Préface

Cet ouvrage présente un large panorama des méthodes algorithmiques utilisées pour traiter le texte. À ce titre il s'agit d'un livre d'algorithmique, mais dont l'objet est focalisé sur la manipulation du texte par ordinateur. L'idée de cette publication résulte du constat qu'il n'existe aucune autre référence en français abordant le sujet avec cette ampleur (et les rares livres en anglais consacrés au sujet sont essentiellement des monographies de recherche). Ce constat est d'autant plus surprenant que les problèmes du domaine sont connus depuis le développement des systèmes d'exploitation évolués et que le recours à des solutions efficaces devient crucial depuis l'utilisation massive de l'informatique et de la bureautique dans de nombreux secteurs de la société.

Sous une forme écrite ou vocale, le texte est le seul véhicule fiable des concepts abstraits. Il reste donc le support privilégié des systèmes d'informations, en dépit d'efforts importants vers l'utilisation d'autres médias (interfaces graphiques, systèmes de réalité virtuelle, films de synthèse, etc.). Cet aspect est encore renforcé par les bases de connaissances littéraires, juridiques, commerciales ou autres qui se développent sur l'Internet grâce en particulier aux services de la Toile.

Le contenu de l'ouvrage porte sur les éléments formels et les bases techniques utilisés dans les domaines de la recherche documentaire, de l'indexation automatique pour les moteurs de recherche, et des logiciels systèmes, ce qui inclut l'édition, le traitement et la compression de textes. Les méthodes qui sont décrites s'appliquent au traitement de la langue naturelle, au traitement et à l'analyse des séquences génétiques, à l'analyse de séquences musicales, aux questions de sécurité liées aux flux de données, et à la gestion des bases de données textuelles, pour ne citer que quelques applications immédiates.

Les sujets retenus comprennent la localisation de motifs textuels, l'indexation de données textuelles, la comparaison de textes par alignement et la recherche de régularités locales. En plus de leur intérêt pratique, ces sujets possèdent des aspects théoriques et combinatoires qui fournissent d'étonnants exemples de solutions algorithmiques.

Le but de l'ouvrage est avant tout pédagogique. Il s'adresse en premier lieu aux étudiants des classes préparatoires aux grandes écoles, aux étudiants des seconds et troisièmes cycles universitaires d'informatique, ainsi qu'aux élèves-ingénieurs en informatique. Mais il peut aussi être utilisé par des ingénieurs de production de logiciels.

Nous remercions vivement les chercheurs qui ont pris le temps de relire et de commenter les ébauches préalables de ce livre. Il s'agit de Saïd Abdeddaïm, Marie-Pierre Béal, Christian Charras, Sabine Mercier, Laurent Mouchard, Johann Pelfrêne, Bruno Petazzoni, Mathieu Raffinot, Giuseppina Rindone, Marie-France Sagot. Les erreurs qui restent sont les nôtres. Le lecteur avisé notera l'utilisation des règles de l'orthographe recommandées par le Conseil supérieur de la langue française (*Journal officiel* du 6 décembre 1990) qui nous fait par exemple écrire « apparaitre » et « ambigüité » sous ces formes.

Enfin, des éléments complémentaires au contenu de l'ouvrage sont accessibles sur le site `http://chl.univ-mlv.fr/` ou depuis les pages des auteurs.

Maxime Crochemore
Christophe Hancart
Thierry Lecroq
Marne-la-Vallée et Rouen,
juin 2001