

Projet ScIn1B 2018

Projet à réaliser en binôme.

Vous devez rendre un rapport (max 10 pages) contenant les éléments suivants

- Explications de votre code,
- Images des résultats obtenus
- Manuel utilisateur indiquant les commandes à utiliser pour exécuter vos programmes

Rapport papier à rendre le lundi 17/12 lors de la démonstration de projet

Vous devez également rendre le code de votre projet (par mail ou clé usb) lors de la démonstration.

La démonstration aura lieu le lundi 17/12 et durera 10 minutes par binôme. Les salles et créneaux horaires vous seront communiqués ultérieurement. Vous devez être présents 10 minutes avant le début de votre passage pour vous installer.

Données

Les données sont disponibles dans l'archive donnees_projet.zip sur Plubel. Les fichiers de données, `page1.html` à `page23.html`, sont les codes HTML de pages web.

Sujet :

1) Vous devez écrire un programme Perl qui recherche dans une page source HTML un mot-clé avec :

- un espace ou une fin de balise `>` avant le mot-clé,
- un `s` optionnel après le mot-clé
- un signe de ponctuation (espace, virgule, point-virgule, point, deux-points) ou un début de balise `<` après le mot-clé,
- en minuscules ou en majuscules (utilisez l'option `i` des expressions régulières en Perl)

Ce programme `exo1.pl` reçoit en paramètre le nom du fichier de données et le mot-clé cherché. Votre programme doit afficher dans le terminal le nom du fichier de données ainsi que le nombre de lignes contenant le mot-clé.

Par exemple : `./exo1.pl page20.html egypt` donne comme résultat
`page20.html 2`

2) A partir du programme précédent, vous devez écrire un programme Perl `exo2.pl` qui recherche dans une page HTML un mot-clé (avec les critères de la question précédente) et attribue un score. Ce score est composé :

- du nombre de lignes contenant le mot-clé (avec les critères de la question 1)
- auquel on ajoute 3 points à chaque fois que le mot-clé (en minuscule ou majuscule) se trouve dans un titre
- et auquel on ajoute 2 points chaque fois que le mot-clé (en minuscule ou majuscule) se trouve dans un des liens défini par href, par exemple :
`href="http://www.encyclopedia.com/places/africa/egyptian-political-geography/cairo"`
pour le mot-clé `cairo`

Votre programme doit afficher dans le terminal le nom du fichier de données ainsi que le score obtenu.

Par exemple : `./exo2.pl page22.html cairo` donne comme résultat
`page22.html 201`

Indications :

- Pour rechercher dans le titre de la page, il faut chercher : `<title>` et `</title>`
- Pour rechercher dans les liens, il faut chercher : `href="..."`

Remarque : si la ligne, le titre ou le lien contiennent plusieurs fois le mot-clé, on ne le comptera qu'une fois.

3) Vous devez créer un programme `exo3.pl` qui va exécuter le programme `exo2.pl` sur tous les fichiers de données dont le numéro est compris entre deux valeurs `deb` et `fin`. Vous devez passer trois paramètres (`$deb`, `$fin` et `$mot`) à votre programme. Le programme calcule le score pour chaque page web. Le résultat est placé dans un fichier `res_$mot`.

Par exemple : `./exo3.pl 1 23 cairo` donne comme résultat le fichier `res_cairo` comme celui donné dans `donnees_projet.zip`.

4) Exécutez le programme précédent sur toutes les pages web afin d'obtenir :

- les fichiers `res_cairo`, `res_egypt`, `res_recursive` et `res_music`.
- puis les fichiers `res_tiling`, `res_picture`, `res_apple`, `res_wave`

5) Partie gnuplot

a) Vous devez créer un programme `trace.gnuplot` pour créer un graphique en ligne des données de `res_cairo`, `res_egypt`, `res_music` et `res_recursive` :

- titre du graphique : scores des pages web sur différents mots-clés

- axe horizontal :

- titre : page web
- étiquettes : les noms des pages web inclinées à 45°

- axe vertical :

- titre : score
- valeurs entre 0 et 300

- légendes respectives : `cairo`, `egypt`, `music` et `recursive`

Vous devez obtenir un graphique similaire à celui qui vous est fourni (`exo6a.png`) pour les mots `cairo`, `egypt`, `music` et `recursive`.

Tracer ensuite le graphique équivalent avec les mots-clés `tiling`, `picture`, `apple` et `wave`.

b) Vous devez créer un programme `histo.gnuplot` devra tracer sous forme d'histogrammes les scores page par page pour les mots-clés `cairo`, `egypt`, `music` et `recursive`. Vous devez obtenir un graphique similaire à celui qui vous est fourni (`exo6b.png`) pour les mots `cairo`, `egypt`, `music` et `recursive`.

Tracer ensuite le graphique équivalent avec les mots-clés `tiling`, `picture`, `apple` et `wave`.