

TP encadré n°1 :

“Extraire et transformer des données textes à partir d’un système de fichier”

Instruction : Les exercices portant l’annotation [secondaire] devront être réalisés une fois les autres exercices terminés.

Exercice 1 : `bash`

a) Ecrivez un script `bash` qui vous permette de recopier la hiérarchie de dossier et des fichiers que contient `/home/depot/2A/gmd/tp1/` dans vos dossier personnels.

b) Ecrivez un script `bash` avec la commande `find` qui permette dans cette hiérarchie de trouver uniquement les fichiers des listes d’étudiants de 1a, 2a et 3a en 2010-2011 (c’est à dire dont le nom contient les chaînes "liste", "étudiant" ou "élève", "2010" et "2011").

c) Ecrivez un script qui trouve également ces trois fichiers et en plus leur donne les droits en lecture et exécution à tous les utilisateurs et en écriture, uniquement à vous.

Exercice 2 : `cat` et `grep`

En utilisant les résultats de l’exercice 1 ainsi que les commandes `cat` et `grep`, écrivez un script qui cherche dans les listes des étudiants inscrit à l’ESIAL en 2010-2011 tous ceux qui suivent l’approfondissement IL ou TRS.

Exercice 3 : `awk`

a) Avec le langage de script `awk` faites un script qui écrive dans un fichier les noms et prénoms des étudiants de 3^{ème} année qui étaient à l’étranger en 2010-2011 (le 7^{ème} champ est différents de : vide, "3A_SR" et "MR2").

b) [secondaire] Ecrivez un script `awk` dont la sortie affiche sur la sortie standard la liste des 5 prénoms les plus fréquents parmi étudiants et enseignants de l’ESIAL. La sortie donnera la liste triée selon la fréquence des prénoms et indiquera pour chaque prénom le nombre de fois qu’il est porté.

Conseil : utilisez les tableaux associatifs.

c) [secondaire] Ecrivez un script qui écrit l’ensemble des données disponibles (redoublant ? PVE validée ? groupes ?, etc.) pour les étudiants délégués de groupe

Exercice 4 : `python`

a) Ecrivez les scripts `python` qui vous permettrons de compter quel est le taux de redoublant par approfondissement en 3a.

Aide :

taux de redoublement par approfondissement = $\frac{\text{nombre d'étudiants redoublants de l'approfondissement}}{\text{le nombre d'étudiant de l'approfondissement}}$

b) [secondaire] En sachant que pour être diplômable, un 3a doit avoir validé sont TOIC (colonne 9 de la liste des 3a) et validé sa PVE (colonne 10 des 3a), pouvez vous écrire un script qui dise si les délégués de 3a sont diplômables.

c) [secondaire] Le fichier `section_de_la_cnu` contient l'intitulé des disciplines dans laquelle les enseignants sont qualifiés alors que la liste des enseignants ne fournit que le numéro de cette discipline. Ecrivez un script qui parcourt ces fichiers pour finalement créer une nouvelle liste d'enseignants dont la structure est la suivante :

"nom";"prénom";"discipline de qualification";numéro de leur poste
et qui est triée dans l'ordre alphabétique de la discipline.

Exercice 5 : `sed`

a) En utilisant `sed`, ôtez les espaces qui peuvent exister avant et après les noms/prénoms des étudiants et enseignants.

b) [secondaire] Toujours avec `sed`, écrivez dans un nouveau fichier la liste de tous les étudiants de 2a et 3a avec les données les concernant écrites dans l'ordre suivant : Approfondissement explicité ("ingénierie logicielle" pour IL par exemple), nom, prénom.