

# République Algérienne Démocratique et Populaire

Ministère de l'Enseignement  
Supérieur

Institut National de Formation  
en Informatique

*Mémoire de fin d'Etudes*

**Pour l'obtention du diplôme d'ingénieur d'état en  
informatique**

Option : Système

**THEME**

**Intégration d'une technique de compression  
de données à un système  
de recherche documentaire**

Présenté par : Mr. MECHKOUR Mourad

Organisme d'accueil : Centre de Recherche sur l'Information  
Scientifique et Technique ( CE.R.I.S.T. )

PROMOTION : 1989

# SOMMAIRE

## INTRODUCTION

### CHAPITRE I LA COMPRESSION DE DONNEES

I.1	DEFINITION . . . . .	I-1
I.2	UTILISATION DE LA COMPRESSION DE DONNEES . . . . .	I-1
I.3	CONCEPTS DE BASE DE LA COMPRESSION DE DONNEES . . . . .	I-2
I.3.1	Le code et ses propriétés . . . . .	I-3
I.3.2	Modèle pour la compression de données . . . . .	I-5
I.3.2.1	Quantité d'information et optimalité d'un code . . . . .	I-5
I.3.2.2	Le schéma de compression . . . . .	I-7

### CHAPITRE II REDONDANCES DES DONNEES ET CLASSIFICATION DES TCD

II.1	LES REDONDANCES DANS LA REPRESENTATION DES DONNEES . . . . .	II-1
II.2	CLASSIFICATION DES TECHNIQUES DE COMPRESSION DE DONNEES (TCD) . . . . .	II-2
II.2.1	Classification par élément de base de la codification . . . . .	II-3
II.2.1.1	Les caractères . . . . .	II-3
II.2.1.2	Les bigrams . . . . .	II-3
II.2.1.3	Les fragments . . . . .	II-4
II.2.1.4	Les mots . . . . .	II-5
II.2.2	Classification suivant la fonction de correspondance . . . . .	II-5
II.2.2.1	Les TCD sémantiquement dépendantes . . . . .	II-6
II.2.2.1.1	Définition . . . . .	II-6
II.2.2.1.2	Les TCD les plus usuelles . . . . .	II-6
II.2.2.1.2.1	La différenciation . . . . .	II-6
II.2.2.1.2.2	Codage des répétitions . . . . .	II-7
II.2.2.1.2.3	Compression programmée . . . . .	II-7
II.2.2.1.2.4	La compression des nombres . . . . .	II-9
II.2.2.1.2.5	Compression des indexes . . . . .	II-10
II.2.2.2	Les TCD sémantiquement indépendantes . . . . .	II-11
II.2.2.2.1	Définition . . . . .	II-11
II.2.2.2.2	L'algorithme de HUFFMAN . . . . .	II-12
II.2.2.2.3	Les TCD statiques . . . . .	II-15
II.2.2.2.4	Les TCD dynamiques (adaptatives) . . . . .	II-17

## CHAPITRE III LES DONNEES BIBLIOGRAPHIQUES ET LES TCD APYHOPRIEES

III.1	ETUDE DES ED DOCUMENTAIRES . . . . .	III-1
III.1.1	Données numériques . . . . .	III-2
III.1.2	Données textuelles . . . . .	III-2
III.1.3	Données non typées . . . . .	III-4
III.2	COMPRESSION DES DONNEES BIBLIOGRAPHIQUES . . . . .	III-5
III.3	DESCRIPTION DU SCHEMA DE COMPRESSION . . . . .	III-6
III.3.1	Compression des données numériques . . . . .	III-6
III.3.2	Compression des données nontypées . . . . .	III-7
III.3.3	La compression des données textuelles . . . . .	III-7

## CHAPITRE IV INTEGRATION DE LA TCD AU SYSTEME DOCUMENTAIRE

IV.1	LA RECHERCHE DOCUMENTAIRE ET LES SYSTEMES INFORMATIQUES DE R.D. . . . .	IV-1
IV.2	DESCRIPTION DU SYSTEME SIBADOC . . . . .	IV-4
IV.2.1	Structure de la base . . . . .	IV-4
IV.2.1.1	Le thésaurus . . . . .	IV-5
IV.2.1.2	Le fichier bibliographique . . . . .	IV-5
IV.2.1.3	Le fichier inverse . . . . .	IV-6
IV.2.2	Modules fonctionnels du système . . . . .	IV-7
IV.2.2.1	Module de création et de mise à Jour . . . . .	IV-7
IV.2.2.2	Module d'interrogation de la Base . . . . .	IV-9
IV.3	INTEGRATION DE LA TCD A SIBADOC . . . . .	IV-10
IV.3.1	La compression . . . . .	IV-11
IV.3.2	La décompression . . . . .	IV-13

CONCLUSION

BIFLIOGRAPHIE