

THESE

1131

C 314

présentée à

Institut National Polytechnique de Grenoble



pour obtenir le grade de

DOCTEUR-INGENIEUR

**CONTRIBUTION A L'ETUDE DES
TECHNIQUES DE COMPRESSION DES DONNEES
ET DE LEURS DOMAINES D'APPLICATION**



Salwa Abdel RAZEK (née ZAKARIA A.W. EL SHARAWY)



Thèse soutenue le 9 novembre 1976 devant la Commission d'Examen

Président : L. BOLLINET

J.P. ANSART

G. CULLMAN

Examineurs : C. DELOBEL

G. LOUIS-GAVET

1131/8956

S O M M A I R E

CHAPITRE 1. INTRODUCTION

1. Intérêt du compactage p. 1
2. Les différentes méthodes utilisées pour compacter les données p. 3
3. Buts de l'étude envisagée p. 5
4. Conclusion p. 6

CHAPITRE 2. DESCRIPTION GENERALE DES PRINCIPALES METHODES DE COMPACTAGE

1. Introduction p. 7
2. Méthodes basées sur la suppression des caractères redondants p. 8
 - 2.1. Présentation générale de ces méthodes p. 8
 - 2.2. Méthode des tabulations logiques p. 9
 - 2.3. Méthode par troncation p. 11
 - 2.4. Méthode par segmentation p. 12
 - 2.5. Méthode par délimiteur p. 14
 - 2.6. Méthode par représentation de tables binaires p. 15
3. Méthodes basées sur une étude statistique du corpus des données p. 18
 - 3.1. Introduction p. 18
 - 3.2. Méthode par substitution p. 19
 - 3.3. Méthode de Brilloin p. 22
 - 3.4. Méthodes de Shannon-Fano et Huffman p. 24
 - 3.5. Méthode de Synderman et Hunt p. 26

4. Méthode basée sur l'étude des langages structurés	p. 30
5. Méthode basée sur des calculs arithmétiques de codes	p. 31

CHAPITRE 3. ETUDE DES EXPERIENCES ET EVALUATION DES METHODES
DE COMPACTAGE REALISEES EN VUE DE LEUR APPLICATION
SUR DES LANGAGES DETERMINES

1. Introduction. Les différentes mesures effectuées	p. 32
2. Evaluation sur les méthodes classiques	p. 36
2.1. Méthode de compactage par troncation : Chopping	p. 36
2.2. Compactage par réduction des caractères répétés	p. 40
2.3. Méthodes liées à la théorie de l'information	p. 43
2.4. Méthode de Hahn	p. 71
2.5. Méthode de Louis-Gavet	p. 89
3. Evaluation sur des méthodes combinées	p. 94
3.1. Introduction	p. 94
3.2. Méthode basée sur les calculs arithmétiques de codes et les délimiteurs	p. 96
3.3. Méthode basée sur des statistiques et des délimiteurs	p. 98
3.4. Méthode basée sur des statistiques et des truncations	p. 100
3.5. Méthode basée sur les truncations et la méthode Louis-Gavet	p. 102
4. Méthode de changement de base de numération	p. 104
4.1. Description de la méthode	p. 104
4.2. Résultats obtenus	p. 116
5. Conclusion	p. 118

CHAPITRE 4. SYNTHÈSE ET CRITIQUE DES DIFFÉRENTES MÉTHODES
DE COMPACTAGE EXPÉRIMENTÉES

- 1. Introduction p. 119
- 2. Synthèse des résultats obtenus p. 120
 - 2.1. Comparaison des taux de compactage p. 120
 - 2.2. Comparaison des temps de compactage/décompactage p. 122
- 3. Étude des gains réels respectifs des différentes méthodes dans le cas d'un transfert de données p. 126
- 4. Conclusion p. 133

CHAPITRE 5. CONCLUSION

- 1. Introduction p. 137
- 2. Classifications possibles p. 138
 - 2.1. Classification selon le caractère de généralité p. 138
 - 2.2. Classification suivant le mode d'élaboration p. 140
- 3. Méthode de différenciation des classes de données p. 142
 - 3.1. Introduction p. 142
 - 3.2. Différenciation des langages par application d'une méthode statistique p. 143
 - 3.3. Conclusion p. 145
- 4. Elaboration de références communes pour les méthodes basées sur une étude statistique préalable p. 146
- 5. Vers un système évolutif "intelligent" p. 147
- 6. Exemples d'application p. 148
 - 6.1. Introduction p. 148
 - 6.2. Implémentation d'un compactage dans un réseau d'ordinateurs p. 149
 - 6.3. Application du compactage pour la réalisation d'un stockage géant de données p. 154

7. Conclusion	p. 155
7.1. Résumé des résultats	p. 155
7.2. Perspectives d'avenir	p. 156

ANNEXES

Méthode de Chopping	p. 157
Méthode Hasp	p. 159
Code Shannon-Fano	p. 162
Méthode de combinaisons numériques (méthode de Hahn)	p. 168
Méthode de Louis-Gavet	p. 174
Coefficients de corrélation entre échantillons de différentes classes de données	p. 177
Code interne et modèle de nouveau rangement de colonnes d'une image de carte	p. 178
Dictionnaires (méthode de Hahn)	p. 182
Longueur moyenne du mot-code selon le type de fichier	p. 183
Résultats comparatifs	p. 184

BIBLIOGRAPHIE

p. 186