THESE

présentée à

L'UNIVERSITE PARIS IX DAUPHINE
U.E.R. SCIENCES DES ORGANISATIONS

pour obtenir

LE DIPLOME DE DOCTEUR EN INFORMATIQUE

par

Djamel Eddine ZEGOUR

Directeur de recherche
Gérard LEVY

Sujet de la thèse :

EXTENSIONS DU HACHAGE DIGITAL : HACHAGE DIGITAL MULTINIVEAUX HACHAGE DIGITAL AVEC REPRESENTATIONS SEQUENTIELLES

Soutenue le 21 juin 1988 devant la commission composée de :

MM	**C. BERTHET**	Président
	C. DELOBEL	Examinateurs
	E. DIDAY	
	G. LEVY	
	W. LITWIN	
	E. PICHAT	

Résumé

Le hachage dynamique a pris naissance dés l'année 1978 avec les premières propositions de Litwin et Larson. Contrairement au hachage classique, la fonction d'accès générée par la méthode varie dynamiquement de telle sorte que l'espace des adresses possibles ne soit plus limité. Depuis, plusieurs techniques d' accès basées sur ce principe ont été developpées principalement pour gérer des fichiers dynamiques. Certaines d'entre elles préservent en plus l' ordre des clés et/ou permettent de gérer des données multidimensionelles.

Le hachage digital , en anglais "trie hashing " conçu en 1981, est une technique d'accès aux fichiers monoclés dynamiques et ordonnés basée sur le concept du hachage dynamique. Son facteur de chargement est de l'ordre de 70%. En plus, un accès disque au plus suffit pour retrouver tout article pouvant appartenir à un fichier d'une contenance moyenne de quelques millions d'articles.

Cette thèse propose une extension du hachage digital aux fichiers volumineux. La nouvelle technique a été baptisée hachage digital multiniveaux. Sa formulation consiste à segmenter l'arbre représentant la fonction d'accès, ne pouvant résider longtemps en mémoire principale, en pages sur le disque. L'organisation correspondante est alors similaire à celle des arbres B. L'analyse des performances d'accès montre que deux accès par recherche de clé suffisent pour des fichiers de plus de 500 000 000 d'articles. Ainsi le hachage digital multiniveaux se situe parmi les techniques d'accès aux fichiers les plus efficaces connues actuellement.

Cette thèse propose également des représentations très conçises de la fonction d'accès engendrée par le hachage digital. Avec ces nouvelles représentations, on peut doubler le contenu du fichier pour le même espace alloué pour la fonction d'accès que si une représentation normale, appelée représentation standard, est utilisée. En contrepartie, l'algorithmique est plus complexe et le temps de calcul d'une adresse est plus long. Cependant, elles peuvent être avantageuses pour certaines applications.

Mots clés

Structures de données - Algorithmes - Méthodes d'accès - Arbres B - Hachage classique - Hachage dynamique - Hachage digital - Hachage digital multiniveaux - Représentations séquentielles.

SOMMAIRE

Première partie

Introduction

Deuxième partie

Les méthodes d'accés

Troisème partie

Le hachage dynamique

Quatrième partie

Hachage digital multiniveaux

Cinquième partie

Hachage digital avec représentations séquentielles

Sixieme partie

Conclusion

Annexes