

Analyse du comportement des clients

Partie 1

Un site web marchand propose de la documentation thématique pour ses clients (jardinage, informatique, ..). Le fichier de log du site nous permet d'identifier le type de documentation qu'a téléchargé chaque usager. Voici un exemple type d'enregistrement dans un fichier de log d'Apache :

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326

Dans cet exemple : le 10 octobre à 13h55 Franck a téléchargé l'image apache_pb.gif

Après « nettoyage », le contenu utile de ce fichier est joint en annexe (log_clients_themes.txt). Dans ce fichier, chaque ligne représente un téléchargement (ordre d'enregistrement, nom de l'usager, thèmes). Par exemple : « **5 ; Pierre ; Bricolage** » signifie que le 5 eme enregistrement indique que Pierre a consulté une information sur le bricolage.

On vous demande d'écrire un programme en JAVA permettant construire la matrice MUT (usagers, thèmes) dans laquelle chaque élément représente le nombre d'accès de chaque usager à un thème donné. Cette matrice vous permettra de répondre aux questions qui suivent. Le programme que vous écrirez vous sera aussi utile dans les exercices suivants.

Votre programme doit extraire **3 fichiers** texte

- La liste des thèmes où chaque thème apparaît une seule fois.
- La liste des usagers où chaque usager apparaît une seule fois
- Une matrice de chiffre MUT (relations entre usagers et thèmes)

La sortie de la matrice MUT doit être de la forme suivante. Le nombre de lignes, de colonnes et le contenu de chaque cellule est fictif dans cet exemple.

```
10;12;50;70;13;  
20;32;10;40;23;  
30;22;30;40;33;  
10;12;50;70;13;
```

Dans cet exemple le premier usager s'est connecté **10** fois sur le premier thème et **12** fois sur le second thème. Les usagers peuvent être en lignes ou en colonnes, l'important est de pouvoir retrouver la relation qui lie chaque usager au nombre de thèmes qu'il a consulté.