

Master IC – Fouille de Données

TP Fouille de Textes - Exercices Perl

05 novembre 2008

Exercice 1

Construisez un tableau propre aux jours de la semaine. Écrire un script Perl qui affiche le contenu de chaque case du tableau.

Exercice 2

Écrire une fonction qui test si un nombre est premier.

Exercice 3

Écrire une procédure qui test si une chaîne de caractères correspond à une heure valide (de la forme hh.mm.ss).

Exercice 4

Considérons un fichier *fic.txt* comportant une phrase par ligne.

Écrire une fonction *est_bien_formee* qui test si une chaîne de caractères commence bien par une majuscule et se termine bien par un point (phrase correctement formée). Écrire une procédure appelant *est_bien_formee* qui affiche le nombre de lignes bien formées du fichier *fic.txt*.

Implanter une procédure qui écrit dans un fichier *fic_sortie.txt* les phrases correctement formées du fichier *fic.txt* en plaçant un espace devant chaque virgule et chaque point (prétraitement nécessaire pour l'étiqueteur de Brill).

Exercice 5

Généraliser les exercices précédents en prenant en entrée un texte quelconque d'un fichier *in.txt*. Ce texte doit être écrit dans un fichier *out.txt* au format adapté pour l'étiqueteur de Brill (toutes les ponctuations et caractères spéciaux sont séparés par un espace avec une phrase par ligne).

Attention !! Pensez à gérer les situations spécifiques pour la segmentation par ligne, par exemple la phrase :

Plusieurs candidats se sont déclarés pour la campagne présidentielle, en particulier Mme Royal et M. Sarkozy qui sont respectivement issus du P.S. et de l'U.M.P.