RAIRO. INFORMATIQUE THÉORIQUE

M. TREHEL

Deux constructions équivalentes d'un jeu d'arrangements pour les fichiers inverses multi-indices

RAIRO. Informatique théorique, tome 12, nº 1 (1978), p. 3-14

http://www.numdam.org/item?id=ITA_1978__12_1_3_0

© AFCET, 1978, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Informatique théorique » 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.



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

DEUX CONSTRUTIONS ÉQUIVALENTES D'UN JEU D'ARRANGEMENTS POUR LES FICHIERS INVERSES MULTI-INDICES (*)

par M. Trehel (1)

Communiqué par R. Cori

Résumé. — D. E. Knuth présente la construction d'un jeu de permutations représentant chacune la structure d'un fichier inversé multi-indices. Ces fichiers inversés permettent de donner la réponse à toutes questions sur le fichier de base, en une seule consultation. On donne ici une interprétation de la construction de D. E. Knuth à partir de la notion de couplage dans les graphes bipartis.

V. Y. Lum [7] propose pour interroger un fichier à plusieurs caractéristiques de construire des fichiers inversés multi-indices. Ainsi, si un fichier porte sur les caractéristiques :

nom,

date de naissance, profession,

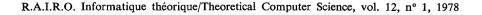
- il faut pouvoir rechercher les gens qui ont :
 - 1) tel nom,
 - 2) telle date de naissance,
 - 3) telle profession,
 - 4) tel nom et telle date de naissance,
 - 5) tel nom et telle profession,
 - 6) telle date de naissance et telle profession,
 - 7) tel nom et telle date de naissance et telle profession.

Un jeu de fichiers inversés multi-indices sera tel que, quel que soit le sousensemble de l'ensemble des caractéristiques, au moins un des fichiers inversés a pour premières caractéristiques celles de ce sous-ensemble.

Ainsi pour l'exemple ci-dessus, il faut et il suffit de 3 fichiers inversés. On peut par exemple prendre les 3 suivants :

- le 1er porte sur nom, date de naissance, profession;
- le 2e porte sur date de naissance, profession;
- le 3e porte sur profession, nom.

⁽¹⁾ Laboratoire d'analyse numérique et d'informatique, Faculté des Sciences et des Techniques, Route de Gray, Besançon.





^(*) Reçu janvier 1977, révisé septembre 1977.

4 M. TREHEL

Les questions 1), 4), 7) seront résolues par une consultation du 1^{er} fichier inversé.

Les questions 2), 6), seront résolues par une consultation du 2e fichier inversé.

Les questions 3), 5), seront résolues par une consultation du 3e fichier inversé.

Si on sait que ne seront posées que des questions portant sur f caractéristiques, il n'est pas utile de s'intéresser au jeu complet de fichiers inversés multi-indices. M. Gondran [4], D. K. Ray-Chaudhuri [8], proposent diverses solutions: les fichiers inversés à 2, 3, puis f indices et les fichiers combinatoires équilibrés.

On trouve dans les thèses de B. Brun [3] et de J. C. Grattarola [5] une étude exhaustive des organisations. Il y a en outre dans la seconde une implémentation de certains fichiers inversés multi-indices.

Si toutes les questions possibles risquent d'être posées, il faut trouver le jeu complet.

Notons I_n l'ensemble $\{1, 2, ..., n\}$. Le problème revient à trouver un jeu d'arrangements de I_n de sorte que, quel que soit le sous-ensemble x de I_n , il existe dans le jeu, au moins un arrangement dont les card (x) premiers éléments soient les éléments de x.

- V. Y. Lum impose, ce qui n'est pas indispensable, que ces arrangements soient des permutations de I_n ; il démontre qu'il faut et suffit de $C_n^{n'}$ telles permutations pour résoudre le problème (où $n' = \lfloor n/2 \rfloor$).
- E. Sperner [10] a prouvé en 1928 qu'il y avait au plus $C_n^{n'}$ sous-ensembles de I_n tels qu'aucun d'eux n'est contenu dans un autre. D. E. Knuth [6] constate que le fait de démontrer que $C_n^{n'}$ permutations suffisent pour les fichiers inversés multi-indices prouve par le fait même le théorème de Sperner. Plus récemment, T. C. Brown [2] a donné une démonstration élégante de ces résultats.

Connaissant le nombre d'arrangements il importe maintenant d'avoir un algorithme pour les énumérer. D. E. Knuth [6] propose une construction. Nous allons la rappeler puis voir comment, à partir de la notion de couplage, on peut en donner une interprétation nouvelle.

I. CONSTRUCTION DE D. E. KNUTH

Celle-ci est présentée sous forme d'exercice page 567 du tome 3 : Sorting and Searching. Des éléments de solution sont donnés page 699. Puisqu'il ne s'agit pas ici d'un exercice destiné au lecteur, nous allons utiliser une présentation différente.

Continuons de noter $I_n = \{1, 2, ..., n\}$.

Soit $x \in I_n$. La construction est destinée à donner un arrangement commençant par les éléments de x. Elle passe par la mise en place de 2 graphes de \mathbb{R}^2 que nous appelons chemin naturel et chemin spécial. Les chemins, naturel et spécial, sont des lignes brisées et continues formées de n segments.

Ils partent de (0,0).

Pour i = 1, 2, ..., n, le *i*-ième segment joint ou bien (i-1, j) à (i, j+1) et dans ce cas il est dit montant, ou bien (i-1, j) à (i, j-1), dans ce cas il est dit descendant.

A. Passage de x au chemin naturel

Si $i \in x$, le *i*-ième segment est descendant sinon il est montant.

B. Passage du chemin naturel au chemin spécial

Un segment montant reste montant, un descendant reste descendant sauf s'il descend vers une ordonnée négative non encore atteinte auquel cas il devient montant.

×	chemin naturel	chemin spécial	arrangement	×	chemin naturel	chemin spécial	arran g ement
Ø			A B 1234	{3}	3	<u>}</u>	A B 3 14
{1}	 v			{1,3}	}		
{1,2}	>			{ 1,3,4}	8		
{1,2,3}	1			{4}	<u>\</u>	<u>_</u>	A B 4 12
{1,2,3,4	K			{1,4}	*		
{2}	1	مما	A B 2 34	{1,2,4}	>		
{2,3}	+			{2,4}	L.		A B 24
{2,3,4}	+			{3,4}			A B 34

Figure 1.

6 m. trehel

C. Passage du chemin spécial à l'arrangement

On crée 2 piles A et B.

Pour i = 1, 2, ..., n, si le *i*-ième segment est montant dans le chemin spécial, empiler i dans B. S'il est descendant, empiler i dans A et désempiler B.

En mettant bout à bout les piles A et B, on obtient l'arrangement souhaité.

Exemples: n = 4 (voir figure 1 page précédente).

On n'a pas répété sur la figure 1 le chemin spécial et l'arrangement quand ils étaient les mêmes qu'au-dessus.

II. CONSTRUCTION PAR COUPLAGE

A. Rappels

La terminologie est celle de C. Berge [1].

Un graphe biparti est donné par le triplet (X, Y, U) où X et Y sont 2 ensembles de sommets avec $X \cap Y = \emptyset$ et U est un ensemble d'arcs dont l'origine est un sommet de X et l'extrémité un sommet de Y.

Un couplage d'un graphe biparti est un sous-ensemble U' de U tel que 2 arcs quelconques de U' ont leurs origines distinctes et leurs extrémités distinctes.

B. Couplage direct

On part aussi d'un graphe biparti (X, Y, U) avec la contrainte supplémentaires que X et Y sont totalement ordonnés :

$$X = \{x_1, x_2, \dots, x_p\},\$$

 $Y = \{y_1, y_2, \dots, y_a\}.$

L'algorithme de couplage direct de X vers Y est le suivant : on prend les x_i dans l'ordre. On couple x_i , si c'est possible, avec le premier y_j non encore couplé pour lequel il existe un arc (x_i, y_j) . Si on ne peut pas coupler x_i , on passe au x_i suivant, et cela jusqu'à x_p .

C. Définition de E_n^i

On ne parlera que des combinaisons d'éléments de I_n . Une combinaison sera écrite comme un mot formé de la suite croissante des éléments qu'elle contient. Ainsi pour n = 3, i = 2, les combinaisons sont 12, 13, 23.

Notons E_n^i , où $1 \le i \le n$ l'ensemble, ordonné lexicographiquement, des combinaisons de i éléments de I_n . $E_n^0 = \{\emptyset\}$. Chaque fois que nous parlerons de E_n^i , nous omettons de préciser, pour alléger l'écriture, que $0 \le i \le n$.

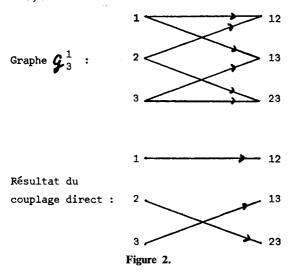
D. Graphes bipartis considérés

Considérons le graphe biparti dont les ensembles de sommets sont E_n^i , E_n^{i+1} , et l'ensemble U_n^i d'arcs est ainsi défini :

$$(x, y) \in U_n^i \leftrightarrow x \in E_n^i, y \in E_n^{i+1}, x \subseteq y.$$

Un tel graphe sera appelé \mathscr{G}_n^i

Exemple: n = 3, i = 1:



Profitons de cet exemple pour constater l'importance de l'ordre lexicographique dans E_{π}^{i} .

Ainsi, si les combinaisons de E_3^2 sont dans l'ordre 23, 13, 12, le couplage direct ne couple que deux arcs.

E Couplage dans \mathcal{G}_n^i

Avant de parler de couplage direct, vérifions que dans le couplage de \mathcal{G}_n^i , si i < n/2, toutes les combinaisons de E_n^i sont couplées et si $i \ge n/2$ toutes les combinaisons de E_n^{i+1} sont couplées.

Le nombre d'arcs partant de x appartenant à E_n^i est n-i. Le nombre d'arcs arrivant en y appartenant à E_n^{i+1} est i+1. Soit un sous-ensemble A de E_n^i de $a = \operatorname{card}(A)$ éléments. Notons $b = \operatorname{card}(\Gamma(A))$. Les arcs arrivant sur les sommets de $\Gamma(A)$ ne proviennent pas tous nécessairement de A et donc

$$a(n-i) \leq b(i+1)$$
.

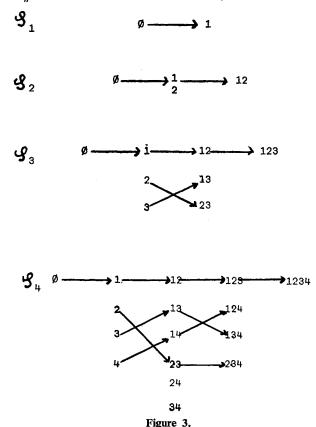
Si i < n/2, $n-i \ge i+1$ et donc $a \le b$ et le théorème de König-Hall permet de dire qu'il y a un couplage permettant de coupler tous les éléments de E_n^i .

vol. 12, nº 1, 1978

Si $i \ge n/2$, en prenant les arcs en sens inverse, on a pour tout B inclus dans E_n^{i+1} , card $(B) < \text{card } (\Gamma(B))$, et par le même théorème, il y a un couplage permettant de coupler tous les éléments de E_n^{i+1} .

F. Graphes \mathcal{S}_n

Nous appellerons \mathcal{S}_n le graphe obtenu avec les couplages directs dans $\mathcal{G}_n^0, \mathcal{G}_n^1, \ldots, \mathcal{G}_n^{n-1}$. Présentons $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4$.



G. Construction des arrangements des fichiers inversés multi-indices

Notons X l'ensemble des éléments de \mathcal{S}_n et Y l'ensemble de ses arcs.

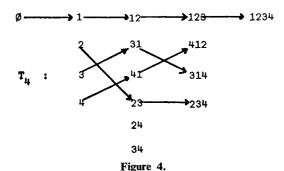
Le graphe T_n est construit à partir du graphe \mathcal{S}_n par l'isomorphisme h suivant :

si $x \in X$ et x n'a pas de prédécesseur alors h(x) = x;

si $x, y \in X$, $(x, y) \in Y$ et $y = x \cup \{a\}$, où $a \in I_n$ alors h(y) = h(x).a; (l'opération est à interpréter comme la concaténation de 2 arrangements pour en former un troisième).

L'ensemble des arrangements cherché est l'ensemble des sommets qui sont des extrémités de chemins de T_n .

Exemple:



ensemble des arrangements cherché : { 1234, 412, 314, 234, 24, 34 }.

On constate qu'on obtient le même ensemble d'arrangements que par la construction de D. E. Knuth. L'objet de la suite va être de justifier ce résultat dans le cas général.

III. PASSAGE DU CHEMIN NATUREL AU COUPLAGE DIRECT

Soit x une partie de I_n et card (x) = i. Notons K(x), si elle existe, la combinaison couplée avec x dans le couplage direct de \mathcal{G}_n^i .

Considérons le chemin naturel de x. Appelons sommets les origines et extrémités des segments successifs. Parmi les sommets de plus faible ordonnée, on appellera sommet sud-est celui qui est situé le plus à droite et sommet sud-ouest celui qui est situé le plus à gauche. Ils peuvent bien sûr être confondus.

Notons alors φ et ψ les 2 applications partielles suivantes :

- si le sommet sud-est est le dernier sommet du chemin naturel, $\varphi(x)$ n'est pas définie, sinon $\varphi(x) = x \cup \{a\}$, où a est le numéro du segment qui suit le sommet sud-est;
- si le sommet sud-ouest est le premier sommet du chemin naturel, $\psi(x)$ n'est pas définie, sinon $\psi(x) = x \{b\}$, où b est le numéro du segment qui précède le sommet sud-ouest.
- $\varphi(x)$ et $\psi(x)$ sont des combinaisons de I_n . Nous les noterons comme les autres combinaisons de I_n à savoir comme un mot formé de la suite croissante des éléments qu'elle contient.

Notre but est de montrer que φ et K ont même domaine de définition et dans ce domaine, on a $\varphi(x) = K(x)$.

Nous avons besoin pour cela de deux lemmes préalables.

LEMME 1 : $Si \psi (y)$ est défini, alors $\varphi . \psi (y) = y$.

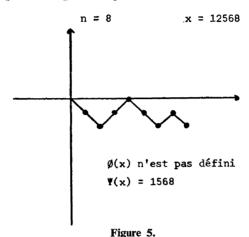
vol. 12, nº 1, 1978

10 m. trehel

Démonstration : Admettons que $\psi(y)$ est défini.

Soit Q le sommet sud-ouest et P le sommet qui le précède dans le chemin naturel de y.

Par définition du sommet sud-ouest, il n'y a aucun sommet plus bas que Q à droite de Q et plus bas que P à gauche de P dans le chemin naturel de y.



Dans le chemin naturel de ψ (y), les sommets à droite de P ont tous monté de 2 en ordonnée par rapport au chemin naturel de y. Le sommet sud-est

de 2 en ordonnee par rapport au chemin naturel de y. Le sommet sud-est du chemin naturel de $\psi(y)$ est donc P et puisque φ consiste à transformer de montant en descendant le segment qui suit le sommet sud-est, on a

$$\varphi.\psi(y)=y.$$

Remarque: On démontrerait de la même façon que:

si $\varphi(x)$ est défini alors $\psi \cdot \varphi(x) = x$.

LEMME 2 : Si $y = x \cup \{a\}$, où a est le numéro d'un segment montant à gauche du sommet sud-est dans le chemin naturel de x alors pour l'ordre lexicographique : $\psi(y) < x$.

Démonstration: Soit S le sommet sud-est dans le chemin naturel de x. Par définition, il n'y a pas à droite de S de sommet d'ordonnée inférieure ou égale à celle de S et à gauche de S de sommet d'ordonnée inférieure à celle de S.

Soit PQ un segment montant situé à gauche de S, a son numéro et $y = x \cup \{a\}$.

Dans le chemin naturel de y, l'ordonnée des sommets à droite de P a diminué de 2 par rapport au chemin naturel de x.

Dans le chemin naturel de y, le point sud-ouest est donc entre Q et S et il peut être Q ou S.

Ainsi $\psi(y) = y - \{b\}$, où b > a et b est inférieur ou égal à l'abscisse de S. $\psi(y)$ contient a et on a donc pour l'ordre lexicographique $\psi(y) < x$.

THÉORÈME 1: φ et K ont le même domaine de définition, et, dans ce domaine, on $a \varphi(x) = K(x)$.

Démonstration : Soit $x \in I_n$ et card (x) = i :

- si i = n, $\varphi(x)$ et K(x) ne sont pas définis;
- soit i < n.

On va faire une récurrence portant sur l'ordre lexicographique des combinaisons de E_n^i .

- Si x est la première combinaison de E_{x}^{i} ,

$$x = 12...i$$
, $K(x) = 12...i(i+1)$.

Dans le chemin naturel de x, les i premiers segments sont descendants les autres montants. Le sommet sud-est suit le i-ième segment et précède le i+1-ième. On a alors : $\varphi(x) = 12 \dots i(i+1)$ et $\varphi(x) = K(x)$.

Supposons que pour chaque combinaison w inférieure à x dans E_n^i , on a $K(w) = \varphi(w)$.

Vérifions que $K(x) = \varphi(x)$.

• Vérifions d'abord que x ne peut pas se coupler avec $y = x \cup \{a\}$ où a est le numéro d'un segment montant à gauche du sommet sud-est.

C'est naturellement vrai si le sommet sud-est est le premier sommet.

Sinon on a par le lemme 2 $x_1 = \psi(y) < x$.

Par hypothèse de récurrence $K(x_1) = \varphi(x_1)$ et par le lemme $1 \varphi(x_1) = y$. On ne peut donc pas avoir K(x) = y.

- Si le sommet sud-est est le dernier sommet, $\varphi(x)$ n'est pas défini. Par ce qu'on vient de voir K(x) ne peut pas être égal à $x \cup \{a\}$, où a est le numéro d'un segment à gauche du sommet sud-est et donc K(x) n'est pas défini non plus.
- Si le sommet sud-est est suivi d'un segment de numéro b on a $\varphi(x) = x \cup \{b\}.$

 $x \cup \{b\}$ ne peut pas être déjà couplé puisque le couplage direct se fait dans l'ordre lexicographique des combinaisons et pour

$$w < x$$
 on a $\varphi(w) = K(w)$.

Donc $x \cup \{b\}$ est libre et $K(x) = x \cup \{b\}$.

On a bien $K(x) = \varphi(x)$.

IV. ÉQUIVALENCE ENTRE LES DEUX CONSTRUCTIONS

NOTATIONS: Les piles de D. E. Knuth pour construire l'arrangement pour une combinaison x donnée seront notées A(x) et B(x).

A(x) et B(x) sont des arrangements disjoints. La concaténation A(x). B(x) est donc un nouvel arrangement.

12 m. trehel

On appelle combinaison initiale une combinaison sans prédécesseur dans \mathcal{S}_n et une combinaison finale une combinaison sans successeur dans \mathcal{S}_n .

Toute combinaison de \mathcal{S}_n étant sur un chemin et un seul, on appellera combinaison initiale (respectivement finale) de x dans \mathcal{S}_n , l'origine (respectivement l'extrémité) du chemin sur lequel se trouve x dans \mathcal{S}_n .

Pour le graphe T_n , on parlera d'arrangement initial, d'arrangement final, d'arrangement initial de x, d'arrangement final de x.

Puisqu'on écrit une combinaison comme la suite croissante des éléments qu'elle contient, nous ne distinguerons pas la combinaison initiale de x dans \mathcal{S}_n et l'arrangement initial de x dans T_n .

OBJET DE CETTE PARTIE : Il s'agit de montrer que les 2 constructions donnent le même résultat, c'est-à-dire que si z est l'arrangement final de x, on a A(x). B(x) = z.

Nous avons besoin des lemmes suivants :

LEMME 3: Si t est une combinaison initiale, A(t) = t.

Démonstration: Par le lemme 1 et le théorème 1, $\psi(t)$ n'est pas défini. Ainsi, dans le chemin naturel de t, le sommet sud-ouest est (0, 0), aucun segment ne descend vers une ordonnée négative et le chemin naturel est égal au chemin spécial. L'ensemble des numéros des segments descendants est t et A(t) = t.

LEMME 4 : Si K(x) est défini, le chemin spécial pour K(x) est le même que pour x.

Démonstration: Si s est le segment qui suit le sommet sud-est dans le chemin naturel de x, rappelons que le chemin naturel de K(x) est obtenu en remplaçant le segment s montant par un segment descendant. Ce segment descend vers une ordonnée non encore atteinte. Il est donc remplacé dans le chemin spécial de K(x) par un segment montant. Les chemins spéciaux pour x et K(x) sont alors les mêmes.

Pour étudier A(x) et B(x), il suffit donc d'étudier A(t) et B(t) où t est la combinaison initiale de x.

LEMME 5 : Si t est une combinaison initiale, B(t) est constitué de la suite des numéros des segments qui montent vers une ordonnée inférieure ou égale à l'ordonnée de tous les éventuels sommets postérieurs.

Démonstration: On a vu au lemme 3 que les chemins naturels et spécial de t étaient les mêmes. Or par la construction de D. E. Knuth, on empile dans B(t) les numéros des segments montants. Mais on désempile un numéro chaque fois qu'on rencontre un segment descendant. Cela revient à garder dans B(t) les numéros des segments qui montent vers une ordonnée inférieure ou égale à celle de tous les éventuels sommets postérieurs.

THÉORÈME 2 : La construction par couplage donne le même ensemble d'arrangements que la construction de D. E. Knuth.

Démonstration: Il suffit de démontrer que si z est la combinaison finale de x, on a A(x). B(x) = z.

Soit t la combinaison initiale de x.

Par les lemmes 3 et 4 on a A(x) = t.

Le lemme 5 revient à dire que les éléments successifs de B(x) sont les numéros successifs des segments qui suivent les sommets sud-est dans les combinaisons successives du chemin de \mathcal{S}_n partant de t.

Les numéros concaténés à t pour former z sont donc les numéros concaténés à A(x) pour former A(x). B(x) et dans le même ordre, on a: z = A(x). B(x).

Remarque: Il est intéressant de noter que si la combinaison initiale a j éléments le chemin naturel arrive à l'ordonnée n-2j. Pour passer au chemin naturel de la combinaison finale, n-2j éléments montants vont changer de sens. La combinaison finale contient donc j+(n-2j) éléments, soit n-j.

INDEX DES NOTATIONS

Notations	Paragraphe ou la notation est introduite		
Arrangement initial (final) de x dans T_n	début de IV		
A(x)	IC et début de IV		
B(x)	»		
Chemin naturel	ΙA		
Chemin spécial	I B		
Combinaison initiale (finale) de x dans \mathcal{S}_n	début de IV		
E_n^i	II C		
$\mathcal{G}_n^{\tilde{l}}$	II D		
I_n	Introduction		
$\ddot{K}(x)$, $\varphi(x)$, $\psi(x)$	début de III		
\mathscr{G}_n	${f II}$ ${f F}$		
T_n	II G		

REMERCIEMENTS

L'auteur remercie chaleureusement R. Cori dont les suggestions lui ont permis, entre autres, d'alléger notablement la démonstration.

Note: L'auteur apprend, alors que l'article est déjà imprimé, que le théorème 1 avait déjà été démontré sous une forme différente, par M. AIGNER, Lexicographic Metchings in Boolean Algebras, J. Comb. Theory, B, vol. 14, n° 3, juin 1973, p. 187-194. Le lien avec la construction de D. E. Knuth et l'application aux fichiers inversés multi-indices ne sont pas présentés dans l'article de M. Aigner.

BIBLIOGRAPHIE

- 1. C. Berge, Graphes et hypergraphes, Dunod, Paris, 1970.
- 2. T. C. Brown, A Proof of Sperner's Lemma via Hall's Theorem, Math. Proc. Camb. Phil. Soc., vol. 78, 1975, p. 387.
- 3. B. Brun, Fichiers inversés sur une caractéristique en vue de la réponse à nue f-question, Thèse d'Université, Besançon, janvier 1976.
- 4. M. Gondran, Structure combinatoire de fichiers, Électricité de France, Bulletin de la direction des études et des recherches, Série C, Mathématiques-Informatiques, nº 2, 1971, p. 55-80.
- 5. J. C. GRATTAROLA, Conception et réalisation d'un système d'interrogation de fichiers en temps réel, Thèse de spécialité, Nice, novembre 1976.
- 6. D. E. KNUTH, Sorting and Searching, vol. 3 de la collection "The Art of Computer Programming", Addison-Wesley, Reading, U.S.A., 1973.
- 7. V. Y. Lum, Multi Attribute Retrieval of Combined Indexes, Comm. A.C.M., vol. 13, no 11, novembre 1970, p. 660-665.
- 8. D. K. RAY-CHAUDHURI, Combinatorial Information Retrieval Systems for Files, S.I.A.M. J. Appl. Math., vol. 16, no 5, 1968, p. 973-992.
- 9. B. Shneiderman, Reduced Combined, Indexes for Efficient Multiple Attribute Retrieval (à paraître dans Information Systems, vol. 2, nº 4).
- 10. E. Sperner, Ein Satz über untermengen einer endlichen Menge, Math. Zeitschrift, 27, 1928, p. 544-548.