

MÉTHODE DE CLASSEMENT ET DE RECHERCHE BIBLIOGRAPHIQUE AUTOMATISÉS DANS UN LABORATOIRE DE PHYSIOLOGIE

Un laboratoire de recherche médicale s'intéresse à un domaine si particulier, que les méthodes de classification bibliographique faisant l'objet de catalogues sont trop générales pour être utiles. D'autre part, les recherches sur ordinateur par des organismes spécialisés ne peuvent pas être de pratique quotidienne car ils sont surtout appropriés à rechercher des références inconnues sur un sujet nouveau. De plus, cette recherche est lente et onéreuse et le pourcentage des références pertinentes est très variable. Elle dépend de la qualité de la communication entre le chercheur spécialisé dans un domaine qu'il manie quotidiennement avec un certain langage et le bibliographe pour qui chaque sujet de recherche bibliographique est d'abord un monde à comprendre. Plus un laboratoire vieillit, plus le nombre d'articles et de références accumulés est important. Au bout de quelques années, une classification par codage des articles est nécessaire par un ensemble de mots-clés en rapport avec l'activité spécialisée du laboratoire.

L'utilisation d'une calculatrice permet d'obtenir immédiatement, sur place, l'ensemble des articles ayant en commun un ou plusieurs mots-clés. L'objet du travail que nous présentons ici était de mettre au point une méthode de stockage de données bibliographiques et d'utilisation rapide de ces données.

MATÉRIEL UTILISÉ ET MÉTHODES D'EXPLOITATION

Nous disposons d'une calculatrice Hewlett-Packard 9820 A munie de 429 pas de programme, d'un bloc de fonctions mathématiques (11221 A) et d'un lecteur de cassettes à bandes magnétiques (9865 A) sur lesquelles l'information sera stockée et récupérée à la demande par la calculatrice. Les bandes magnétiques permettent de stocker un maximum d'informations : plus de 6 000 registres par cassette au lieu de 429 registres pour la calculatrice. Elles évitent ainsi la présence continue d'un opérateur, celui-ci se contentant de lancer le programme et de recueillir les informations désirées en comptant environ 10 minutes de recherche pour 1 000 articles. Le programme de recherche peut cependant être stocké séparément sur carte magnétique que l'on fera lire à la calculatrice avant la recherche bibliographique.

La calculatrice ne travaillant pas en alphabétique, mais en numérique, nous avons été contraints de créer un code établissant une correspondance entre les « mots-clés » caractérisant un article et des équivalents

numériques pour la calculatrice. Les mots-clés sont des termes français employés en médecine et plus spécialement en physiologie, en mécanique avec un intérêt centré sur la biomécanique du tube digestif et de ses sphincters.

Les mots-clés sont classés en 11 colonnes de 98 mots codés de 1 à 98 et figurant sur un catalogue que nous utilisons au moment du codage des articles. La colonne « 0 » contient la liste des « appareils » (circulatoire, respiratoire, urinaire, génital...). La colonne « 1 » est celle des organes concernés (côlon, foie, poumons...). La colonne « 2 » est celle de la partie d'organes (cholédoque, pylore, coecum, etc.). La colonne « 3 » contient des mots qui déterminent l'organe ou la recherche effectuée (antérieur, bilatéral, inférieur, inversé, lent, lisse...). La colonne « 4 » est celle des disciplines (enzymologie, génétique, mécanique, informatique...). La colonne « 5 » est celle des techniques utilisées ou de l'objet précis de l'étude (ablation, adaptation, biopsie, combustion, compliance...). La colonne « 6 » précise la colonne précédente par un adjectif (expérimental, enregistré, bénin, naturel, malin...). La colonne « 7 » est celle des anomalies et plus particulièrement des maladies (cancer, calcul, colite...). La colonne « 8 » est celle du matériel sur lequel s'est effectué le travail relaté par l'article (canard, chien, cobaye) et nous avons différencié dans cette colonne l'homme, la femme, l'enfant et le nouveau-né. La colonne « 9 » est celle du milieu chimique, biologique ou pharmacologique concerné dans le travail relaté par l'article (alcool, angiotensine, calcitonine...). La colonne « 10 » est celle des instruments utilisés dans le travail (débitmètre, diodes...). La colonne « 11 », enfin, est celle des auteurs dont nous lisons préférentiellement les articles. Cette colonne est très utile, car elle permet de retrouver facilement un article avec peu d'autres mots-clés que celui du nom de l'auteur.

Tout article n'a pas obligatoirement de référence dans chaque colonne-type du code, d'où l'obligation de désigner dans chacune de celles-ci un mot-clé appelé néant. Ce nombre sera « 00 ».

Il est intéressant de pouvoir faire un test sur un nombre restreint de colonnes. Les colonnes que l'on ne désire pas tester feront l'objet d'un « non-test » dans le numéro de code. On tapera alors le numéro de code « 99 ».

Les différentes manipulations que l'on peut effectuer à l'aide de la calculatrice et de son équipement sont :

- Le découpage d'une cassette vierge
- La vérification du contenu d'une cassette

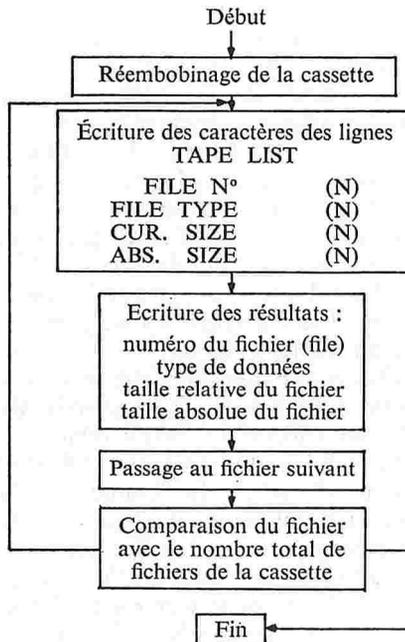
- . La vérification du contenu d'un fichier (ou « file » ou « bloc »)
- . Le chargement d'un article dans un fichier de la cassette
- . La recherche d'un ou plusieurs articles.

Le découpage d'une cassette vierge se fait à l'aide de la calculatrice en 84 fichiers de 100 registres. Chacun de ces registres peut contenir 50 articles.

La vérification du contenu d'une cassette se fait à l'aide d'un programme enregistré sur carte magnétique et injectable dans la calculatrice (organigramme n° 1).

La vérification du contenu d'un fichier se fait à l'aide d'un deuxième programme enregistré sur carte magnétique et injectable dans la calculatrice (organigramme n° 2). Ce programme permet d'obtenir sur l'imprimante :

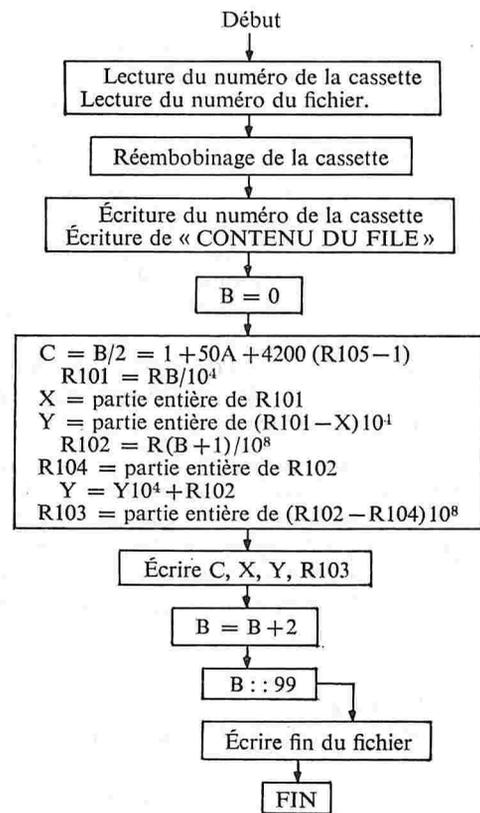
ORGANIGRAMME N° 1

PROGRAMME DE VÉRIFICATION
DU CONTENU D'UNE CASSETTE

A l'aide de l'imprimante incorporée dans la calculatrice, on peut obtenir :

- le numéro du fichier
- le type d'information qui s'y trouve avec comme codage :
 - 0 = fichier vide
 - 2 = donnée numérique
 - 20 = donnée d'un programme
- le nombre de registres effectivement utilisés dans le fichier.
- le nombre maximum de registres utilisables dans le fichier.

ORGANIGRAMME N° 2

PROGRAMME DE VÉRIFICATION
DU CONTENU D'UN FICHIER

C : numéro de l'article
 X : 8 premiers chiffres du numéro de code de l'article
 Y : 8 chiffres suivants du numéro de code de l'article
 R103 : 8 derniers chiffres du numéro de code de l'article

- le numéro du fichier recherché
- le numéro de classement de l'article dans la bibliothèque
- les 12 numéros de 2 chiffres du codage de l'article qui figurent chacun dans une des 12 colonnes du catalogue.

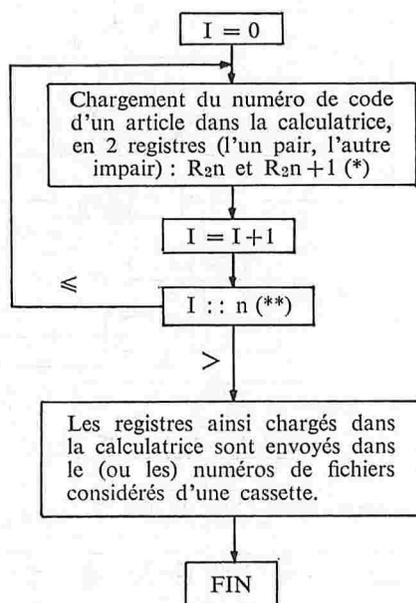
Pour charger un article dans une cassette, il faut préalablement vérifier que le fichier concerné ne contient pas d'information et que tous les registres du fichier précédent sont utilisés car le programme de recherche d'un article prévoit que la recherche s'arrête dès que la calculatrice rencontre deux registres nuls. Si l'on veut charger 50 articles dans un fichier, on affecte chaque article du numéro n ($1 \leq n \leq 50$). Chaque article est codé par deux séries de 12 chiffres que l'on stocke dans deux registres R_{2n-1} et R_{2n-2} . On envoie alors l'information dans les registres R_0 à R_{99} du fichier N de la cassette ($0 \leq N \leq 83$) (organigramme n° 3). Il est possible aussi de modifier ou de compléter un fichier dans lequel il y a des données enregistrées.

Pour rechercher un ou plusieurs articles, il faut d'abord injecter dans la calculatrice un programme enregistré

sur une carte magnétique (organigramme n° 4). Ce programme interroge sur le n° de la cassette utilisée, puis sur le n° de code du mot pertinent contenu dans chacune des 12 colonnes du catalogue. Pendant que la cassette se réembobine, ces informations sont imprimées par l'imprimante de la calculatrice. Celle-ci enregistre le premier fichier de 50 articles. Ensuite, la cassette s'arrête, rien n'apparaît sur l'écran. La calculatrice teste alors le premier fichier. Si elle y trouve les numéros d'article désirés, elle imprime sur le papier de l'imprimante le n° de classement de l'article dans la bibliothèque. Elle enregistre ensuite le second fichier et procède à la même analyse, puis, passe aux fichiers suivants jusqu'à ce qu'elle rencontre deux registres vides.

ORGANIGRAMME N° 3

CHARGEMENT D'ARTICLES DANS UN FICHIER DE LA CASSETTE



Le chargement peut être fait sous forme de programme ou bien en exécution simple.

(*) Ces registres ont les premiers numéros de registres libres dans un fichier. Afin de les connaître si le fichier est incomplet, il est nécessaire d'effectuer un listing de celui-ci à l'aide du programme « CONTENU DU FICHIER ».

(**) n'est le nombre d'articles que l'on peut enregistrer dans le fichier ($n = 49$ si le fichier est vierge, et $n < 49$ suivant la place occupée si le fichier est incomplètement rempli).

COMMENTAIRES ET CONCLUSION

La présence d'une calculatrice de petit volume mais suffisamment performante dans un laboratoire permet de stocker une information dont le volume est certes limité et dépendant des possibilités de lecture des chercheurs et techniciens qui y travaillent, mais elle est directement liée aux préoccupations de ces derniers et rapidement disponible quand ils ont besoin de toutes leurs références répondant à un ou plusieurs mots-clés. Au-delà d'un nombre d'articles supérieur à 2 000, cette nouvelle classification automatisée nous a paru plus facile à exploiter que celle que nous utilisons jusqu'à présent. Dans notre expérience, cette classification automatisée est intervenue alors que nous avions déjà établi la classification « manuelle » de quelque 2 000 articles par organe (foie, estomac, cœur, etc.). Nous avons donc codé chaque article dans l'ordre où il était déjà classé antérieurement. Par conséquent, pour cet ensemble d'articles le codage par mots-clés ne fait que doubler la première classification qui est toujours exploitable manuellement. Mais les articles nouvellement reçus au laboratoire doivent obligatoirement être stockés dans la bibliothèque dans l'ordre de leur réception. Toute recherche manuelle sur un sujet donné devient donc totalement impossible.

Cette méthode présente aussi un certain nombre d'inconvénients en ce qui concerne le codage des articles. En effet, un mot-clé d'une colonne donnée peut-être synonyme d'un mot-clé d'une autre colonne. Pour récupérer les articles codés avec l'un et l'autre mot, il faudra faire plusieurs recherches par la calculatrice. De même, un article peut être à la fois codé par deux mots d'une même colonne si l'on en juge par son titre et son contenu : cela est impossible,

mais il faudra éventuellement en tenir compte dans les recherches de références en testant l'un et l'autre mot par deux recherches successives.

Il faut donc choisir judicieusement les mots-clés de chaque colonne. Il importe donc d'éviter les synonymes sur une même colonne et d'une colonne à l'autre et les mots qui ne codent qu'un très petit nombre d'articles. Mais, le grand avantage de cette méthode de classification est sa possible modification à la demande. Si un mot-clé d'une colonne est peu ou pas employé, il peut être remplacé par un synonyme plus pertinent. On peut aussi le supprimer. Mais dans cette dernière hypothèse, il faut rechercher l'article, modifier sa cotation et en informer la calculatrice qui transmettra la modification à la bande magnétique. Il existe pour cela un programme dont nous avons déjà parlé. Certaines colonnes de 98 mots-clés ont été dans notre expérience rapidement saturées : c'est le cas de la colonne « 9 » qui s'occupe des « milieux chimiques, biologiques ou pharmacologiques ». On a donc intérêt, pour conserver une place à des mots-clés nouveaux à supprimer deux ou trois mots-clés et à les remplacer par un seul qui est le nom de l'ensemble qu'ils forment (exemple : remplacer butanol, ethanol et methanol par « alcools »).

La possibilité de réviser continuellement cette classification et celle d'obtenir en quelques secondes des références pertinentes sont liées entre elles ainsi qu'à la perspicacité et à l'entraînement du responsable de la bibliothèque.

Pierre ARHAN, Joëlle BORIONE,
Francis EUVE, Yves POUSSIN,
Claude FAVERDIN, Catherine DORNIC,
Denys PELLERIN.