

Rendu individuel n°1

Durant ces 2 semaines, j'ai fait l'analyse de la base de données d'IMDB. Pour cet analyse, j'ai donc fait le diagramme UML de la base de données afin de voir les liens entre les tables. Lors de la première réunion, nous avons discuter du design (Quelles sont les couleurs que nous allons utiliser pour l'ensemble de notre site ?).

Les fichiers comportant les données sont en format tsv, nous avons donc eu l'idée de convertir ces fichier de tsv en csv, c'est-à-dire de mettre un ";" à la place de "\t" pour la séparation des colonnes et de remplacer les "\N" en "null", cette tâche a été attribuée à Bastien. Je l'ai aidé à écrire le script mais malheureusement nous avons finalement remarqué qu'il n'est pas nécessaire de faire cette conversion car la base de données peut directement utiliser ce format.

Suite à l'analyse, j'ai discuté avec mon groupe lors de la réunion, nous avons constaté qu'il y a une possibilité de fusionner quelques fichiers, j'ai donc fait les scripts pour faire ce fusionnement. Après avoir fait le script, j'ai remarqué qu'il y aura beaucoup de valeurs "\N" dans les colonnes et le filtrage des données peut être plus compliqué dû aux nombres de données supplémentaires. De plus il y a un fichier qui fait 2Go, nous avons tenté de réduire sa taille en le divisant en plusieurs tables mais cela compliquera l'accès aux données. Nous avons ainsi décidé de ne pas faire de changement.

ZHAN Pascal