

Préface

VOICI un nouveau livre d'introduction à la théorie des langages, de la calculabilité et de la complexité. Un de plus ? non, car il s'inscrit dans l'évolution et l'enrichissement continu de ce sujet et qu'il témoigne de sa vitalité et de son mûrissement. Loin d'être encore un domaine standardisé, il n'est plus non plus à ses débuts et son exposition s'est enrichie au fur et à mesure de l'expérience acquise en l'enseignant à des publics variés.

Il n'existe pas pour l'instant de présentation standardisée de cette théorie, au sens où le livre de Feller a pu en constituer une pour les probabilités. La tentative d'Eilenberg d'en constituer une s'est arrêtée en chemin puisque les deux premiers volumes (automates puis semigroupes) ne furent jamais suivis des deux autres prévus initialement (langages algébriques puis calculabilité).

Ce livre représente un choix pragmatique fondé sur un enseignement réalisé en interaction avec des étudiants. C'est un choix original de sujets et il dévoile au fil des pages les prolongements des éléments de base vers des sujets plus spécialisés. En ce sens, il constitue une introduction stimulante, qui donne envie d'aller plus loin aux débutants et réserve des surprises aux spécialistes du domaine.

Il plaira spécialement aux étudiants ayant le goût des mathématiques discrètes, mais aussi à ceux qui aiment les algorithmes. On y trouvera ainsi aussi bien un résultat peu connu de Guibas et Odlyzko sur les polynômes de corrélation qu'un programme en C s'écrivant lui-même comme illustration du théorème de récursion.

La présentation est exceptionnellement claire et les preuves sont données avec grand soin. Signe des temps de recherche de qualité, les exercices sont accompagnés de solutions qui sont la seule garantie que l'exercice est faisable (on se souvient d'un manuel proposant en exercice de résoudre le problème de correspondance de Post sur un alphabet de moins de quatre lettres...).

Bon voyage donc au lecteur qui aborde ce sujet et auquel je souhaite autant de plaisir que j'en ai éprouvé à la lecture des *Notions sur les grammaires formelles* de Gross et Lentin qui m'a permis de le découvrir moi-même – je crois bien que c'était en 1968.

Dominique Perrin

Introduction

S'il n'y pas de solution, c'est qu'il n'y a pas de problème

Devise Shadoks

CE LIVRE est une introduction à l'informatique théorique qui couvre aussi bien les langages formels que la calculabilité et la complexité. Il existe d'excellents ouvrages en anglais couvrant ces différents sujets mais très peu en français. Celui-ci tente humblement de combler cette lacune.

L'objectif est d'appréhender les limites des ordinateurs et de comprendre ce qu'il est possible ou impossible de faire avec eux. Cette question est délicate mais fondamentale puisqu'elle touche à l'essence de la notion de calculable. Elle est considérée tout au long de cet ouvrage à travers différents modèles de calculs : automates, automates à pile et machines de Turing.

Les questions de calculabilité et de complexité ne sont abordées que dans la seconde partie de l'ouvrage. La première partie est consacrée à la théorie des automates qui est un préambule indispensable. La théorie des automates s'est développée dès la naissance de l'informatique théorique. Elle en est maintenant le socle sur lequel s'appuie le reste. Bien que les automates soient un modèle très simple, ils jouissent de remarquables propriétés et leur étude est encore très active aujourd'hui. La théorie des automates a des liens étroits avec l'arithmétique, la combinatoire, l'algèbre, la topologie, la logique, la théorie des jeux et l'algorithmique. Les automates sont aussi au cœur de nombreuses applications en informatique du génome, en traitement des langues naturelles et en arithmétique des ordinateurs. La théorie de la complexité s'est épanouie plus récemment que la théorie des automates mais c'est maintenant un domaine très dynamique. La seconde partie de cet ouvrage couvre les thématiques centrales de ce vaste sujet. Quelques thématiques plus spécialisées comme les approximations et les algorithmes randomisés ne sont pas abordées.

La première partie est consacrée à la théorie des langages formels. Des éléments de combinatoire des mots et des ordres font office d'introduction. Ces sujets sont trop rarement abordés dans ce type d'ouvrages. Les automates finis et les langages rationnels forment la trame de ce chapitre. Le théorème de Kleene, les automates déterministes et la minimisation sont incontournables et occupent une place importante. Viennent ensuite des prolongements dans des sujets moins classiques. La théorie algébrique des langages et en particulier l'étude des langages sans étoile terminent le premier chapitre. Le second chapitre est consacré aux grammaires et aux langages algébriques. Les liens

avec les systèmes d'équations y sont particulièrement développés pour obtenir le théorème de Parikh. Après les propriétés principales de clôture des langages algébriques, les automates à pile sont introduits et étudiés. Cette étude passe par l'équivalence avec les grammaires et les automates déterministes. Des compléments sur les contenus de pile et le groupe libre concluent cette première partie.

La seconde partie est consacrée à la calculabilité et à la complexité. Ces deux notions fondamentales sont introduites à travers les machines de Turing mais les fonctions récursives sont également abordées. L'équivalence entre ces deux définitions de la notion de calculable est démontrée. Le problème de Post et son indécidabilité permettent ensuite d'exhiber des questions plus naturelles, sur les grammaires par exemple, auxquelles les réponses ne sont pas calculables. Un détour par les machines linéairement bornées conduit au théorème d'Immerman et Szelepcsényi. La NP-complétude constitue le cœur du second chapitre mais la complexité en espace est aussi largement développée. L'espace logarithmique et l'espace polynomial sont l'un et l'autre étudiés en détail à travers la notion de complétude. Ce chapitre se termine par une introduction aux machines alternantes et par la preuve du théorème d'Hartmanis sur les machines à une bande en temps $O(n \log n)$.

Ce livre s'adresse aux étudiants en master d'informatique ou de mathématiques. Tout master d'informatique comporte un cours sur les automates finis, la compilation et les automates à pile. La calculabilité et la complexité sont indispensables à tout étudiant désireux d'avoir une base solide en informatique fondamentale. L'approche rigoureuse conviendra parfaitement aux étudiants en mathématiques ayant une inclination vers l'informatique. Cet ouvrage couvre une grande part du programme de l'option informatique de l'agrégation de mathématiques. Il sera sans nul doute propice à de nombreux développements pour les leçons d'oral. Des exercices corrigés permettent une bonne assimilation.

Aucune connaissance préalable n'est requise. Il est seulement supposé que le lecteur maîtrise quelques éléments de mathématiques enseignés en première année universitaire. Le premier objectif de cet ouvrage est d'introduire et d'expliquer les premières notions d'informatique fondamentale. De nombreux exemples illustrent les définitions. Les différentes notions sont mises en perspective et des connexions entre des concepts en apparence éloignés sont établies. Le second objectif est de permettre d'aller plus loin. Chacun des chapitres comprend des compléments. Ces compléments ont souvent été choisis parce qu'ils permettent d'aborder des résultats élégants sans toutefois nécessiter la mise en place d'une machinerie lourde. Le prix à payer est la technicité de quelques passages. Les choix ont aussi été guidés par des préférences personnelles assumées. Ces compléments sont l'occasion de montrer quelques joyaux parmi lesquels, le théorème de Schützenberger, le théorème de Parikh et le théorème d'Immerman et Szelepcsényi.

Internet

Une page WEB contenant une liste d'errata et d'autres informations se trouve à l'adresse suivante :

<http://www.liafa.jussieu.fr/~carton/Lfcc/>.