ligne de l'écran de la saisie (2 caractères)
TN : tabulation du message (2 caractères)
TS : tabulation de la saisie (2 caractères)

- . La deuxième partie regroupe les informations facultatives :
 - libellé à afficher,
 - zone de saisie obligatoire ou non,
 - valeur par défaut (retour charriot immédiat),
 - longueur minimale de saisie,
 - contrôles de fourchette,
 - liste d'appartenance.

Le gestionnaire d'écran possède 4 fonctions :

- . première saisie d'une page, à l'appel de la fonction ZM% = 0
- . consultation/modification d'une page, à l'appel de la fonction ZM% = 1
- . consultation d'une page à l'appel de la fonction ZM% = 2
- . saisie d'une rubrique isolée : dans ce cas la page n'a pas à être spécifiée.
Z% contient le numéro de la rubrique dans le fichier descripteur.

ANNEXE 11 - PROGRAMME TRI

Le programme de tri utilisé est le programme SORTMERGE commercialisé par la SMT, la CEDIS, ... sur le GOUPIL. (SORTMERGE COPYRIGHT 1979, by Technical Systems Consultants Inc., PO Box 2570, West Lafayette, Indiana 47906).

Le fichier DICO est trié par ordre alphabétique sur les mots-clés.

Le fichier TRIDICO.BIN contient toutes les spécifications relatives au TRI. Ce fichier est créé par le programme SORT.

Le programme de tri proprement dit, PSORT, prend la place de l'interpréteur BASIC ; on utilise la procédure EXEC pour lancer le tri, recharger le BASIC et lancer le programme suivant.

Le fichier TRIDICO.TXT contient le nom des commandes ou programmes qui se succèdent :

```
TRIDICO.TXT = 1.PSORT|_|1.TRIDICO.BIN|_|1.DICO.DAT
              XBG64|_|1.ORGADICO.BAC
```

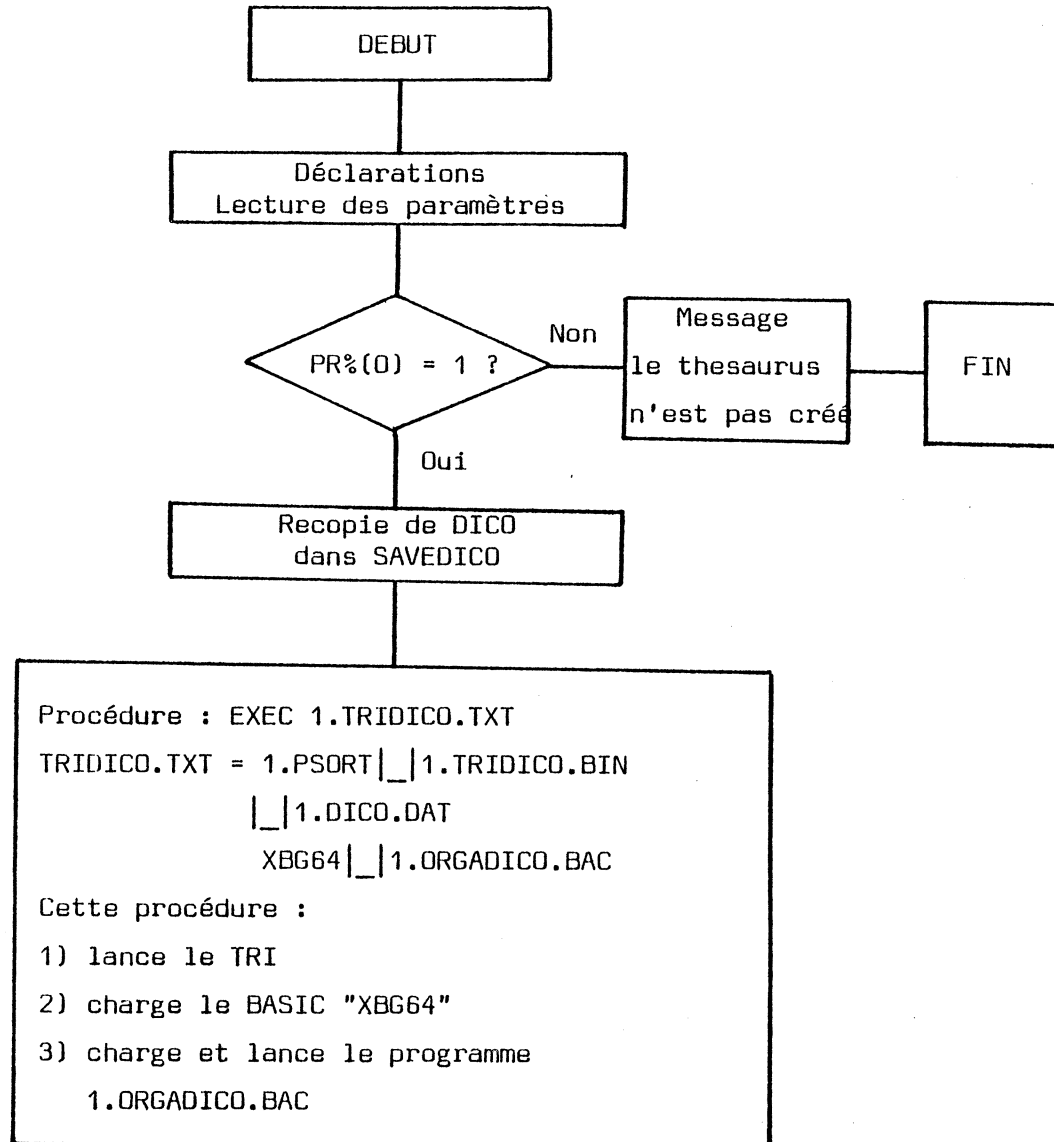
1.PSORT|_|1.TRIDICO.BIN|_|1.DICO.DAT : cette commande lance le tri PSORT sur le fichier DICO en suivant les spécifications contenues dans 1.TRIDICO.BIN. XBG64 est le nom du BASIC qui sera chargé après le tri.

1.ORGADICO.BAC désigne le programme qui succède à TRI.

Par mesure de sécurité, avant le lancement du TRI, le fichier DICO est recopié dans un fichier SAVEDICO qui est donc la dernière version de DICO juste avant le TRI.

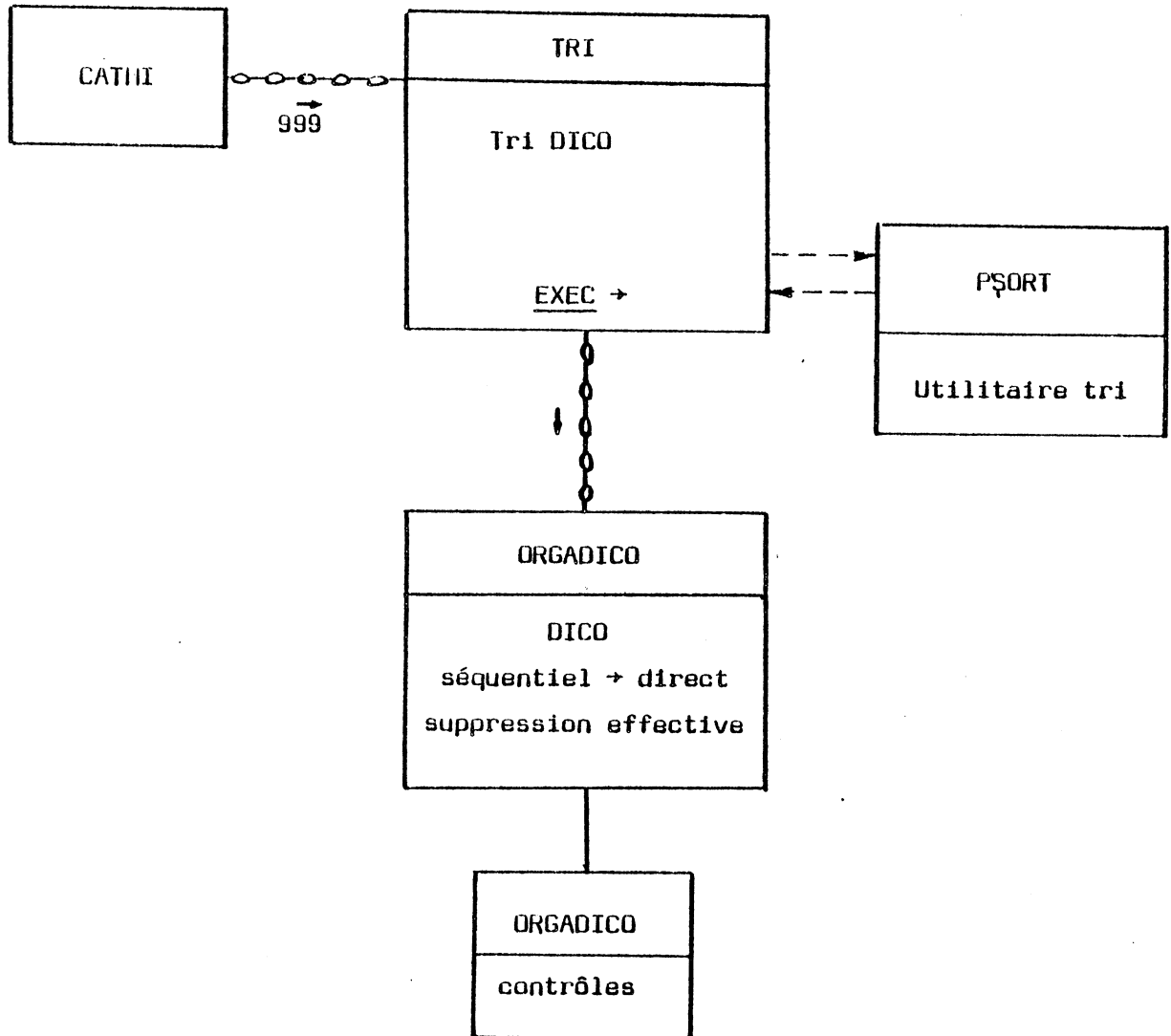
Le programme PSORT ne touche pas au fichier DICO, mais le recopie en le triant dans un autre fichier, le fichier DICOTRIE ; ce dernier est un fichier binaire et séquentiel.

Organigramme du programme TRI :



N.B. La valeur 1 est affectée à PR%(0) lors de la création du thesaurus.

Programmes TRI + CONTROLE : TRI - ORGADICO (ou ORGAEDIT)



Source : ICARE.



ANNEXE 12 - PROGRAMME ORGADICO

Première partie du programme ORGADICO

Le TRI n'a pas modifié le fichier DICO mais l'a recopié en le triant dans le fichier DICOTRIE, il faut donc recopier celui-ci dans DICO en respectant la structure interne de DICO (fichier accès direct, quatre enregistrements par secteur).

Les mots-clés à éliminer ne sont pas recopiés dans DICO ; on efface en même temps leurs enregistrements dans THESUS1 et THESUS2 et on récupère les adresses de ces enregistrements vacants dans PILE.

Pour les autres mots-clés, on met à jour les numéros d'ordre :

NO $\%$ dans DICO

ND $\%$ dans THESUS1

DD $\%$ dans THESUS2

AZ $\%(0)$ dans INVERSE

et on initialise les variables NL $\%$ et NM $\%.$

On recrée le fichier TABLEAU après l'avoir détruit : ce fichier est construit en prenant un mot-clé tous les 20 mots-clés (le nombre 20 est un paramètre qui ne peut être modifié).

On recalcule les paramètres sur le tri et l'élimination de certains mots-clés ont pu les modifier (PR $\%(3)$ et PR $\%(4)$ ne sont pas modifiés). On a PR $\%(1) =$ PR $\%(8)$, PR $\%(2) =$ PR $\%(7)$, PR $\%(5) =$ PR $\%(12)$, PR $\%(9) = 0.$

Deuxième partie du programme ORGADICO

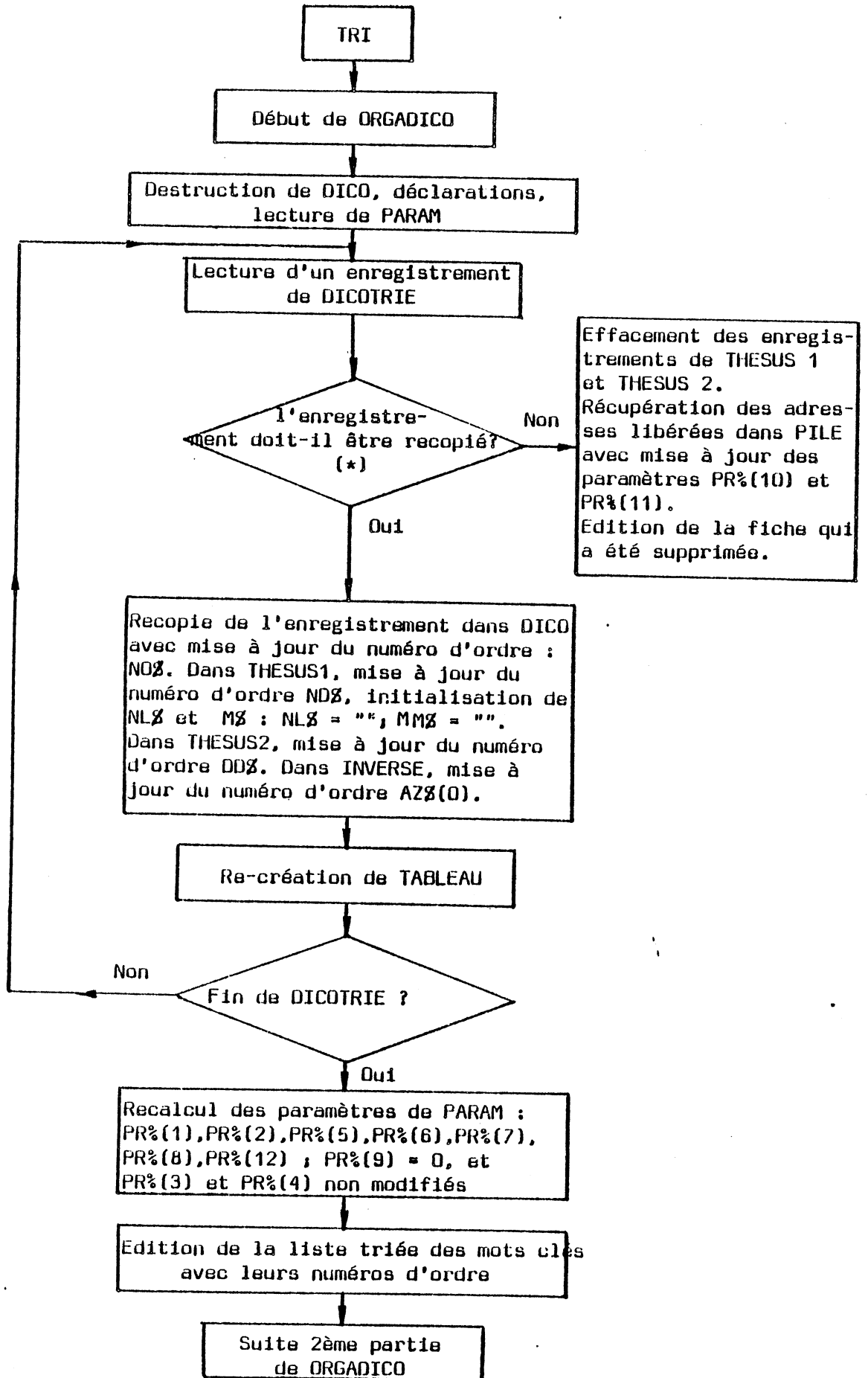
On détecte les anomalies du type 1, 2, 3, et on les stocke dans la première partie du fichier ANOMALIE qui est recréé à cette occasion.

On met à jour :

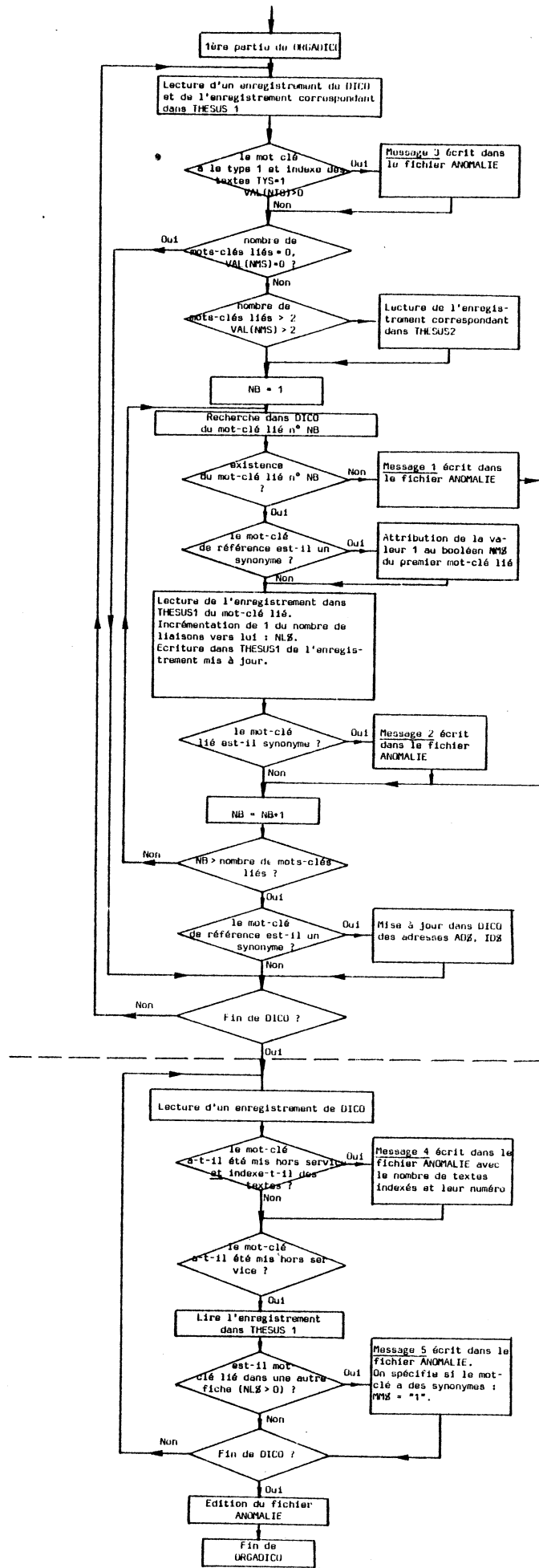
- . la variable NL% qui indique le nombre de liaisons vers le mot-clé de référence,
- . la variable MM%
- . les variables AD%, ID%.

On détecte les anomalies de type 4 et 5. La liste des mots-clés qui ont été mis hors service, mais qui ne peuvent être supprimés (*) constitue la deuxième partie du fichier ANOMALIE recréé à cette occasion.

(*) Ce sont les mots-clés mis hors service mais qui indexent encore des textes ou qui sont présents comme mots-clés liés dans les fiches thesaurus.



(*) VAL(IU%) ≠ 1 ou
 VAL(NL%) ≠ 0 ou
 VAL(NT%) ≠ 0



Organigramme de la 2ème partie de ORGADICO

ANNEXE 13 - EXEMPLES D'ANOMALIES DÉTECTÉES DANS LE PROGRAMME GÉNÉRAL DE CONTRÔLE

Exemples d'anomalies du type 1, 2, 3

Supposons que les mots-clés C et E n'aient pas été créés et que l'on ait les fiches suivantes, le mot-clé B indexe par ailleurs plusieurs textes :

A n° 11	D n° 203	B n° 54 type : 1	F n° 9	G n° 15
2 C 3 D 4 E	1 F	4 G		

Nous avons les messages suivants :

- . Mot-clé de référence A, n° fiche 11
 - mot-clé lié C
 - (1) mot-clé lié non encore saisi
 - mot-clé lié D, n° fiche 203
 - (2) incohérence, mot synonyme de F
 - mot-clé lié E
 - (1) mot-clé lié non encore saisi
- . Mot-clé de référence B, n° fiche 54
 - (3) incohérence entre type et liaison texte

N.B. Les numéros ci-dessus entre parenthèses indiquent le type d'erreur.

Exemples d'anomalies du type 4 ou 5

Supposons maintenant que les mots-clés F, G et B aient été mis hors service ($VAL(IU\mathcal{Z}) = 1$). Par ailleurs, F est présent comme mot-clé lié dans deux fiches thesaurus (dont D) et G est présent comme mot-clé lié dans quatre fiches thesaurus (dont B) (voir schéma des fiches thesaurus ci-dessus). En outre, B indexe trois textes (les textes n° 13, 30, 60).

On a les messages suivants :

édition des mots-clés dont la suppression est demandée mais ne peut encore être faite (ces mots-clés sont hors service, mais restent dans le thesaurus

. Mot-clé de référence F, n° fiche 9

(5) ce mot-clé est présent dans deux fiches thesaurus, il a un (des) synonyme(s)

. Mot-clé de référence G, n° de fiche 15

(5) ce mot-clé est présent dans quatre fiches thesaurus.

. Mot-clé de référence B, n° de fiche 54

(4) on ne peut supprimer ce mot-clé qui indexe encore les textes suivants 13, 30, 60.

Le nombre de fiches thesaurus où figuré F comme mot-clé lié est $NL\mathcal{Z}(F)$.

On sait que F a des synonymes, car $MM\mathcal{Z}(F) = "1"$.

Le nombre de textes indexés par B est donné par $NT\mathcal{Z}(B)$ et les numéros des textes sont les adresses inscrites dans le fichier INVERSE dans l'enregistrement correspondant au mot-clé B.

ANNEXE 14 - MODIFICATION DES FICHIERS DANS GESTDICO

1. Création d'un enregistrement dans THESUS 1

La création d'un nouvel enregistrement dans THESUS 1 n'a lieu que dans la partie création.

PR%(10) est le paramètre indiquant le nombre d'adresses libres dans THESUS 1.

PR%(4) est le paramètre qui donne l'adresse relative du dernier enregistrement de THESUS 1.

Quand un nouvel enregistrement doit être créé dans THESUS 1 :

- . soit $PR\%(10) = 0$, l'enregistrement est mis à la fin du fichier et $PR\%(4)$ est incrémenté,
- . soit on met l'enregistrement à la première adresse libre de THESUS 1, adresse qui se trouve dans le fichier PILE, puis $PR\%(10)$ est décrémenté. $PR\%(10)$ n'est jamais incrémenté dans ce programme.

2. Création d'un enregistrement dans THESUS 2

Le principe est le même que pour la création d'un nouvel enregistrement dans THESUS 1. On récupère d'abord les adresses libres qui se trouvent dans le fichier PILE ; s'il n'y en a pas, on met le nouvel enregistrement à la fin du fichier.

Dans le premier cas, $PR\%(11)$ est décrémenté.

Dans le deuxième cas, $PR\%(3)$ est incrémenté.

$PR\%(11)$ est le paramètre donnant le nombre d'adresses libres dans THESUS 2.

$PR\%(3)$ est le paramètre donnant l'adresse relative du dernier enregistrement de THESUS 2.

La création d'un nouvel enregistrement dans THESUS 2 peut avoir lieu dans les parties création et modification.

3. Suppression d'un enregistrement de THESUS 2

Cela peut se produire lors de la modification d'une fiche thesaurus ou lorsqu'un mot-clé mis hors service est recréé, si le nombre de mots-clés liés était > 2 et devient ≤ 2 .

Dans ce cas :

- . on stocke l'adresse qui sera libérée dans le fichier PILE,
 - . on incrémente PR%(11),
 - . l'enregistrement de THESUS 2 est effacé, ainsi que l'adresse de chaînage dans THESUS 2.
- (PR%(3) n'est jamais décrémenté).

4. Modification du fichier PARAM

Les paramètres PR%(0), PR%(6), PR%(7), PR%(8) ne sont pas modifiés dans GESTDICO.

Pour PR%(4) et PR%(10), se reporter à la création d'un enregistrement dans THESUS 1.

Pour PR%(3) et PR%(11), se reporter à la création et à la suppression d'un enregistrement dans THESUS 2.

Les paramètres PR%(1), PR%(2), PR%(5) et PR%(9), PR%(12) ne sont modifiés que lorsqu'un nouvel enregistrement est créé dans DICO (inexistence du mot-clé de référence) et uniquement dans ce cas.

ANNEXE 15 - FICHIER DESCRIPTEUR DE LA PAGE ÉCRAN : DESDICO

DESDICO est le fichier descripteur des différentes zones des pages écran utilisées dans le programme GESTDICO.

Il y a une seule page écran représentant le masque d'une fiche thesaurus vierge.

Le numéro d'ordre est affiché par programme et non par le gestionnaire d'écran. C'est pourquoi il ne figure pas comme rubrique dans DESDICO.

Ce fichier est un tableau virtuel DS δ (K), K = 0 à 29.

Chaque élément de ce tableau est une variable alphanumérique contenant les caractéristiques d'une rubrique (libellé à afficher, tabulation, emplacement, nombre maximum de caractères à saisir, type et information à saisir et informations facultatives).

En création :

on rentre en mode création (ZM% = 0) ; si c'est une re-crédation, tout se passe comme si c'était une première saisie.

En modification :

le mot-clé est saisi en rubrique isolée pour vérifier qu'il existe et qu'il n'a pas été mis hors service. On rentre en mode consultation-modification (ZM% = 1).

En suppression :

le mot-clé est saisi en rubrique isolée ; on rentre en mode consultation (ZM% = 2).

DESDICO-DAT

- 1 -
- 2 - 0101900022020#1MOT-CLE DE REFERENCE :#
- 3 - 0102138032020#20#
- 4 - 0104201052036#1TYPE DE MOT-CLE:#70/1#
- 5 - 0106900062000#1NOTE D'APPLICATION:#
- 6 - 0107138072020#
- 7 - 0108138082020#
- 8 - 0109138092020#
- 9 - 0110138102020#
- 10 - 0111900122000#1MOT-CLES DIRECTEMENT LIES:#
- 11 - 0112101132020#71/2/3/4#
- 12 - 0113138132022#
- 13 - 0114101142020#71/2/3/4#
- 14 - 0115138142022#
- 15 - 0116101152020#71/2/3/4#
- 16 - 0117138152022#
- 17 - 0118101162020#71/2/3/4#
- 18 - 0119138162022#
- 19 - 0120101172020#71/2/3/4#
- 20 - 0121138172022#
- 21 - 0122101182020#71/2/3/4#
- 22 - 0123138182022#
- 23 - 0124101192020#71/2/3/4#
- 24 - 0125138192022#
- 25 - 0126101202020#71/2/3/4#
- 26 - 0127138202022#
- 27 - 9928101232053#1EDITION DE CETTE FICHE?(O,N)#20#70/N#
- 28 - 9931900022020#1MOT-CLE DE REFERENCE :#
- 29 - 9932138032020#20#

ANNEXE 16 - CRÉATION DE GESTDICO

Il y a deux types de création : soit on crée un mot-clé dont la fiche n'a jamais été saisie, soit on re-crée un mot-clé préalablement mis hors service.

1. Le mot-clé n'existe pas dans DICO

a) Il n'y a pas d'enregistrement correspondant à ce mot-clé : on en crée un dans DICO à la fin du fichier avec mise à jour des paramètres PR%(1) et PR%(2).

On crée aussi un enregistrement dans THESUS 1 (voir remarques générales Annexe 14), et au besoin (si le nombre de mots-clés liés est supérieur à 2), on crée un nouvel enregistrement dans THESUS 2 (annexe 14).

b) Il y a trois sortes de variables écrites dans les fichiers :

- . celles qui sont *saisies* par transaction écran-clavier :
 - mot-clé, type, note d'application, mots-clés liés et leur code liaison ;
- . celles qui sont *initialisées* par un blanc (" ") :
 - l'adresse dans le fichier INVERSE (AIØ = " "),
 - le nombre de textes indexés (NTØ = " ")
 - le booléen mis hors service du mot-clé (IUØ = " "),
 - le nombre de liaisons vers le mot-clé (NLØ = " ").
- . celles qui sont *gérées* par programme :
 - le nombre de mots-clés liés (NMØ) qui est calculé par programme,
 - le numéro d'ordre dont la valeur est attribuée par programme, c'est pourquoi il est affiché,
 - l'adresse relative dans DICO du mot-clé principal : ADØ, initialisée par " " si le mot-clé n'est pas un synonyme, et par "ZZ" si le mot-clé est un synonyme et si le mot-clé principal n'existe pas, sinon on met à jour ADØ en lui affectant l'adresse du mot-clé principal.
 - INDICE du mot-clé principal dans son secteur (IDØ)
 - si le mot-clé de référence est un synonyme et si le mot-clé principal a été trouvé, on met à jour IDØ ; dans tous les autres cas, IDØ = " ".
 - Les adresse de chaînage dans les fichiers THESUS 1 (ATØ) et THESUS 2 (CHØ) sont attribuées automatiquement.

c) Les paramètres :

- . nombre d'enregistrements dans DICO : PR%(5),
 - . nombre de créations d'enregistrements depuis le dernier tri : PR%(9),
 - . plus grand numéro d'ordre : PR%(12),
- sont incrémentés.

2. Re-crédation d'un mot-clé qui a été mis hors service

On sait que les enregistrements d'un mot-clé mis hors service n'ont pas été effacés ; quand l'utilisateur veut recréer un mot clé mis hors service, on garde les anciens enregistrements (dans DICO, THESUS 1, THESUS 2).

a) En modifiant les champs qui ont pu être modifiés au cours de la nouvelle saisie :

- . il s'agit des variables saisies :
 - type, note d'application, mots-clés liés et leur code liaison ;
- . ainsi que :
 - le nombre de mots-clés liés (NM%) qui a pu changer,
 - l'adresse dans THESUS 2 (CH%) car le mot-clé peut avoir un nombre de mots-clés liés qui soit passé de ≤ 2 à > 2 , ou inversement ;
 - l'adresse du mot-clé principal (AD%) et de son indice d'enregistrement (ID%).

b) Par contre, on conserve le numéro d'ordre, le mot-clé de référence, l'adresse de chaînage dans THESUS 1, l'adresse dans le fichier INVERSE (AI%), le nombre de textes indexés (NT%), le nombre de liaisons vers le mot-clé (NL%), le booléen indiquant si c'est un mot-clé principal ou non.

c) Le booléen "mis hors service" IU% est remis à blanc (IU% = " ") puisqu'on recrée le mot-clé.

d) Les paramètres PR%(5), PR%(9) et PR%(12) ne sont pas incrémentés puisqu'il n'y a pas création d'un nouvel enregistrement dans DICO. PR%(1) et PR%(2) ne sont pas modifiés.

e) L'écriture dans les fichiers est commune avec celle de la partie MODIFICATION, car tout se passe comme si on modifiait la fiche du mot-clé (sauf pour la variable IU% qui est réinitialisée dans le cas d'une re-crédation).



ANNEXE 17 - SUPPRESSION DE GESTDICO

Avant de procéder à la mise hors service d'un mot-clé, on vérifie qu'il existe dans DICO et qu'il n'a pas déjà été mis hors service.

Cette vérification a rendu obligatoire la saisie du mot-clé en rubrique isolée, car le gestionnaire d'écran ne permet pas un contrôle zone par zone, c'est pour cette raison que le masque vierge ne s'affiche pas en entier.

Après avoir saisi le mot-clé en rubrique isolée, et vérifié que le mot-clé existe, on visualise le reste de la fiche thesaurus.

On est en mode consultation (ZM% = 2) et la première rubrique affichée est le type correspondant à la rubrique numéro 4 (D% = 4).

Le fichier PARAM n'est pas modifié dans cette partie, puisqu'aucune suppression n'est effective.



ANNEXE 18 - MODIFICATION DE GESTDICO

Ecriture dans les fichiers après contrôles et validation :

- a) En dehors des variables qui correspondent à des zones saisies de la fiche thesaurus : type (TY%), note d'application (N1%, N2%, N3%, N4%), mots-clés liés et codes liaison (MI%, CI%, I = 1, 8).
- b) Un certain nombre de variables dans les fichiers sont susceptibles d'être modifiées, il s'agit :
- . de l'adresse du mot-clé principal (AD%) et de son indice d'enregistrement dans le secteur (ID%) ; en effet, un mot-clé qui était un synonyme peut ne plus l'être ou inversement, ou bien le mot-clé principal peut voir changé.
 - . Le nombre de mots-clés liés (NM%).
 - . L'adresse de chaînage dans THESUS 2 (CHS) : si le nombre de mots-clés liés était > 2 et est devenu ≤ 2 après modification (ou inversement) (annexe 14).
- c) Les autres variables ne sont pas modifiées :
- les numéros d'ordre : NO%, ND%, DO%
 - le mot-clé de référence : MC%
 - l'adresse dans THESUS 1 : AT%
 - le nombre de textes : NT%
 - l'adresse dans le fichier INVERSE : AI%
 - le nombre de liaisons vers le mot-clé de référence : NL%
 - le booléen indiquant si c'est un mot-clé principal ou non (MM%)
 - le booléen "mise hors service du mot-clé" : IU%.



ANNEXE 19 - CRÉATION DU FONDS DOCUMENTAIRE (PROGRAMME INTEX)

Le numéro de fiche est attribué automatiquement avec la même valeur pour les deux bordereaux ; si l'utilisateur modifie cette valeur, il y a émission d'un message d'erreur et l'utilisateur doit corriger sa faute (ceci est à revoir ; il faudrait empêcher l'utilisateur de pouvoir positionner le curseur sur cette zone).

Les descripteurs après validation des deux bordereaux, sont tassés et comptés.

L'insertion dans les fichiers se fait toujours en appliquant le même principe : récupération des adresses vacantes stockées dans PILTEX, si elles existent, sinon le nouvel enregistrement est écrit à la fin du fichier.

L'insertion dans les fichiers se fait dans l'ordre suivant :

- 1) MODES : aucun enregistrement si nombre de descripteurs ≤ 3
 - 1 enregistrement si nombre de descripteurs > 3 et ≤ 9
 - 2 enregistrements si nombre de descripteurs > 9

le deuxième enregistrement est chaîné au premier par C9%.

Le numéro de fiche (F2%) est le même que dans le fichier TEXTE.
- 2) ANALYSE : il y a toujours deux enregistrements par fiche documentaire.

Le deuxième est chaîné au premier par l'adresse CE%.

Le numéro de fiche (F1%) est le même que dans le fichier TEXTE.
- 3) TEXTE : un enregistrement par fiche documentaire.

Chainage avec fichier MODES par l'adresse CD%

chainage avec fichier ANALYSE par l'adresse CA%

Le booléen indiquant si la fiche doit être supprimée "ST%", est initialisé par un blanc.

- 4) INVERSE : le dernier enregistrement dans le fichier INVERSE correspondant à chaque descripteur accepté sera mis à jour immédiatement en rajoutant l'adresse (dans TEXTE) de la nouvelle fiche documentaire créée et qu'il indexe. (Quand il y a création d'un nouvel enregistrement dans INVERSE, la zone AZ%(0) qui contient le numéro d'ordre du descripteur est remplie)

Dans DICO,

- pour chaque descripteur, la variable NT%(nombre de textes indexés) est incrémentée ;
- quand il y a création du premier enregistrement dans INVERSE (c'est-à-dire quand un mot-clé devient pour la première fois un descripteur), la zone AI%(adresse dans le fichier INVERSE) est mise à jour.

Les fichiers PILTEX et PARAM sont également modifiés dans ce programme.

La frappe de MENU fait revenir au menu général de GESTEX1.

Le fichier descripteur associé aux deux pages écran correspondant à la fiche documentaire, est le fichier ECRAN : page n° 1 pour le bordereau n° 1
page n° 2 pour le bordereau n° 2.

Critiques sur ce programme :

- . L'utilisateur ne devrait pas pouvoir modifier le numéro de fiche attribué par programme ;
- . retour automatique à la première page écran quand l'utilisateur veut corriger le deuxième bordereau (il devrait pouvoir rester sur ce deuxième bordereau) ;
- . empêcher les descripteurs de figurer en double (cohérence avec GESTEX2) ;
- . utiliser les places disponibles dans le fichier TEXTE, au lieu de mettre systématiquement le nouvel enregistrement à la fin du fichier. En effet, quand les suppressions effectives auront lieu, il faudra reprendre ce programme.

ANNEXE 20 - MODIFICATION D'UNE FICHE DOCUMENTAIRE

GESTEX 1 (modification du premier bordereau)

Même remarque à propos du numéro de fiche : si l'utilisateur modifie l'ancien numéro, il y a message d'erreur et obligation pour lui de retaper l'ancienne valeur ; il serait préférable d'empêcher l'utilisateur de positionner le curseur sur cette zone.

Fichiers modifiés dans GESTEX 1 :

- . TEXTE : variables susceptibles de modification dans ce fichier : SD%, RT%, RF%, L1%, L2%, DC%
variables non modifiées : NF%, N5%, ST%, CA%, D1%, D2%, D3%, CD%
- . ANALYSE : variables susceptibles de modification dans ce fichier : L3%, L4%, L5%, L6%, L7%, L8%, dans les deux enregistrements associés au texte ;
variables non modifiées : F1%, CE%, dans les deux enregistrements associés au texte.

Ces fichiers sont mis à jour immédiatement.

La frappe de MENU (ou MEN) par l'utilisateur fait revenir au menu général de GESTEX1.

Fichier descripteur de la page écran associé au premier bordereau est le fichier ECRAN première page.

Il n'y a pas de suppression ou d'ajout d'enregistrement dans GESTEX1, puisqu'il n'y a pas de création de fiche documentaire d'une part, et d'autre part à chaque fiche documentaire sont associés deux enregistrements dans le fichier ANALYSE (même si la partie analyse est vide).

Critiques :

- . L'utilisateur ne devrait pas pouvoir modifier le numéro de fiche ;
- . après validation des corrections apportées, l'utilisateur devrait pouvoir corriger une autre fiche, sans être obligé de repasser par le menu général.

GESTEX 2 (Modification du deuxième bordereau)

Pour le numéro de fiche, voir INTEX et GESTEX1.

Fichier descripteur associé au deuxième bordereau : ECRAN (2ème page).

Fichiers susceptibles d'être modifiés dans GESTEX 2 :

- . TEXTE : variables modifiables : N5%, D1%, D2%, D3%, CD%, suppression/ajout de descripteurs ;
variables non modifiées dans ce programme : NF%, SD%, RF%, RT%, ST%, L1%, L2%, CA%, DC%
- . MODES : variables modifiables : D4%, D5%, D6%, D7%, D8%, D9%, C9%, dans les deux enregistrements,
variable non modifiée : F2%, dans les deux enregistrements.
(suppression/ajout de descripteurs).
Au cours de ce programme, des enregistrements du fichier MODES sont susceptibles d'être supprimés ou rajoutés (toujours en utilisant le principe de récupération des adresses vacantes stockées dans PILTEX).
- . INVERSE : ajout, suppression, tassement, et tri des adresses de textes stockées. Au cours de ce programme, des enregistrements de INVERSE sont susceptibles d'être supprimés ou rajoutés (toujours en utilisant les adresses vacantes stockées dans PILTEX).
Remarques : les tri et le tassement des adresses pourraient se faire au cours du programme de réorganisation et de contrôle (ORGADICO).
- . DICO : deux variables peuvent être modifiées dans l'enregistrement d'un mot-clé descripteur ou anciennement descripteur.
NT% : nombre de textes indexés, cette variable peut être incrémentée ou décrétementée
AI% : adresse dans le fichier INVERSE. Seule modification : remis à blanc si le mot-clé n'indexe plus de texte.

Bien sûr les fichiers PILTEX et PARAM sont mis à jour dans ce programme.

La commande MENU fait revenir au menu général de GESTEX 1.

Critiques sur GESTEX 2 :

- . l'utilisateur ne devrait pas pouvoir modifier le numéro de fiche.
- . Permettre que GESTEX 1 et GESTEX 2 soient en mémoire ensemble, ce qui éviterait à l'utilisateur d'attendre le chargement de GESTEX 1 quand il a fini de corriger le bordereau 2 d'une fiche et qu'il veut corriger une autre fiche.

Proposition :

Pour diminuer la taille de GESTEX 2, le tri et le tassement des adresses dans INVERSE devraient se faire à la sortie de GESTEX 1 avant de revenir à la page d'accueil, ou dans le programme de réorganisation.



ANNEXE 21 - CRITIQUES TECHNIQUES DU LOGICIEL

Ecriture des nombres et des adresses dans les fichiers

Les nombres ont été écrits dans les fichiers en utilisant la fonction STR\$(X) qui transforme un nombre en chaîne de caractères. En utilisant d'autres fonctions de conversion (CHR\$ ou CVT\$), certaines variables pourraient être écrites sur moins d'octets

Exemple :

nombre de textes indexés NT\$: sur 2 octets au lieu de 4,
 nombre de mots-clés liés NM\$: sur 1 octet au lieu de 2,
 nombre de liaisons vers le mot-clé NL\$: sur 2 octets au lieu de 4,
 adresse de chaînage dans THESUS 1 AT\$: sur 2 octets au lieu de 5,
 adresse de chaînage dans INVERSE AI\$: sur 2 octets au lieu de 5,
 adresse du mot-clé principal AD\$: sur 2 octets au lieu de 4,
 adresse de chaînage dans THESUS 2 CH\$: sur 2 octets au lieu de 5,
 adresses des textes dans le fichier INVERSE : sur 2 octets au lieu de 4.

Certains programmeurs utilisant le GOUPIIL nous avaient mis en garde contre l'utilisation des fonctions CVT\$ et nous avaient recommandé l'utilisation de la fonction de conversion STR\$. Par ailleurs, les numéros d'ordre des mots-clés et les numéros des fiches étaient affichés en utilisant le gestionnaire d'écran, or celui-ci utilise la fonction de conversion STR\$. Pour uniformiser le mode d'écriture des nombres dans les fichiers, nous avons donc décidé d'utiliser la fonction STR\$.

En fait, l'utilisation des fonctions CVT\$ et CHR\$ s'avère sans risque et de plus les numéros des fiches étant attribués par programme, pourquoi ne pas les faire afficher sans utiliser le gestionnaire d'écran. Solution qui permettrait d'empêcher toute modification de ces numéros par l'utilisateur et ferait gagner de la place.

Nouvelle organisation proposée :

DICO

Nombre d'octets :	2	38	2	2	1	2	1	2	= 50 octets
Variable :	NOØ	MCØ	ATØ	AIØ	IUØ	ADØ	IDØ	NTØ	
	↓	↓	↓	↓	↓	↓	↓	↓	
	↑	↑	↑	↑	↑	↑	↑	↑	
	Numéro d'ordre	Mot-clé de référence	Adresse dans THESUS 1	Adresse dans INVERSE	Booléen mis hors service ou non	Adresse du mot-clé principal	Indice d'enregistrement du mot-clé principal dans son secteur	Nombre de textes indexés	

Par rapport à l'ancienne version, seul le type n'est pas inscrit dans DICO, mais pour savoir si un mot-clé est descripteur ou non, la variable NTØ suffit et est plus exacte (car gérée par programme à l'inverse de TYØ).

THESUS 1 :

L'enregistrement de ce fichier contiendrait les mêmes variables avec le type en supplément, en stockant les adresses et nombres sur 1 ou 2 octets, comme indiqué ci-dessus.

THESUS 2 :

Inchangé, sauf numéro d'ordre sur 2 octets au lieu de 4.

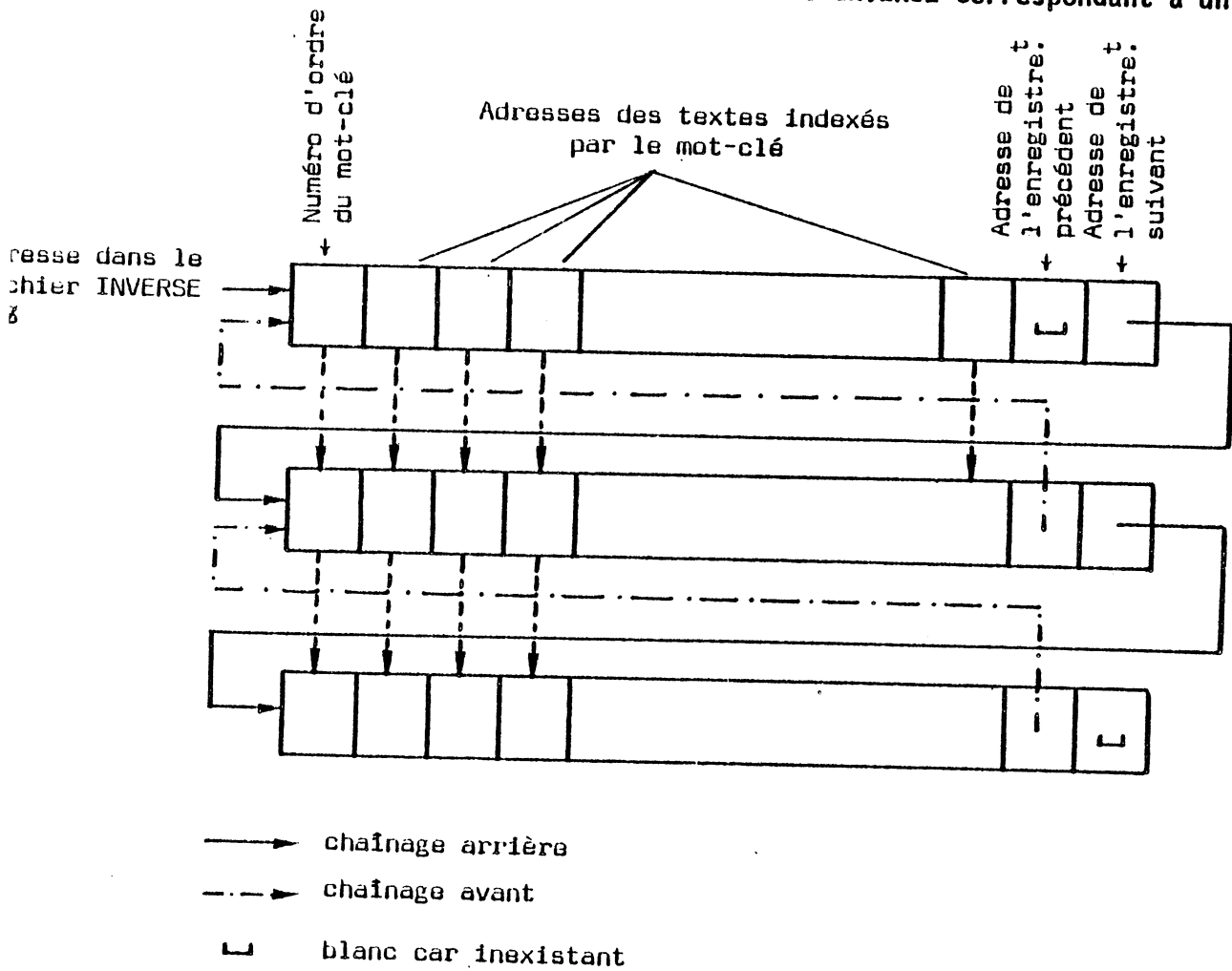
INVERSE :

Trois modifications à apporter dans l'enregistrement de ce fichier :

- 1) adresses textes stockées sur 2 octets au lieu de 4,
- 2) numéro d'ordre sur 2 octets,

- 3) chaînage avant pour faciliter le tassement des adresses des textes lors d'une suppression d'une des adresses.

Nouveau dessin des enregistrements dans INVERSE correspondant à un mot-clé



Nombre d'adresses de textes stockées par enregistrement :
 $(252 - 3 \times 2) / 2 = 123$ au lieu de 31 à l'heure actuelle.

Cette écriture des entiers sur 1 ou 2 octets au lieu de 2, 4 ou 5, ne ferait pas gagner de place au niveau des fichiers du fonds documentaire.

Fonctions de conversion

CHR\$(X) : cette fonction fournit un simple caractère ASCII (une chaîne de 1 caractère) qui est le code ASCII de l'argument X ; l'argument doit être un nombre compris entre 0 et 255.

La fonction inverse est **ASC(X\$)** qui donne la valeur numérique ASCII du premier caractère de la chaîne.

STR\$(X) : fonction qui retourne une chaîne de caractères qui représente l'expression numérique de X. La chaîne est construite en imitant tous les DIGITS contenant les espaces.

Fonction inverse de **STR\$(X)** : **VAL(X\$)**.

CVT\$(X) : ces fonctions ne réalisent pas à proprement parler des conversions puisque les données numériques conservent leur représentation binaire dans l'enregistrement. Il s'agit plutôt d'un artifice permettant d'affecter une valeur numérique à une variable de type chaîne et inversement.

Si X est un entier, **CVT\$(X)** occupera 2 octets.

AUTORISATION DE SOUTENANCE

U les dispositions de l'article 3 de l'arrêté du 16 avril 1974,

U le rapport de présentation de Monsieur L. BOLLIET, Professeur

Madame FILLON Agnès épouse JAUNEAU

est autorisée à présenter une thèse en soutenance pour l'obtention du titre de
DOCTEUR de TROISIEME CYCLE, spécialité "Informatique".

Fait à Grenoble, le 16 juin 1983

Le Président de l'I.N.P.-G. 7

D. BLOCH
Président
de l'Institut National Polytechnique
de Grenoble

P.O. le 16/06/1983



