

## Prétraitement du fichier data.csv

Groupe:	John	THOMPSON-WELLS
L3 Informatique	Louis	LE LANN

Opérations effectuées:

Importation du fichier dans une table pré-construite avec des attributs de type text et sans clé primaire, afin de traiter les données depuis postgres et renvoyer les données sur une table en 1FN avec les bons types d'attributs, en ajoutant des attributs manquant pour passer en 2FN et 3FN et une clé primaire correcte.

L'importation des données est effectué en enlevant la première ligne du fichier contenant l'entête des colonnes et en séparant les données sur le caractère "tabulation".

Chaque attribut de chaque enregistrement sera traité de façon à retirer le contenu inutile de celui-ci, par exemple: avoir 2 fois le même nom de ville dans l'attribut ville, n'est pas utile et peut causer des problèmes s'ils ne sont pas orthographiés de la même façon.

Seuls les attributs de description n'ayant pas de données inutiles ne seront pas traités de cette façon.

Pour les attributs "Négatif ou inversible" et "Couleur ou noir et blanc", des domaines seront créés pour avoir des données correctes.

La date n'étant pas complète dans la plupart des enregistrements, sera stockée sous forme de 3 attributs supplémentaire qui remplaceront l'attribut "Date" du fichier, à savoir "Jour", "Mois" et "Annee", le texte précédant la date indiquée ("Prise de vue : ") sera retiré pour pouvoir effectuer des requêtes facilement sur les attributs de la date.

Le canton sera déplacé de l'attribut ville dans un attribut canton.

Deux attributs contenant les coordonnées des villes seront rajoutés.

Les données dans l'attribut "Nombre de clichés" seront analysées, et si elles sont du type : "chiffre , chiffre , .." alors les données des attributs "Négatif ou inversible", "Couleur ou noir et blanc" et "Taille du/des clichés" seront analysées, si et seulement si chacun de ces attributs contient autant de données distinctes, alors elles seront séparées dans de nouveau attributs qui remplaceront ceux du fichier originel, et auront pour nom l'attribut correspondant avec un numéro pour les différencier.

Il a été repéré qu'il n'y avait pas plus de 3 données distinctes par enregistrement, par conséquent on aura au final 3 "Nombre de clichés X", 3 "Négatif ou inversible X", 3 "Couleur ou noir et blanc X" et 3 "Taille du/des clichés X", X étant un numéro allant de 1 à 3 pour bien séparer les données distinctes.

Des attributs qui serviront de clé primaire seront ajoutés pour passer la table en 2FN et finalement en 3FN.