

Langage Perl : Exercices de révision - TD2

DESS TEXTE

Exercice 1 : Entrées / Sorties

Ecrire le programme *inout.pl* recevant en paramètre un entier N, puis demandant N valeurs afin de calculer leur moyenne, et de l'afficher à l'écran.

Exercice 2 : Manipulation de fichiers

Ecrire le programme *compt.pl* comptant les lignes non-vides d'un fichier texte.

Exercice 3 : Tri de données

Ecrire le programme *tri.pl* recevant une série de 10 entiers, les triant par ordre décroissant au moyen d'un algorithme que vous décrirez, puis réalisant l'écriture des entiers ainsi classés dans un fichier.

Exercice 4 : Expressions régulières

Ecrire le programme *accents.pl*, chargé d'extraire les accents écrits au format HTML d'un corpus non-étiqueté, et de les transformer en leur version "normale". Le fichier contenant le corpus et celui contenant le résultat seront passés en paramètres.

Exercice 5 : Tableaux associatifs

Ecrire le programme *tion.pl* qui extrait les mots se terminant en "tion", les classe par ordre alphabétique, compte leur nombre d'occurrences, puis les affiche par ordre de fréquence décroissante.

Exercice 6 : Substitutions

Ecrire le programme *minmaj.pl* qui remplace tous les caractères majuscules d'un fichier par leur équivalent minuscule.

Problème : réalisation d'un étiquetteur

On veut réaliser un étiquetteur de texte. Pour cela, on dispose des ressources suivantes :

- un corpus non-étiqueté : fichier *corp_nontagge.txt*,
- un dictionnaire au format "mot - étiquette" : fichier *corp_tala.ascii*.

Les tâches à accomplir sont les suivantes :

1. analyser le dictionnaire → sortir le fichier des étiquettes,
2. écrire un programme qui "lit" le dictionnaire, puis à chaque mot saisi, renvoie l'étiquette correspondante,
3. créer un lexique unique, écrire un dictionnaire qui associe un mot à chaque étiquette,
4. réfléchir aux moyens de déterminer les étiquettes des mots inconnus,
5. en proposer une implémentation en Perl.