

# LA GESTION DOCUMENTAIRE RAPIDE au service d'une petite équipe de recherche

Jean-Olivier JOB

Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM)  
213 rue Lafayette, 75010 Paris

Touhama MOUHEICH

Arab Center for the Study of Arid Zones and Dry Lands (ACSAD)  
B.P. 2440, Damas, Syrie.

*Le système bibliographique BAC (Bibliographie Avec Commentaires) doit permettre de gérer un fonds documentaire de plusieurs milliers d'articles sur les sols salés. Le fonds lui-même est constitué par des publications, des mémoires ou des thèses, de la littérature souterraine, ou tout article traitant de l'étude des sols salés au sens le plus large du terme. Le système de gestion que nous avons choisi a été mis au point sur mille enregistrements qui constituaient les références permanentes du laboratoire d'analyse des sols de l'ACSAD (Arab Center for the Study of Arid Zones and Dry Lands, Damas, Syrie) en 1981. Le laboratoire ayant une vocation de formation, le but était de permettre à des stagiaires un accès sélectif et très rapide à des documents de base leur permettant d'aborder un aspect très précis de l'analyse des sols salés en laboratoire. Un premier langage à quatre niveaux a d'abord été utilisé sur des fiches cartonnées pendant deux ans [1]. Le système a été ensuite informatisé en utilisant le logiciel de gestion de base de données dBase III (Ashton-Tate), tout en gardant le principe de départ, à quatre niveaux. Le fonds s'est ensuite enrichi de nouveaux documents plus théoriques concernant l'étude des sols salés, dans le contexte sol-eau-plante et dans l'optique recherche appliquée. Nous présentons ici ce fichier qui contient actuellement deux mille sept cents références. Les statistiques de structure du fichier ont été établies sur les mille premiers enregistrements.*

*A system for classification and retrieval of scientific references is presented. The system has been originally developed at ACSAD (The Arab Center for the Study of Arid Zones and Dry Lands, Damascus, Syria). The first objective was to classify summaries of published papers or articles concerning the study and the analysis of salt affected soils [1]. Theses summaries were made by the authors and then sent to interested analysts as training facility. Soon appeared the need of classifying these summaries and the original references as well and a system was developed using punched cards. As the number of references and summaries increased, it became necessary to use a database management system, dBase III and to develop an original codification to classify the documents. This system, called BAC (for Bibliography and Comments) is presented in this paper.*

## 1. PRINCIPE DE CLASSEMENT

Le principe d'indexation retenu consiste à affecter à chacun des documents classés un code en six parties :

1. l'élément chimique ou physique étudié,
2. sa dynamique dans le système sol-eau-plante,
3. le type d'étude faite,
4. le milieu (sol, eau ou plante),
5. le lieu d'étude (terrain ou laboratoire...),
6. le type de document (article, thèse, rapport...).

Chaque partie du code définit un niveau représenté par un nombre de deux ou trois chiffres nécessaire à la description du contenu de l'article.

La partie 1 explore le champ des éléments dont l'étude est possible (calcium, texture, structure, fer, par exemple). Les parties 2 et 3 précisent le rôle de l'élément dans le milieu. Elles sont subdivisées verticalement pour situer à quel niveau de la dynamique géochimique on se place (exemple : l'origine, ensuite la migration, puis l'évolution...) et également horizontalement pour explorer tout le champ sémantique (migration = évaporation, drainage ou lessivage, par exemple). La partie 4 permet de cerner la discipline scientifique la plus concernée par l'étude. La partie 5 précise si l'étude est faite en laboratoire ou sur le terrain. La partie 6 sert à classer physiquement les articles. Les livres sont dans une bibliothèque, les revues sur une étagère, les rapports dans un classeur, etc.

66  
M

## 2. CODIFICATION DES DOCUMENTS

Chaque document est parcouru rapidement par un lecteur capable de juger son contenu. Il le résume suivant un cadre destiné à en faciliter la lecture, et lui attribue un code. Ce code est composé en explorant successivement les niveaux 1 à 6, un nombre de trois chiffres étant donné pour chaque niveau suivant le tableau suivant :

### Niveau 1 : élément étudié

Le principe retenu est de survoler les éléments étudiés dans le sol suivant l'ordre alphabétique des éléments d'abord et suivant leur degré de complexité physico-chimique ensuite.

Al : 100	Cl : 117	Na : 150
B : 102	Fe : 120	P : 156
Ba : 103	K : 125	S : 160
C : 104	Mg : 130	Si : 165
Ca : 110	N : 140	

  

Eléments de transition : 168	pH Redox : 178
Eléments rares : 169	Surface : 179
Sels en général : 172	Structure : 180
Argiles : 173	Température : 182
Matière organique : 174	Toutes propriétés : 190
Granulométrie : 175	Tous cations : 166
Conductivité : 176	Tous anions : 167
Capacité d'échange : 177	H <sub>2</sub> O : 170

Un astérisque est ajouté au numéro choisi si l'étude fait appel à des éléments radioactifs (ex. : N15 est représenté par 140\*). Si nécessaire, chaque sous-niveau peut être à son tour subdivisé, par exemple :

Fe : 120                      Fe ++ : 121                      Fe +++ : 122

pour préciser à quel degré d'oxydation se trouve la forme de fer qui est étudiée. Huit sous-niveaux sont définis actuellement. Ils permettent entre autres d'introduire les composés de Ca, Na, Mg... comme le gypse, le chlorure de sodium et d'autres sels importants.

### Niveau 2 : dynamique de l'élément

Origine : 20	Assimilation : 22	Distribution : 24
Migration : 21	Echange : 23	Evolution : 25

A ce niveau il est nécessaire de subdiviser le mot Distribution en deux sous-ensembles : la distribution verticale dans le profil du sol (241) et la distribution horizontale dans le paysage (242).

### Niveau 3 : type d'étude

Le principe est de classer les études des plus théoriques aux plus pratiques.

Théorie : 30	Expérimentation : 34	Matériel utilisé : 37
Simulation : 31	Echantillonnage : 34	Datation : 38
Description : 32	Analyse : 35	Archivage : 39
Taxonomie : 33	Interprétation : 36	

### Niveau 4 : milieu

Sols : 430	amendements : 431	érosion : 432
Plantes : 440	toxicité : 441	physiologie : 442
Eau : 450	évaporation : 451	irrigation : 452
	drainage : 453	

Ce niveau est subdivisé pour tenir compte des études spécifiques au milieu. Un élément n'est pas toxique pour le sol mais l'est pour les plantes, l'érosion est spécifique aux sols pour ne prendre que deux exemples.

### Niveau 5 : lieu

Situe la distance entre le chercheur et l'objet de recherche. Ce niveau détermine le type de matériel

utilisé, le type de résultats obtenus et l'utilisation que l'on peut en faire.

51 : in situ                      52 : au labo                      53 : télédétection

### Complément 6 : document

61 : publication                      62 : mémoire

## 3. EXEMPLES

Le code donné ne correspond pas en général à celui que l'on pourrait attribuer au vu du titre seul, ou à la recherche de descripteurs dans le résumé. Il ne tient pas compte des mots-clés attribués par les auteurs de l'article. En effet, la tendance est soit de définir trop de mots-clés, soit d'en définir dont le sens est trop général. Dans les deux cas, l'effet produit est un bruit dans les recherches automatiques par mot-clé ou par descripteur.

Seuls les niveaux significatifs sont pris en compte. Le code du niveau 1 est séparé des autres par une barre oblique pour indiquer que ce niveau est le principal.

Soit par exemple l'article ayant pour titre :

AUBERT G. - 1976. La morphologie des sols affectés par le sel. Reconnaissance, prévision, surveillance continue. In : Prognosis of Salinity and Alkalinity, F.A.O Soil Bull., n° 31, pp : 187-194.

Son code sera : 180/32-33-430-51,

180 car l'article étudie en détail la *structure* de chaque horizon des différents types de sols salés, 32 et 33 parce que c'est une étude *générale* et qu'elle aborde largement les problèmes de classification de ces sols (*taxonomie*), 430 puisque la *sol* est l'unique objet de l'étude et 51 parce qu'il s'agit de la morphologie visible du sol en place (*in situ*).

De même :

BELLAIR P. - 1954. Sur l'origine des dépôts de sulfate de calcium actuels et anciens. C.r., acad., Sci., séance du 27 oct., pp : 1059-1061.

sera codé : 112/20-32-430,

et :

COMBEAU A., VERDIER J., OLLAT D. - 1964. Remarques sur les facteurs de variation des limites d'Atterberg. Cah., ORSTOM, ser., pédol, II, n° 4, pp : 29-39.

sera codé : 180/35-430-52.

## 4. SYNTAXE

Chaque partie du code, ou codon, est reliée aux autres par une syntaxe implicite. Ainsi, « Théorie du lessivage du silicium dans le sol » ou « Simulation de l'échange Na-Ca » ont une syntaxe de type :

(niveau 3) *du* (niveau 2) *du* (niveau 1) *dans* (niveau 4).

C'est la syntaxe générale. Par contre, les articles « Etude générale de l'érosion » et « Etude de végétation par télédétection » auront respectivement une syntaxe du type :

(niveau 3) *de* (niveau 4)

et

(niveau 4) *au* (niveau 5)

## 5. AMBIGUÏTÉS ET REDONDANCES DU SYSTÈME

Si la place des éléments étudiés est naturellement en tête de codification, il subsiste à ce niveau une redondance apparente. C'est la place de l'eau qui est en même temps un constituant permanent du sol (eau hygroscopique), un élément extérieur (pluie, nappe...) et un milieu en soi (elle peut contenir tous

les autres éléments). On la retrouve donc aux niveaux 1 et 4, sans que cela soit gênant d'ailleurs. Il n'y a par contre aucune ambiguïté, c'est l'un des avantages du système.

## 6. PHYSIONOMIE DU FONDS DOCUMENTAIRE

Le fonds documentaire a été testé sur 1000 enregistrements qui constituaient les références permanentes du laboratoire d'analyse des sols de l'ACSAD (Arab Center for the Study of Arid Zones and Dry Lands) à Damas, en Syrie, en 1981.

a) Eléments étudiés (en nombre d'articles traitant d'un sujet donné, pour mille) :

- >5 et ≤ 10 : Structure.
- >10 et ≤ 20 : Al, B, F, Si, Cations, Granulométrie, Analyse physique.
- >20 et ≤ 30 : Cl, Fe, K, Eléments rares, Matière organique, pH.
- >30 et ≤ 50 : Ca, Mg, P, Tous résultats.
- > 50 : S (SO<sub>4</sub> surtout), Transition (Co, Ni, Mn essentiellement), H<sub>2</sub>O, Sels, Conductivité, Capacité d'échange.

b) Type d'étude :

- ≤ 10 : Taxonomie.
- >10 et ≤ 100 : Description, Echantillonnage, Etude expérimentale, Matériel, Simulation.
- > 100 : Théorie, Analyse.

On voit que les préoccupations essentielles du laboratoire sont la formation aux techniques analytiques et que les sujets de recherche traitent des sels dans le sol. Cette physionomie a évolué depuis, pour tenir compte de recherches nouvelles.

## 7. STRUCTURE DU FONDS DOCUMENTAIRE

Dans le tableau I sont comparées les structures de notre fonds, noté (1), et celle d'un laboratoire de recherche de physiologie comparée (D. GUINIER [1]), noté (2). Les paramètres sont comparés dans les deux cas pour les mille premiers articles des deux fonds.

On remarquera que tous les paramètres de BAC sont inférieurs à ceux qui sont cités par [1]. De nombreux ouvrages ont été saisis avec comme nom d'auteur un sigle (FAO, ORSTOM...) et comme titre un en-tête de chapitre, qui sont plus courts que les titres de publications.

Tabl. I : Structure comparée de deux fonds documentaires.

Paramètre	Fonds (1)	Fonds (2)
Nombre d'auteurs	1250	2000
Nombre de caractères moyen	19	28
Caractères par article	145	320
Longueur des titres	74	98
Nombre de revues	220	250
Nombre de mots-clés	—	1800
Nombre de codons	80	—

## BIBLIOGRAPHIE

[1] GUINIER (D.). — « Analyse statistique des structures, du contenu et de l'évolution d'un fichier bibliographique d'un laboratoire de recherche en biologie ». — *Documentaliste*, vol. 21, n° 3, mai-juin 1984, p. 101-105.

La différence la plus significative est le nombre de caractères par article qui est 2,2 fois plus petit dans le système BAC à cause de l'absence de mots-clés en grande partie.

## 8. EVOLUTION

Depuis sa création en 1981, le fonds BAC est passé de 800 à 2700 références. A l'origine, il y avait quatre niveaux de codification, séparés en 80 sous-niveaux. Le champ d'étude s'est ensuite étendu à la tolérance aux sels des espèces cultivées, à l'halophilie des plantes naturelles, à la modélisation du mouvement des sels dans le sol. Le nombre de sous-niveaux a été réduit à 66, par suppression de tous ceux dont l'occurrence était inférieure à 0,5 %, malgré l'apparition des thèmes nouveaux. Le nombre des niveaux est passé à 6.

## 9. PERFORMANCES

Les références sont stockées sur le disque dur d'un micro-ordinateur Micral 30. La saisie et l'édition se font en utilisant le logiciel de gestion de base de données dBase III. Le tri se fait sur le numéro de code par un fichier programme géré par le même logiciel à l'aide de menus qui sont présentés à l'utilisateur pour chacun des niveaux. L'utilisateur peut faire son choix, les enregistrements sont transférés dans un fichier de travail à partir duquel on peut éliminer les enregistrements en surnombre, les visionner.

Le bruit peut être réduit par l'interrogation elle-même. Soit, par exemple, une recherche sur le sodium échangeable dans les sols. On peut formuler l'interrogation « Sodium-échangeable », ou : 150/23. Le programme traduit la question par :

LIST RECNO( ) FOR «150/23» \$ CODEBAC

et on obtiendra l'enregistrement n° 207 codé 150/23.37.430.450, mais aussi le n° 165 codé 110,150/23.38.430. Il se produit un léger bruit car la référence n° 165 parle également du calcium échangeable. On évite ce bruit en interrogeant « Sodium échangeable seul » qui est traduit par le système par :

LIST RECNO( ) FOR «150/23» SUBSTR (CODEBAC,1,6)

et seul l'enregistrement n° 207 est listé. Le bruit peut donc être réduit au minimum. Le taux de silence dépend uniquement de la manière dont a été faite la codification. Le nombre de caractères moyens du code est 12 avec un écart type de 3,2.

## 10. CONCLUSIONS

Le système que nous avons décrit est très souple et permet à une petite équipe de chercheurs de gérer leurs références bibliographiques ou leur fonds documentaire rapidement avec un minimum de bruit et de silence. On peut à tout moment ajouter des niveaux de classification soit pour affiner la recherche, mais le gain en précision se fait nécessairement au détriment de la légèreté du procédé [3], soit pour introduire des thèmes nouveaux, sans dans ce cas alourdir le système.

Manuscrit reçu en janvier 1986.

[2] JOB (J.O.), MOUHEICH (T.). — « Le système BAC ». — Pub. tech. ACSAD/ORSTOM, novembre 1981, p. 1-5.

[3] MANIEZ (J.). — « Problèmes de syntaxe dans les systèmes de recherche documentaire ». — *Documentaliste*, vol. 20, n° 2, mars-avril 1983, p. 52-58.