

M. JACOBZONE

G. CARBONNEL

**La gestion des fichiers micropaléontologiques par ordinateur
Une application du programme Mistral**

Publications du Département de Mathématiques de Lyon, 1976, tome 13, fascicule 4
, p. 101-132

http://www.numdam.org/item?id=PDML_1976__13_4_101_0

© Université de Lyon, 1976, tous droits réservés.

L'accès aux archives de la série « Publications du Département de mathématiques de Lyon » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LA GESTION DES FICHIERS MICROPALÉONTOLOGIQUES PAR ORDINATEUR
UNE APPLICATION DU PROGRAMME MISTRAL
par M. JACOBZONE et G. CARBONNEL

TABLE DES MATIÈRES

CHAPITRE 1	LA RECHERCHE DOCUMENTAIRE
	1. Définition
	2. Moyens utilisés
	3. Formulation des demandes
CHAPITRE 2	MISTRAL
	1. Le stock documentaire
	2. Le thésaurus
	3. Implantation de Mistral
CHAPITRE 3	EXPLOITATION PRATIQUE DE MISTRAL
	1. Phase préliminaire
	2. Deuxième phase
	3. Troisième phase/Interrogation
CHAPITRE 4	MODE D'INTERROGATION / EXEMPLES
	1. Traitement par lots
	2. Traitement en temps partagé (TS)
CONCLUSIONS	

RESUME

Le programme MISTRAL de la compagnie C. I. I. (programme de recherche documentaire) a été adapté à un fichier micropaléontologique d'ostracodes. Celui-ci comprend environ 12 000 documents de base correspondant chacun à une espèce. Les possibilités d'exploitation : procédure d'interrogation et recherche rétrospective sont analysées.

L'accès au fichier est possible sur demande adressée aux auteurs.

ABSTRACT

The MISTRAL program (program for documentary research) of the C. I. I. company, has been adapted to micropaleontological punchcards for ostracodes, which now contain about 12 000 basic units, each for one species. Processes of recall and retrospective research have been analysed as possibilities for use. Access to the file can be obtained through the authors.

INTRODUCTION

La recherche micropaléontologique, en particulier celle concernant les ostracodes, s'enrichit chaque année, depuis dix ans, en moyenne de 500 à 600 espèces nouvelles. Ces créations s'ajoutent au nombre important d'ostracodes déjà répertoriés avant 1960.

Ainsi à ce jour, on dénombre dans le monde environ 12 000 espèces nouvelles, valides ou non, dont la répartition stratigraphique s'étend du Tertiaire inférieur à l'Actuel.

De ce nombre sont exclues les espèces créées par les auteurs soviétiques.

Devant l'abondance de ces documents, le besoin impérieux de maîtriser ces informations nous oblige à appliquer les procédés de la recherche documentaire sur ordinateur. Le recours à une trieuse mécanique des cartes perforées (cf. Cl. Chevalier, 1974) se révèle en effet insuffisant et d'un emploi malaisé.

Le problème fondamental que doit résoudre le micropaléontologiste peut se formuler ainsi : obtenir une liste d'ostracodes répondant à certains critères stratigraphiques, géographiques et paléoécologiques ou une combinaison des trois ensembles.

Comment le résoudre ? Les procédés appliqués à la recherche documentaire nous permettent d'y parvenir.

CHAPITRE 1

LA RECHERCHE DOCUMENTAIRE

Ce chapitre traitera, d'un point de vue théorique, des conditions nécessaires à la recherche documentaire automatique.

1. Définition

La recherche documentaire est une technique nouvelle qui se propose de fournir les références des documents satisfaisant à une question donnée. Elle permet de sélectionner les documents répondant à un des deux types de demande suivante :

- Diffusion sélective : mise au courant des nouveaux documents.
- Recherche rétrospective : connaissance des références des documents, en l'occurrence des ostracodes, répondant aux critères souhaités.

Ce deuxième type de demande est le seul à être utilisé en micropaléontologie. C'est donc celui que nous allons développer.

2. Moyens utilisés

Le nombre important de documents ne nous permet pas de les traiter par une méthode classique. Aussi l'ordinateur nous a-t-il paru apte à résoudre ces problèmes grâce à sa souplesse d'emploi, sa rapidité de réponse et son aptitude à gérer des volumes importants d'informations.

2.1. Langage de description

Le langage utilisé doit être à la fois accessible à l'ordinateur et surtout au chercheur, afin que ce dernier puisse s'exprimer commodément. Aussi nous avons choisi un système de classification.

2.1.1. Classification par mot-clé

Afin de faciliter les échanges d'analyses de demandes et de réponses, il fallait normaliser le système utilisé. Aussi est-on conduit théoriquement à rechercher un système fondé, non pas sur un code, mais décrivant le sujet du document à partir du contenu même de celui-ci.

On utilise donc des "mots-clés" ou "mots-vedettes" choisis par le chercheur afin de rendre compte le plus fidèlement possible du sujet traité.

Exemple : Dans le cas d'un fichier d'ostracodes le nom de l'ostracode (espèce et genre) sera un mot-clé le décrivant.

Par la suite nous dirons que caractériser un document par des "mots-clés", c'est l'indexer.

2.1.2. Synonymie des mots-clés

Nous constatons lors de l'analyse des documents qu'un chercheur peut utiliser un mot-clé alors qu'un autre chercheur, dans le même domaine, pourra utiliser un autre mot-clé très proche.

Aussi l'existence de synonymie étant inévitable, il fallait créer un "thésaurus" prévoyant ces rapports de synonymie. Chaque chercheur pourra employer les "mots-clés" synonymes figurant au thésaurus. L'exemple suivant illustrera cette notion.

Soit l'ostracode dénommé Loxoconcha créé par un auteur que certains utilisateurs appelleront BRADY et d'autres BRAD, le thésaurus des mots-clés prévoiera que le nom de l'auteur sera indifféremment BRADY = BRAD.

Le thésaurus des mots-clés contient donc l'ensemble des mots-clés concernant un sujet donné.

2.1.3. Description complémentaire

Lorsqu'on "traite" un document on peut le décrire par mot-clé mais on peut établir d'autres renseignements caractérisant le document.

Un ostracode représenté par son genre, son espèce, peut également être caractérisé par l'auteur qui l'a découvert, l'année de sa création, les références bibliographiques où il a été cité, etc. . .

C'est donc en fonction des besoins que l'on établit la liste des renseignements à conserver.

3. Formulation des demandes

3.1. Fonction des mots-clés

Nous constatons que l'énumération des mots-clés ne suffit pas à constituer une question. Nous pouvons utiliser la logique booléenne sur les mots-clés. Ainsi nous souhaiterions avoir la liste des ostracodes appartenant au genre Loxoconcha, à l'espèce X ou Y, et vivant à la profondeur codée 2. La question sera rédigée ainsi :

(LOXOCONCHA) (X OU Y) ET PROF 2.

Une première sélection trouvera le genre LOXOCONCHA. Une deuxième sélection ne gardera que ceux répondant à l'espèce X ou Y dont la profondeur sera 2.

3.2. Autres critères de sélection

Les mots-clés ne sont pas les seuls employés et d'autres modes de sélection pourront être choisis. Tout dépendra du sujet traité.

CHAPITRE 2

MISTRAL

La Compagnie internationale pour l'informatique (C. I. I.) a élaboré un programme¹ de recherche documentaire nommé MISTRAL : Mémorisation d'Information, Sélection, Traitement et Recherche Automatique. Il a été conçu pour venir en aide aux documentalistes.

L'utilisation de Mistral impose de constituer les informations nécessaires pour obtenir une réponse à une question demandée, c'est-à-dire le stock documentaire.

1. Le stock documentaire

1.1. Généralités

Il faut tout d'abord décomposer les informations concernant un document (ici une espèce d'ostracode) en autant d'unités souhaitées. Ces unités d'information s'appellent des champs.

1.1.1. Décomposition en champs

D'une façon générale, ce choix dépend des questions que le chercheur désire poser. En effet, les informations peuvent servir :

- de critères de sélection ou
- de sources d'information pour une édition.

Exemple : Prenons un document relatif à une espèce d'ostracode (ici, Cythereis dentata, MÜLLER, 1894) et montrons les possibilités de créations de champs. Ce document, à l'instar de tous les autres, comprend :

1 genre et espèce

2 auteur et année de création

1 Nous remercions Monsieur M. DUCLOS (C. I. I.) pour l'aide apportée lors de sa mise en oeuvre.

- 3 âge : ère, période, étage
- 4 provenance
- 5 références bibliographiques : périodique (s), volume (s), fascicule (s), pages
- 6 paramètres écologiques et paléoécologiques.

Ces informations étant globales et mélangées, on peut ensuite affiner la décomposition afin de ne plus avoir d'ambiguïté de la façon suivante :

- 1 genre
- 2 espèce
- 3 auteur de création
- 4 année de création
- 5 ère
- 6 période
- 7 étage
- 8 périodique contenant les références du document
- 9 volume dans lequel est cité l'ostracode
- 10 numérotation des pages se rapportant à l'article.

La distinction de plusieurs données nécessite l'attribution d'un champ à chacune. Les champs pourront être disposés dans n'importe quel ordre car chacun est reconnu par son nom.

1.1.2. Contenu du champ

1.1.2.1. Nature du champ

Un champ peut être alphanumérique, tous les caractères de l'alphabet sont admis ainsi que les chiffres. Exemple : G E N R E, A U T 1 3

Il peut aussi être numérique en évitant toutefois les points, les virgules, les blancs.

1.1.2.2. Valeur d'un champ

Un champ ne forme pas un tout, il est décomposé en un certain nombre de valeurs. Soit par exemple le champ AUTEUR. Ce champ peut être constitué de plusieurs noms séparés par un séparateur

choisi par le documentaliste. Ici nous avons choisi la virgule :

Exemple : BRADY, ROBERTSON ... chaque nom est une valeur du champ AUTEUR.

Cette notion de valeur est utilisée lors de la formulation des questions.

1.1.2.3. Mots réservés

Certains mots et sigles ont une signification particulière dans Mistral :

// (2 traits) représentent le séparateur de champ

(virgule) est le séparateur de valeurs de champ

R C H est un opérateur permettant de rechercher une chaîne de caractères.

Cette liste de mots réservés est fournie par le programme MISTRAL.

1.2. Constitution d'un stock documentaire

L'entrée des documents dans l'ordinateur peut s'effectuer à partir de supports variés :

- . cartes perforées
- . bandes magnétiques
- . disques.

Seul sera décrit ici le support par cartes perforées.

En effet, nous avons utilisé le fichier ostracodes, réalisé entièrement sur cartes perforées par Cl. Chevalier, (1974). Il constitue la base du stock documentaire.

Le type du document est représenté par la figure 1 (bordereau de perforation et carte perforée correspondante).

La totalité des espèces d'ostracodes du fichier de cartes perforées du Laboratoire ostracode de Lyon, comprennent ou peuvent comprendre les indications suivantes :

genre, espèce, sous-espèce, auteur, année de création, auteur ayant cité l'espèce, année de citation, en notation alphabétique ou alphanumérique.

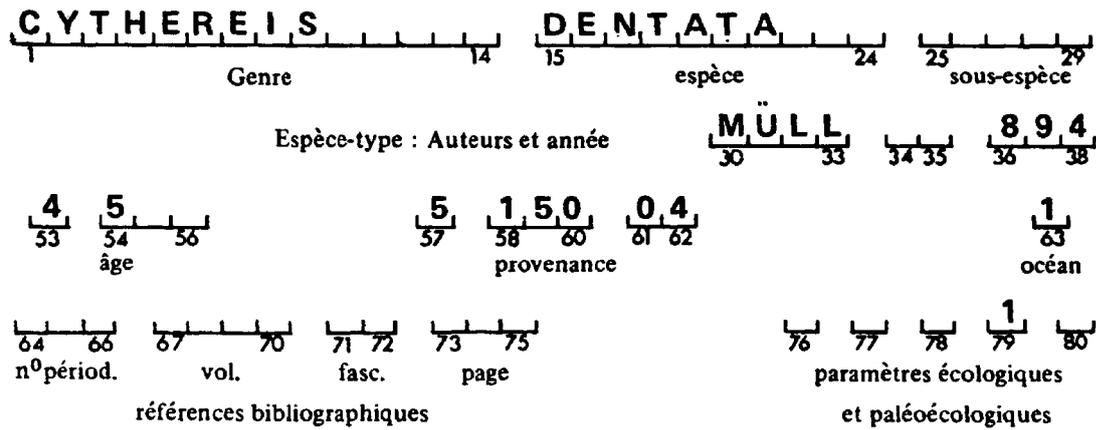
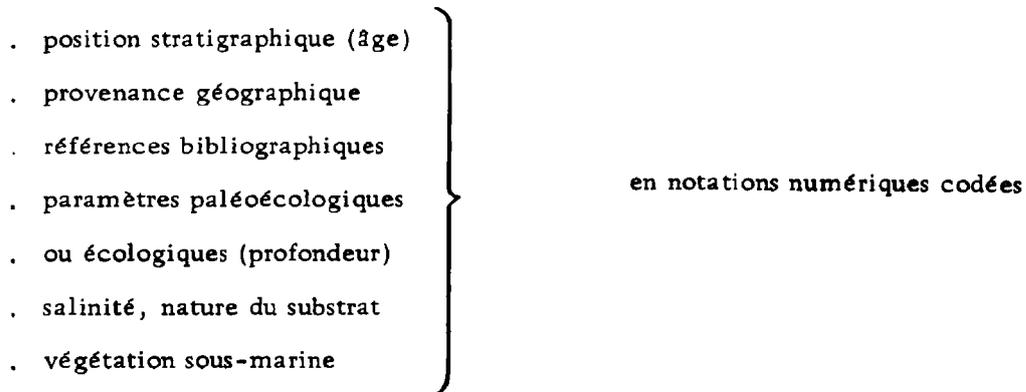


Fig. 1 - Document de base du stock documentaire (bordereau et carte perforée correspondante).



Ce fichier a été remodelé afin de le rendre utilisable pour le programme Mistral suivant le modèle de la figure 2.

```

0,EN,OSIR      2://GEN  CYTHEROMORPHA //ESPE  ACIUNCIAT
1,EN,OSIR      2://AUT1  BRAB
2,EN,OSIR      2://ANCP  880
3,EN,OSIR      2://AUCI  ISHI
4,EN,OSIR      2://ANCI  966
5,EN,OSIR      2://ERE  4

6,EN,OSIR      2://PERI  5
7,EN,OSIR      2://CONT  4
8,EN,OSIR      2://PAYS  100
9,EN,OSIR      2://DEPT  30
10,EN,OSIR     2://OCEN  3
11,EN,OSIR     2://REV   99
12,EN,OSIR     2://RIHP  99 99  9

13,EN,OSIR     2://TEMP  3
14,EN,OSIR     2://SUBST  5
15,EN,OSIR     2://SALM  4
16,EN,OSIR     2://PROF  2
    
```

Fig. 2 - Schéma de la restructuration des données.

Supposons en effet que nous voulions introduire une espèce d'ostracode :

genre : CYTEROMORPHA

espèce : ACUPUNCTA

auteur : M Ü L L (E R) ¹

année : (1) ¹ 869

Nous utiliserons les cartes suivantes que nous commenterons ensuite :

O, B N, O S T R 1 : // GEN ^CYTHEROMORPHA // ESPE ^ACUPUNCTA

1, B N, O S T R 1 : // AUT ^MÜL

2, B N, O S T R 1 : // ANN ^869

La première carte correspondant à la première ligne se décompose ainsi :

O, numéro de séquence

BN, sigle

OSTR 1 : référence du document (ici document ostracode entré sous le n° 1)

// : séparateur

GEN : indicatif de champ

CYTEROMORPHA : contenu du champ

// : séparateur

ESPE : indicatif du champ

ACUPUNCTA : contenu du champ

Nous constatons que chaque carte est partagée en deux parties :

- une partie identification du document
- une partie contenu du document.

1. 2. 1. Identification du document : (O, BN, OSTR 1)

L'ordre d'écriture de cette partie est obligatoire.

1 La parenthèse limite les indications omises sur la carte perforée mais qu'il est nécessaire de restituer ici.

1.2.1.1. Numéro de séquence (O , 1 , 2 ...)

Il donne la place de la carte dans le groupe de cartes décrivant le document. C'est une numérotation continue à l'intérieur du document.

1.2.1.2. Sigle de la carte (BN)

Il permet de préciser la nature du document. Le programme Mistral a adopté les sigles suivants :

- . BN : création d'un document
- . BM : modification d'un document
- . BR : retrait d'un document.

1.2.1.3. Référence du document (OSTR 1 :)

Elle permet de le repérer dans le fichier. Elle doit être répétée sur chacune des cartes composant le document. Elle peut comporter tous les caractères alphabétiques et numériques à l'exception de la virgule et des deux points, 'e signe : deux points étant réservé pour signifier la fin de la référence. En l'absence des 2 points la carte est rejetée.

1.2.2. Contenu du document

Le contenu d'un document est un ensemble de champs. Plusieurs champs peuvent être placés sur une même carte.

1.2.2.1. Indicatif de champ : (ESPE)

Il est obligatoire et comporte de 1 à 8 caractères.

1.2.2.2. Séparateur de champ : (/ /)

Représenté par deux barres de fraction, il est obligatoire pour séparer deux champs.

1.2.2.3. Valeur du champ : (ACUPUNCTA)

C'est une chaîne de caractères comprise entre le nom du champ et le séparateur. Un mot-clé peut constituer une valeur de champ.

1.3. Gestion d'un stock

Le programme MISTRAL donne la possibilité de gérer un stock. Il offre trois possibilités.

- Sigle BN : Introduction d'un nouveau document.
Un document nouveau étant un document jamais introduit auparavant en mémoire.
- Sigle BM : Modification d'un document.
Ceci porte uniquement sur un document déjà introduit en mémoire.
On veut par exemple modifier le nom d'un auteur : BRADY à la place de MÜLLER
L'indication
2, BM, OSTR10 : // AUT BRAD remplacera
2, BN, OSTR10 : // AUT MÜLL.
- Sigle BR : Retrait d'un document.
Il faut que le document soit déjà en mémoire, par exemple le document 3.
L'opération de retrait s'écrit :
3, BR, OSTR10 : // ...

1.4. Edition d'un stock

1.4.1. Généralités

Le programme MISTRAL permet d'éditer le stock documentaire. Il offre plusieurs possibilités. Il permet l'établissement de listes dans lesquelles les documents sont triés par champs. Il conduit à la connaissance du contenu d'un ou plusieurs documents.

1.4.2. Fonctions de l'édition

1.4.2.1. Sélection des documents

Les documents édités peuvent être choisis

- soit en donnant une suite de références de documents
- soit en donnant les bornes des intervalles dans lesquelles les références des documents se trouvent.

1.4.2.2. Choix des champs

On peut éditer les champs choisis dans l'ordre désiré.

L'édition de la totalité du document sans connaître à priori les champs qu'il contient nécessite l'emploi du paramètre EDIT TOUT

1.4.3. Langage de commande

Nous verrons par la suite, lors de l'explication de l'exploitation MISTRAL, la formulation des questions.

2. Le thésaurus

Nous savons que le chercheur peut utiliser des mots-clés pour caractériser ses documents et formuler ses questions.

Le programme MISTRAL permet de gérer un thésaurus.

2.1. Définition du thésaurus

Un fichier est un ensemble organisé d'informations en nous plaçant uniquement au niveau logique sans préjuger de la nature du support des informations. La définition d'un mot-clé ayant été précisée, nous dirons qu'un fichier est dit "inverse" lorsqu'il conserve pour chaque mot-clé la liste des documents qui ont utilisé ce mot.

Dans le fichier d'ostracodes nous n'avons pas créé de champs "mots-clés" mais MISTRAL permet à chaque champ de disposer d'un fichier inverse et d'un thésaurus personnel.

Prenons le champ AUTEUR dont la valeur est BRAD (Y), chaque document à un champ AUTEUR dont la valeur diffère suivant les documents. Mais plusieurs documents peuvent avoir la même valeur BRAD (Y). BRAD (Y) est donc créé une seule fois dans le thésaurus et le fichier inverse conservera la liste des documents ayant utilisé ce nom d'auteur.

2.2. Création et action du thésaurus

Il est créé à l'aide du paramètre : THES.

Le thésaurus n'est pas un simple recueil de mots dont le documentaliste s'aiderait pour indexer au mieux les documents, il est employé pour répondre aux questions et il permet d'accélérer le processus de recherche d'une question.

De plus c'est grâce au thésaurus que les synonymies sont créées.

2.3. Gestion du thésaurus

Pour gérer un thésaurus le chercheur a deux possibilités :

- . tout mot-clé utilisé dans l'indexation d'un document et inconnu du thésaurus peut être incorporé automatiquement dans celui-ci.
- . tout mot-clé utilisé dans l'indexation d'un document et inconnu du thésaurus est rejeté.

Le programme Mistral peut gérer plusieurs thésaurus.

2.4. Edition du thésaurus

A l'aide de paramètres spéciaux il est possible d'éditer le thésaurus.

2.5. Conclusion

Dans ce travail nous n'avons pas eu à créer de thésaurus de mots-clés en raison de la structure des documents de base, mais nous avons muni la plupart de nos champs d'un fichier inverse et d'un thésaurus, ceci afin de ne pas avoir la gestion des thésaurus et de rendre le temps de réponses aux questions plus court.

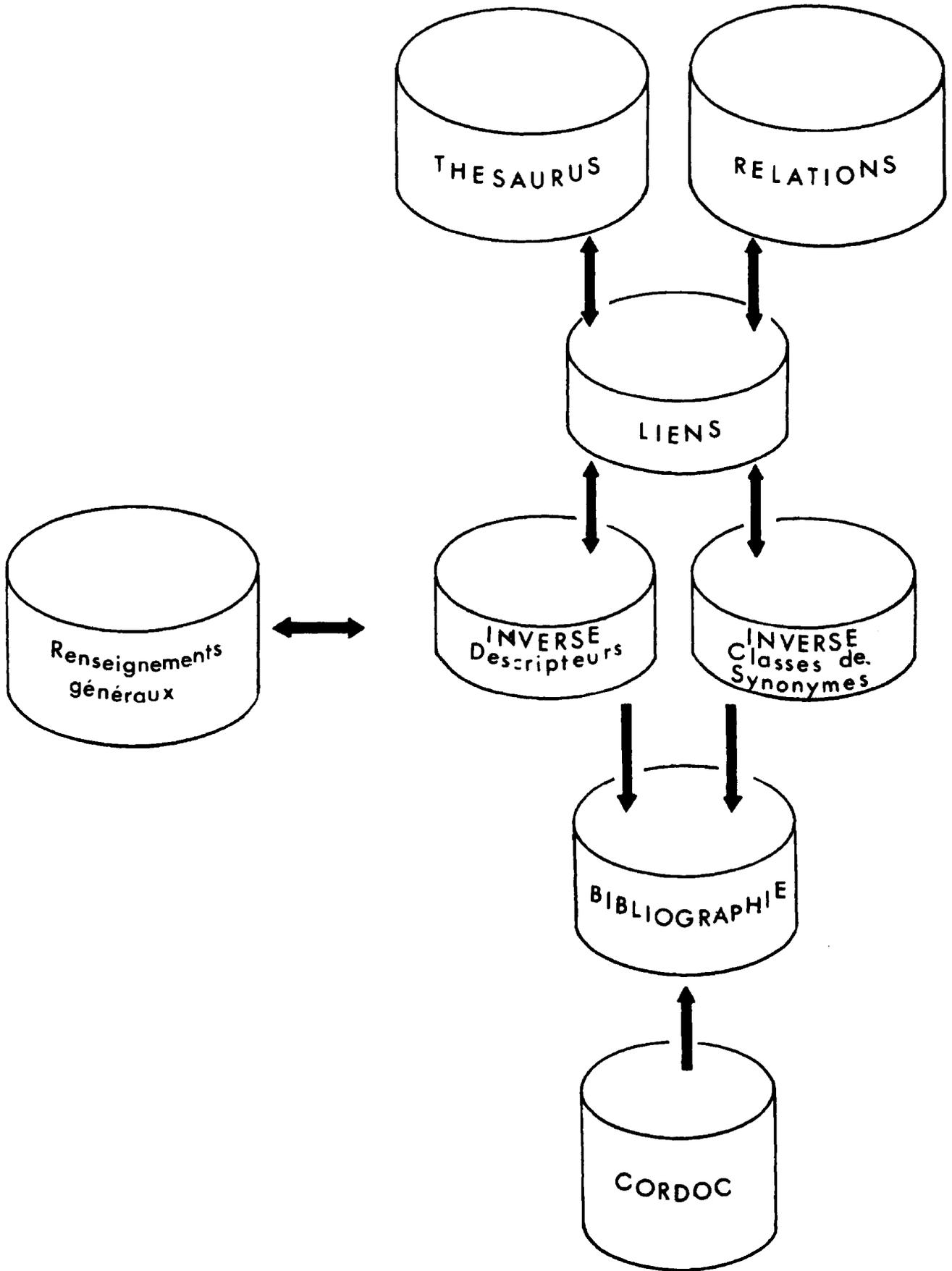
3. Implantation de MISTRAL

Le programme MISTRAL nécessite la constitution et l'utilisation de nombreux fichiers :

- Fichiers permanents
- Fichiers de manoeuvre
- Fichier de tri
- Fichier d'entrée - sortie.

3.1. Fichiers permanents (fig. 3)

Ils sont au nombre de onze et obligatoirement sur un disque. L'utilisateur doit en déterminer la taille en fonction du stock documentaire traité et des thésaurus. Pour calculer la taille de ces fichiers, la Compagnie d'informatique C. I. I. met à la disposition des utilisateurs des formules à appliquer. Les paramètres définissant la taille, le nom, le support des fichiers sont choisis par l'utilisateur à l'exécution de la chaîne de programme/d'initialisation des fichiers permanents.



ORGANISATION GENERALE

Fig. 3 - Organisation générale de fichiers utilisés lors de l'implantation des données.

Fichier des renseignements généraux.

Il contient la liste des champs, des domaines, les paramètres utilisés, etc...

Fichier bibliographie

Au nombre de 2, ils contiennent le stock documentaire.

Fichiers correspondance documents (CORDOC)

Egalement au nombre de 2 : un fichier principal et un fichier débordement.

Fichiers thésaurus

Au nombre de 2. Il contient les thésaurus définis par l'utilisateur.

Fichier relations

Contient les relations (synonymies) définies par l'utilisateur entre les mots du thésaurus.

Fichier inverse des descripteurs

Il contient par descripteur la liste des documents indexés.

Fichier inverse des synonymes

Contient par classe de synonymie, la liste des documents indexés.

Fichier des liens

Contient les liens internes existant entre les fichiers thésaurus, relations, inverse des descripteurs et synonymes.

3.2. Fichiers de manoeuvre

Ils sont au nombre de trois et sont utilisés pour des stockages temporaires. Ils peuvent être placés sur bande ou disque.

3.3. Fichiers de tri

Au nombre de trois, ils permettent d'exécuter le tri des documents. Le choix du tri est défini par carte paramètre.

3.4. Fichiers d'entrée - sortie

Fichiers contenant les données et les réponses.

La mise en oeuvre de ces fichiers n'est réalisable que par un informaticien.

CHAPITRE 3

EXPLOITATION PRATIQUE DE MISTRAL

1. Phase préliminaire

Elle comprend l'établissement du stock documentaire. L'établissement du thésaurus des mots-clés n'a pas été réalisé. En effet ayant à notre disposition un fichier d'ostracodes, fichier par ailleurs codé, monté sur cartes selon des normes qui n'étaient pas celles acceptées par Mistral, nous avons dû tout d'abord le trier suivant un ordre alphabétique choisi par le chercheur, et le monter sur support magnétique (bande magnétique).

Ensuite nous l'avons restructuré. Nous avons défini

- . la référence du document
- . les indicatifs des champs
- . la nature des champs.

Un programme FORTRAN a été établi permettant de constituer ce stock, de le monter sur bande et de le recopier sur disque.

2. Deuxième phase

L'informaticien implante tous les fichiers du programme MISTRAL sur les différents supports magnétiques ; il détermine ses paramètres et calcule la taille des fichiers. Nous avons utilisé le matériel de la compagnie d'informatique (C. I. I.) du centre de calcul interuniversitaire de Lyon - Saint-Etienne. Le fichier de données et les fichiers permanents de Mistral ont été montés sur disque M'D25 (25 millions d'octets).

3. Troisième phase : l'interrogation

Une question comporte trois parties :

- . Identification
- . Critères de sélection
- . Renseignements pour l'édition.

3.1. Identification

Elle comporte une seule carte. Sa structure est la suivante : NS, IDEN, JACARB.

NS : numéro de séquence obligatoire, car les cartes d'une question sont numérotées par ce nombre. Il sert à reclasser les cartes en cas de mélange. Les ruptures de séquences sont donc interdites. Ce numéro ne peut jamais être égal à zéro.
Un numéro de séquence existe sur chacune des cartes de la question.

IDEN. Sigle de la carte, il signifie identification ; c'est un paramètre.

JACARB : Référence de la question ; en général le nom du questionneur.

3.2. Commentaires

Cette zone est facultative et peut servir de titre au chercheur. Sa structure est : NS, COMM, Texte.

NS : Numéro de séquence

COMM : Paramètre, sigle de la carte.

3.3. Critères primaires

Les critères primaires ne s'appliquent qu'aux champs munis de fichiers inverses. Toute question doit comporter au moins un critère primaire sinon elle est rejetée. La structure de la carte est : NS, P, /ICP/, Texte.

NS : Numéro de séquence

- P** : Paramètre indiquant qu'il s'agit du critère primaire
/ICP/ : Indicatif de champ, toujours entre deux barres de fraction
Texte : L'énoncé du critère est la valeur du champ demandé, ou une expression booléenne où les termes sont les éléments entiers ou tronqués du champ inverse demandé.

Des critères sur deux champs différents doivent être écrits sur deux cartes différentes.

Opérateurs :

Les opérateurs utilisés sont :

- AND** : Implique la présence effective des deux éléments du champ
OR : Implique la présence d'un des deux éléments du champ
ANOT : Implique la présence du premier à l'exclusion du deuxième
+ : Troncature.

Pour être valables, les opérateurs doivent toujours être entourés d'un blanc.

3.4. Critères secondaires

Ils sont facultatifs et s'appliquent à l'un quelconque des champs munis ou non d'un fichier inverse.

Structure de la carte : NS, S, /ICP/, Texte.

- NS** : Numéro de séquence
S : Paramètre indiquant qu'il s'agit d'un critère secondaire
/ICP/ : Indicatif de champ toujours entre deux barres de fraction
Texte : L'énoncé du critère et la valeur du champ demandé ou une expression booléenne.

Opérateurs :

Nous avons ceux utilisés pour la sélection primaire

AND, OR, ANOT, +

Il faut y ajouter :

Les opérateurs s'appliquant à la valeur d'un champ :

> strictement supérieur à

< strictement inférieur à

= égal

DIF différent de

Les opérateurs s'appliquent à l'ensemble du champ

RCH : Recherche d'une chaîne de caractères donnés dans l'ensemble du champ

ABS : Absence d'une chaîne de caractères

NFLD : Prend la valeur vraie si le champ demandé est absent. Il peut être combiné avec OR et ANOT mais n'a pas de sens avec ET.

3.5. Récapitulations

Les récapitulations primaires ou secondaires permettent de combiner entre eux les différents critères de la question.

La récapitulation primaire concerne les critères primaires et la récapitulation secondaire les critères secondaires.

Structure de la carte

NS, RP, Texte récapitulation primaire

NS, RS, Texte récapitulation secondaire

NS : numéro de séquence

RP ou RS: paramètre indiquant la récapitulation demandée

Texte : expression booléenne.

La récapitulation primaire ou secondaire n'est obligatoire que dans le cas où l'une ou l'autre des récapitulations comprend plusieurs critères.

3.6. Edition

Cette commande permet de choisir la nature et l'ordre des champs imprimés ainsi que le nombre maximum de documents édités.

Structure de la carte

NS, EDIT, n, Texte ¹

NS	:	numéro de séquence
EDIT	:	paramètre indiquant l'édition
n	:	nombre de documents édités. Si ce chiffre est absent, le programme Mistral prendra par défaut le paramètre qui a été défini en 2ème phase.
Texte	:	liste des champs à éditer séparés par une virgule. Si l'on désire le document complet, on met le paramètre TOUT.

¹ Toute question doit se terminer par un astérisque.

CHAPITRE 4

MODES D'INTERROGATION - EXEMPLES

La question une fois établie, l'utilisateur a deux moyens de la transmettre à l'ordinateur à l'aide du programme RETROS :

- traitement par lots
- traitement en temps partagé (T. S.)

1. Traitement par lots

La question est perforée sur cartes.

Après dépôt de cette question au centre de calcul et dans un délai variable mais ne dépassant pas une journée, le listing contenant les résultats, fonction des paramètres d'édition, est récupérable (p. e. listing des figures 4 et 5).

On obtient également des statistiques de sortie, données par le programme Mistral (listing de la figure 6).

La figure 4 représente le listing de la question telle que l'on doit la perforer.

La ligne 1 identifie le travail (ici JACOB nom abrégé de l'informaticien)

Les lignes 2 et 3 posent le problème. On souhaite connaître les ostracodes des continents européens (codé 5) ou d'Afrique (codé 1).

On ne retient par contre que certains états codés

	410	=	Algérie
ou			
	480	=	Tunisie
ou			
	460	=	Maroc

La gestion des fichiers micropaléontologiques ...

AUTEUR1 : SALA
ANNECREATION : 966

AUTEUR1 : SALA
ANNECREATION : 966

AUTEUR1 : EOPA
AUTEUR2 : MA
ANNECREATION : 968
REVUE : 25
BIBLIOGRA : 36

AUTEUR1 : SALA
ANNECREATION : 966

AUTEUR1 : SALA
ANNECREATION : 966

AUTEUR1 : SALA
ANNECREATION : 966

Fig. 6 - Exemple d'un listing de résultats (partiels).

FORMULATION DE LA QUESTION

!MISTRAL
BONJOUR
POUR EXPLICATIONS TAPÉZ SOS
QUE DESIREZ-VOUS FAIRE ?
?Q
DOMAINE ?
?GEOLOGIE1
REFERENCE DE LA QUESTION
?11 DECMBRE YUGOSLAVIA
CRITERE PRIMAIRE
?/CONT/5
CRITERE PRIMAIRE
?/PAYS/300
CRITERE PRIMAIRE
?/ERE/3
CRITERE PRIMAIRE
?/PERI/0 OR 1
CRITERE PRIMAIRE
?
RECAPITULATION PRIMAIRE
?CONT AND PAYS AND ERE AND PERI
CRITERE SECONDAIRE
?/SALN/1 OR 2 OR 5 OR 4
CRITERE SECONDAIRE
?
RECAPITULATION SECONDAIRE
?
EDITION : NOMBRE MAXI. DE DOCUMENTS EN REponse (,LISTE DE CHAMPS)
?REFE,GEN,ESPE,AUT1,ANCR,REV,RUBR,TEMP,SUBST,VEGE,SALN,PERI

LISTING DE LA QUESTION

PARM IN:CARD,OUT:LPRT
OPTN JUMP:NO,LINE:70,DOMA:GEOLOGIE1
1,IDEN,11 DECMBRE YUGOSLAVIA
2,P,/CONT/5
3,P,/PAYS/300
4,P,/ERE/3
5,P,/PERI/0 OR 1
6,RP,CONT AND PAYS AND ERE AND PERI
7,S,/SALN/1 OR 2 OR 5 OR 4
8,EDIT,REFE,GEN,ESPE,AUT1,ANCR,REV,RUBR,TEMP,SUBST,VEGE,SALN,PERI*
QUE DESIREZ-VOUS FAIRE ?

Fig. 7 - Exemple d'un listing de questions et de résultats (partiels) à la suite d'une exploitation en temps partagé.

LANCEMENT DE LA QUESTION

?GØ

FICHER DES REPONSES

?R

*** SOS, NUMERO(S), MOT DE PASSE, S, P, C OU FIN
POUR L'INSTANT, LA REPONSE COURANTE EST LA PREMIERE

3 DOCUMENTS SATISFY PRIMARY CRITERIA

3 DOCUMENTS SATISFY ALL CRITERIA

3 USED KEY WORDS

1 CPU TIME (100/MIN)

63 I/O CALL

COMMANDE ?

?

VOTRE COMMANDE EST INCONNUE; NOUS VOUS LES RAPPELONS

*** SOS, NUMERO(S), MOT DE PASSE, S, P, C OU FIN
POUR L'INSTANT, LA REPONSE COURANTE EST LA PREMIERE

COMMANDE ?

?1,3

REFERENCE : OSTR 8221

AUTEUR1 : STCH

ANNECREATION : 963

REFERENCE : OSTR 7643

AUTEUR1 : KEIJ

ANNECREATION : 955

REFERENCE : OSTR 4892

AUTEUR1 : CARB

ANNECREATION : 969

COMMANDE ?

?FIN

QUE DESIREZ-VOUS FAIRE ?

?L

QUESTION COURANTE ABSENTE

QUE DESIREZ-VOUS FAIRE ?

?FIN

AU REVOIR

!L

Les différents programmes réalisés pour utiliser une console, dans le cadre du programme MISTRAL ont été écrits par H. Vigne ¹ et D. Balland ¹.

La formulation de la question se fait sous forme de dialogue. Elle est ensuite construite en mémoire centrale selon la syntaxe MISTRAL et transférée sur un fichier par la commande.

Le fichier des réponses est listé sur la console, éventuellement imprimé comme le montre le listing de la figure 7.

¹ que les auteurs remercient pour l'autorisation accordée d'utiliser ces programmes.

CONCLUSIONS

Comme l'ont montré les chapitres traitant de la mise en oeuvre du programme MISTRAL appliqué à la recherche micropaléontologique, sa réalisation est ardue. Elle nécessite une parfaite maîtrise des techniques informatiques les plus diverses pour aboutir à l'utilisation correcte du stock documentaire. De ce fait, le recours à ce mode de recherche ne doit pas être effectué pour un ensemble de données (ici unités taxinomiques valides ou non) inférieur à 10 000 unités.

Il ressort de ce travail que le micropaléontologiste "classique" ne peut s'affranchir de l'aide constante d'un informaticien du plus haut niveau. Cette collaboration est indispensable lors du montage des données d'un fichier micropaléontologique habituel (type catalogue ELLIS et MESSINA, fiches mécanalytiques ou cartes perforées) et lors de sa mise à jour périodique. Elle l'est également avant, lors de la définition des champs et de leur contenu. On évitera, de ce fait, la création d'un fichier plus ou moins inadapté aux règles de fonctionnement du programme MISTRAL. Ce fut le cas, en partie, avec le fichier ostracodes traité ici. Il avait, en effet, été conçu précédemment pour une mise en oeuvre sur trieuse mécanique de cartes perforées (type IBM 108) dont les exigences sont beaucoup plus contraignantes et plus limitées.

A l'inverse l'exploitation d'un fichier micropaléontologique ainsi monté est aisée et accessible à n'importe quel utilisateur.

Qu'il s'agisse de l'utilisation en temps partagé (T. S.) avec visualisation des résultats sur console, ou par lots (perforations des questions sur cartes), la facilité d'emploi est évidente sous réserve de suivre rigoureusement les conditions d'emploi.

La recherche documentaire micropaléontologique se trouve considérablement améliorée. Le temps d'investigation, fonction de la complexité des questions, est de l'ordre de 2/100 à 5/100 de minute d'unité centrale (Iris 80). Toutes les combinaisons entre les divers champs sont réalisables. Le micropaléontologiste doit souvent répondre à ce problème fondamental : quelles unités taxinomiques se trouvent à un moment et en un lieu donné, sous certaines conditions paléoécologiques ? La réponse à cette question est immédiatement fournie par le programme MISTRAL.

Testé sur des ostracodes, ce système de recherche documentaire peut être répliqué, sans aucune difficulté, sur des Foraminifères, ou d'autres microorganismes.

BIBLIOGRAPHIE

CHEVALIER Cl. (1974) - Les cartes perforées et les recherches paléontologiques et paléoécologiques. Applications aux ostracodes. Arch. Sc. Genève, vol. 25, fasc. 1, p. 101-112, 12 fig.

DAVID L. , POUYET S. et VIGNE H. (1974) - Documentation automatique et bryozoaires in "Bryozoa 1974", Docum. Lab. Géol. Fac. Sci. Lyon, H. S. 3, fasc. 2, p. 627-645, 10 fig.

M. JACOBZONE
Laboratoire de Probabilités et Statistiques
Université Lyon 1
43, bd du 11 novembre 1918
69621 VILLEURBANNE

G. CARBONNEL
Département des Sciences de la Terre
Université Lyon 1
et Laboratoire associé au C.N.R.S.
43 bd du 11 novembre 1918
69621 VILLEURBANNE