MINISTERE DE L' EDUCATION NATIONALE

SECRETARIAT D' ETAT A LA RECHERCHE

THESE

Présentée au

CENTRE DE DEVELOPPEMENT DES TECHNOLOGIES AVANCEES

Pour l'obtention du grade de

MAGISTER EN CYBERNETIQUE

(Option : Architecture des systèmes)

OPTIMISATION DU STOCKAGE ET DE L' ACCES DANS LES BASES DE DONNEES TRES VOLUMINEUSES

Par

Mme Fatma-Zohra BESSAI
née MECHMACHE

Soutenue le 15 Novanbre 1992 Devant le jury composé de

Mr H. Bessalah	.....	Président
Mr A. Abdellaoui	.....	Examinateur
Mme Z. Ali-Mazighi	.....	Examinateur
Mr M. Benhamadi	.....	Examinateur
Mme F. Cherief	.....	Examinateur
Melle F. Azrou	.....	Rapporteur

Remerciements

Je tiens a remercier,

Monsieur H. BESSALAH, Directeur du centre de developpement des technologies avancees (CDTA), pour avoir bien voulu me faire l'honneur de presider le jury.

Monsieur M. BENHAMADI, Directeur du CERIST, pour avoir mis a ma disposition les moyens necessaires ***8*** la bonne marche de mon travail, pour la confiance qu'il m'a temoignee, et pour sa participation au jury.

Je suis tres honoree de la participation a ce jury de Madame Z. ALI-MAZIGHI, chargee de cours **à** l'USTHB, Madame F. CHERIEF, maître de conference (USTHB) et de Monsieur A. ABDELLAOUI, maitre de conference (USTHB).

Je tiens egalement a remercier Mademoiselle F. AZROU, chargee de recherche au CERIST, pour avoir propose le sujet de recherche et dirigé mon travail.

Je ne peux pas oublier ici H. LABIOD, Ingénieur au CERIST, pour son aide et sa collaboration pendant l'intégration de mon travail au systeme SIBADOC.

Je ne saurais clore ces pages sans adresser mon expression de sympathie la plus sincere a tous mes collegues et amis du Laboratoire de Recherche et Developpement en Informatique (LRDI), ainsi qu'à l'ensemble du personnel du CERIST. Que toutes celles et ceux qui m'ont témoigné leur disponibilité et apporte leur aide trouvent ici une expression de reconnaissance.

Enfin je tiens a remercier tous mes proches pour leur soutien et patience tout au long de l'élaboration de ce travail. Qu'ils trouvent ici une recompense a leurs efforts.

RESUME

La masse d'information mise a la disposition de la communaute scientifique s'accroît de façon continuelle, etant donnée l'évolution et la diversification des disciplines scientifiques et techniques. Cette grande masse d'information nécessite l'optimisation des moyens de stockage et le developpement de nouvelles methodes d'accès car les methodes classiques se revelent insuffisantes.

L'objectif de notre travail est de contribuer a l'amelioration des techniques d'accès aux bases de donnees volumineuses. Or une methode d'accès est etroitement liée a la methode de stockage utilisée. Aussi, une partie de notre travail a consiste a etudier le compactage de donnees, utilisé au CERIST pour resoudre le probleme de stockage de grands volumes de donnees.

Apres l'étude des diverses techniques de compression, notre travail a essentiellement consiste a concevoir et à realiser un systeme d'accès aux donnees volumineuses. Ce systeme utilise une technique d'indexation basee sur l'inversion totale, ce qui permet aux utilisateurs d'accéder a la base de donnees par n'importe quel mot contenu dans cette dernière et offre un acces rapide a l'information grâce a l'utilisation du Q-arbre qui est un arbre équilibré de profondeur 1. En plus de ces caracteristiques, le systeme a une structure modulaire lui permettant de s'adapter a tous les systemes gerant de grands volumes de donnees textuelles.

MOTS CLES

Base de donnees documentaire, systeme de recherche documentaire, systeme de gestion d'objets complexes, stockage des donnees, acces aux bases de donnees, compression de donnees, indexation, Q-arbre.

ABSTRACT

The amount of scientific information is growing steadily. This huge amount of information requires efficient storage management and new access methods.

Our goal is to improve the current access techniques to very large data bases. Since an access technique is closely related to the storage method used, in the first part of our work we study the data compression techniques already used by the CERIST in order to store a large amount of information.

After an analysis of the current techniques of data compression, we concentrate on the design and implementation of an access system to very large data bases. Our system uses an indexing scheme based on full inversion which allows users to query the data base using any word contained in the data base.
This system is based on the Q-tree (B-tree of depth one) that indeed allow fast access to the sought information. In addition, the system modular structure makes it adaptable to all system managing very large textual data.

KEY WORDS

Bibliographic data base, document-retrieval system, complex object management system, data storage, data base access, data compression, indexing, Q-tree.

S O M M A I R E