

# TD noté, 12 décembre 2007

*Le programme construit au fur et à mesure des questions devra être imprimé à la fin du TD et rendu au chargé de TD. Il ne faut pas oublier de mentionner son nom en commentaires au début du programme. Des commentaires insérés dans le programme permettront de retrouver les réponses de chaque question.*

Lorsqu'on se connecte à un site internet, celui-ci enregistre votre adresse IP, l'heure et la date de connexion ainsi que la page ou le fichier désiré. Ainsi le fichier `logpdf.txt`<sup>1</sup> recense ces quatre informations séparées par des espaces pour les fichiers d'extension `pdf` téléchargés sur le site <http://www.xavierdupre.fr/>.

1) La première étape consiste à charger les informations depuis le fichier texte dans une matrice de 4 colonnes (liste de listes, liste de t-uples). On utilisera pour cela les fonctions `open` et les méthodes `split`, `replace` associées aux chaînes de caractères. On pourra s'aider des corrections des TD précédents. (2 points)

2) On souhaite dans un premier temps faire des statistiques sur les dates : on veut compter le nombre de fichiers téléchargés pour chaque date présente dans les données. On pourra pour cela utiliser un dictionnaire dont la clé sera bien choisie. Le résultat devra être inséré dans une fonction prenant comme entrée une matrice (ou liste de listes, liste de t-uples) et retournant un dictionnaire comme résultat. (4 points)

3) On s'intéresse au programme suivant :

```
l = [ (1, "un"), (3, "deux"), (2, "deux"), (1, "hun"), (-1, "moinsun") ]
l.sort(reverse = True)
print l
```

Donc le résultat est :

```
[(3, 'deux'), (2, 'deux'), (1, 'un'), (1, 'hun'), (-1, 'moinsun')]
```

Que s'est-il passé ? (2 points)

4) On désire maintenant connaître les 10 dates pour lesquelles il y a eu le plus de téléchargements ces jours-là. L'inconvénient est que le dictionnaire élaboré à la question 2 ne retourne pas les réponses dans l'ordre souhaité : il faut classer les dates par nombre de téléchargements croissants. Il faut ici imaginer une fonction qui retourne ces dix meilleures dates et des dix fréquentations correspondantes en se servant de la remarque de la question 3. (4 points) A quels événements correspondent les quatre premières dates ? (Cette petite question ne rapporte pas de point.)

5) Effectuez le même travail pour déterminer les dix documents les plus téléchargés. (2 points)

6) Ecrire une fonction qui retourne l'heure sous forme d'entier à partir d'une date définie par une chaîne de caractères au format `"hh : mm : ss"`. Par exemple, pour `"14 : 55 : 34"`, la fonction doit retourner 14 sous forme d'entier. L'instruction `int("14")` convertit une chaîne de caractères en un entier. (2 points)

7) Calculer le nombre de documents téléchargés pour chaque heure de la journée. Le site est-il consulté plutôt le matin ou le soir ? Serait-il possible de conclure aussi rapidement pour un site d'audience internationale ? (4 points)

---

1. Il faut télécharger ce fichier depuis l'adresse <http://www.xavierdupre.fr/enseignement/initiation/logpdf.txt>