



Maxime Crochemore
Christophe Hancart
Thierry Lécroq

Algorithmique du texte

Vuibert Informatique





Algorithmique du texte – Maxime Crochemore, Christophe Hancart et Thierry Lecroq

Conception de la couverture : Jean Widmer

Les programmes et exemples figurant dans cet ouvrage ont pour but d'illustrer les sujets traités. Il n'est donné aucune garantie quant à leur utilisation dans le cadre d'une activité professionnelle ou commerciale.

Contact : `informatique@vuibert.fr`
Web : `www.vuibert.fr`

© Vuibert, Paris, 2001
ISBN 2-7117-8628-5

Toute représentation ou reproduction intégrale ou partielle, faite sans le consentement de l'auteur, etc.



Table des matières

Préface VII

1 Outils 1

- 1.1 Mots et automates 2
- 1.2 Un peu de combinatoire 8
- 1.3 Algorithmes et complexité 17
- 1.4 Mise en mémoire d'automates 21
- 1.5 Techniques de base 26
- 1.6 Base de techniques élaborées 38

2 Automates de localisation 53

- 2.1 Arbre d'un dictionnaire 54
- 2.2 Localisation de plusieurs mots 55
- 2.3 Implantation avec fonction de suppléance 63
- 2.4 Implantation avec successeur par défaut 69
- 2.5 Localisation d'un mot 79
- 2.6 Localisation d'un mot et fonction de suppléance 81
- 2.7 Localisation d'un mot et successeur par défaut 88

3 Localisation avec fenêtre glissante 97

- 3.1 Localisation sans mémoire 98
- 3.2 Temps de recherche 103
- 3.3 Calcul de la table du bon suffixe 108
- 3.4 Automate du meilleur facteur 111
- 3.5 Localisation avec une mémoire 115
- 3.6 Localisation avec plusieurs mémoires 121
- 3.7 Localisation d'un dictionnaire 130

4 Table des suffixes 139

- 4.1 Recherche dans une table de mots 140
- 4.2 Recherche avec les préfixes communs 143
- 4.3 Préparation de la liste 148

4.4	Tri des suffixes	149
4.5	Préfixes communs des suffixes	155
5	Structures pour index	161
5.1	Arbre des suffixes	162
5.2	Arbre compact des suffixes	167
5.3	Contextes des facteurs	176
5.4	Automate des suffixes	182
5.5	Automate compact des suffixes	193
6	Index	201
6.1	Implantation d'un index	201
6.2	Opérations de base	204
6.3	Transducteur des positions	209
6.4	Répétitions	211
6.5	Mots interdits	212
6.6	Machine de recherche	215
6.7	Recherche d'un conjugué	220
7	Alignements	223
7.1	Comparaison de mots	224
7.2	Alignement optimal	231
7.3	Plus long sous-mot commun	242
7.4	Alignement avec brèches	252
7.5	Meilleur alignement local	255
7.6	Heuristique pour l'alignement local	258
8	Motifs approchés	265
8.1	Recherche de mots à jokers	266
8.2	Recherche avec différences	270
8.3	Recherche avec inégalités	282
8.4	Recherche de motifs courts	291
8.5	Heuristique pour la recherche avec différences	300
9	Périodes locales	307
9.1	Partitionnement des facteurs	307
9.2	Localisation des puissances	316
9.3	Recherche de carrés	320
9.4	Tri des suffixes	328
	Bibliographie	337
	Index	341