

Vous rendrez au plus tard le jour de l'examen une archive compressée contenant :

- un rapport au format **PDF** comprenant une description des structures de données et des fonctions utilisées et un commentaire détaillé de vos résultats ;
  - les sources de vos programmes contenant impérativement un **makefile** ;
  - un fichier **README.txt** contenant une aide pour utiliser vos programmes ;
- et ne contenant ABSOLUMENT pas :
- les exécutables ;
  - les fichiers générés.

## Travail demandé

1. Implanter l'algorithme d'Aho-Corasick pour compter le nombre d'occurrences exactes d'un ensemble de  $k$  mots dans un texte.

Utiliser trois méthodes pour représenter l'arbre :

- une matrice de transitions ;
- un tableau de listes d'adjacence ;
- une table de transitions pour la racine et un tableau de listes d'adjacence pour les autres nœuds de l'arbre.

Vos exécutables doivent impérativement être nommés **ac-matrice**, **ac-liste**, **ac-mixte** respectivement et prendre deux paramètres, d'abord le nom du fichier qui contient les mots à rechercher (un par ligne) puis le nom du fichier qui contient le texte. Ainsi les exécutions de :

- **ac-matrice mots texte**
- **ac-liste mots texte**
- **ac-mixte mots texte**

devront afficher 78 (les fichiers **mots** et **texte** sont les fichiers accessibles aux adresses respectives [https://sympa.univ-rouen.fr/sympa/d\\_read/m1info/Algorithmique%20du%20texte/mots](https://sympa.univ-rouen.fr/sympa/d_read/m1info/Algorithmique%20du%20texte/mots) et [https://sympa.univ-rouen.fr/sympa/d\\_read/m1info/Algorithmique%20du%20texte/texte](https://sympa.univ-rouen.fr/sympa/d_read/m1info/Algorithmique%20du%20texte/texte)).

2. Écrire un générateur pseudo-aléatoire de textes. Votre exécutable doit impérativement être nommé **genere-texte** et prendre uniquement deux paramètres d'abord la longueur du texte à générer puis la taille de l'alphabet. Il doit écrire ses résultats sur la sortie standard.
3. Écrire un générateur pseudo-aléatoire de mots. Votre exécutable doit impérativement être nommé **genere-mots** et prendre uniquement quatre paramètres d'abord le nombre de mots à générer puis la longueur minimale et la longueur maximale des mots et enfin la taille de l'alphabet. Il doit écrire ses résultats sur la sortie standard.

4. Utiliser **genere-texte** pour générer pseudo-aléatoirement des textes de longueur 5 000 000 sur des alphabets de taille 2, 4, 20 et 70.
5. Pour chacun des alphabets utiliser **genere-mots** pour générer pseudo-aléatoirement 3 ensembles de 100 mots de longueur entre 5 et 15, entre 15 et 30 et entre 30 et 60 respectivement.
6. Pour chacun des textes, utiliser **ac-matrice**, **ac-liste** et **ac-mixte** pour effectuer la recherche des 3 ensembles dans les textes avec les 3 implantations de l'algorithme d'Aho-Corasick.
7. Relever les temps d'exécution de chacune de ces recherches, faire des courbes et commenter.