

L3 Informatique, UE DGINF363 (ALGO6, Algorithmique et modélisation)

Vendredi 10 février : utilisation de l'aléatoire

Ce travail est à faire en binôme. Le compte rendu (un par binôme, indiquant vos **noms et groupes de TD**), est à déposer avant vendredi 10 février à 17h sur la page Moodle du cours.

Objectifs

Générer automatiquement un compte rendu d'apnée.

Ressources fournies

- [EssaiPipo.java](#), programme principal.
- [Pipo.java](#), implémentations du générateur (cf. *TD2 d'algo 6*).
- [CRexemple.txt](#), un fichier texte contenant du texte permettant de "calibrer" le générateur aléatoire de texte (i.e calculer les probabilités de transition d'un mot à un autre).

Exercice 1. Compréhension du programme fourni

Le programme Java [EssaiPipo.java](#) prend deux arguments en entrée :

- 1) le nom d'un fichier (format texte) contenant du texte permettant de calculer les probabilités du modèle.
- 2) le nombre de mots à générer aléatoirement.

Lisez `EssaiPipo.java` et `Pipo.java`.

Exercice 2. Compléter le fichier `Pipo.java`

- 1) Écrire la méthode qui met à jour la structure de donnée permettant de compter le nombre de fois qu'un mot est suivi d'un autre.
- 2) Écrire la méthode qui génère des séquences de mots

conformément aux probabilités du modèle.

Exercice 3. Validation du comportement de votre programme

Pour vérifier que votre programme génère bien des mots d'une manière qui est «conforme» au modèle créé dans la phase d'apprentissage il serait nécessaire de réaliser un/des tests statistiques.

On se propose ici de vérifier «simplement» que les fréquences observées pour les mots de plus hautes fréquences sont conformes au modèle (i.e. à la fréquence observée dans le texte d'apprentissage). La fréquence f d'un mot dans un texte est son nombre d'occurrence n dans le texte divisé par le nombre de mots total nt du texte : $f=n/nt$

Pour cela construire le tableau suivant où vous reporterez les fréquences observées dans chacun des textes :

	Texte d'apprentissage	Texte généré 100 mots	Texte généré 1 kmots	Texte généré 100 kmots	Texte généré 1Mmots
Mot le plus fréquent					
2 ^{ième} mot le plus fréquent					
3 ^{ième} mot le plus fréquent					
...					
10 ^{ième} mot le plus fréquent					

Compte-rendu

Le compte-rendu doit être déposé au format PDF. Il devra comporter, sur 2 pages maximum :

- du texte généré par le programme de l'exercice 2;
- le tableau de l'exercice 3;
- une éventuelle conclusion.

Le fichier Pipo.java complété doit être joint à ce compte-rendu.