

TP3: Échéance : 22 Avril 2010 (à faire en équipes de 2 ou 3 personnes)

L'objectif de ce travail est d'écrire un programme Perl permettant de calculer les différentes statistiques sur les mots d'un texte. Le programme lira à l'entrée un fichier texte dont le nom sera *Input.txt* (assurez-vous de sauvegarder le texte dans le format .txt si vous travaillez avec le MS Word). Pour un texte quelconque (sans lettres accentuées) votre programme doit inscrire dans le fichier *Output.txt* les statistiques suivantes :

1. Le nombre de phrases dans ce texte. Une phrases se termine par les caractères « . », « ! » ou « ? ».
2. Le nombre de caractères (espaces non compris) dans ce texte.
3. Le nombre de lettres majuscules dans ce texte.
4. Le nombre de mots dans ce texte. Un mot est une suite de caractères comprenant au moins une lettre et se terminant soit par un caractère d'espacement (i. e., un espace, une tabulation ou une fin de ligne), soit par les caractères « . », « ! » ou « ? ».
5. Le nombre de mots qui se terminent par « er ».
6. Le nombre de mots contenant le motif « th ».
7. Le nombre de chiffres dans ce texte.
8. Le nombre d'occurrences des mots (sans distinguer les minuscules et les majuscules dans un mot) et les numéros de leurs phrases d'apparition sous le format suivant :
mot : nombre d'occurrences -> ligne №1, ligne №2, etc.

À remettre:

Le fichier unique TP.pl (contenant le code source du programme) doit être envoyé à mon adresse de courriel le 22 avril 2010 au plus tard (une copie papier n'est pas nécessaire). Votre code doit être bien commenté.

Un exemple :

Soit le texte suivant :

The Montreal Canadiens made doubly sure they didn't squander a late two-goal lead for a second night in a row. Brian Gionta scored twice -- including Montreal's second empty-net goal -- and Jaroslav Halak came within 3:36 of recording a shutout, leading the Canadiens to a 4-1 victory over the Florida Panthers on Thursday night. Halak, who made 23 saves, lost his bid for a shutout when Radek Dvorak scored 16:24 into the third to draw Florida to 2-1.

Les résultats obtenus pour le texte ci-dessus:

1. Le nombre de phrases dans ce texte est : 3.
2. Le nombre de caractères (espaces non compris) : 374.
3. Le nombre de lettres majuscules dans ce texte est : 15.
4. Le nombre de mots dans ce texte est : 72.
5. Le nombre de mots qui se terminent par « er » dans ce texte est : 2.
6. Le nombre de mots contenant le motif « th » est: 9.
7. Le nombre de chiffres dans ce texte est : 13.
8. Le nombre d'occurrences des mots :

The : 4 -> 1, 2, 2, 3
Montreal : 1 -> 1
Canadiens : 2 -> 1, 2
made : 2 -> 1, 3
doubly : 1 -> 1
sure : 1 -> 1
they : 1 -> 1
didn't : : 1 -> 1
.....