

JDBC – TP – REALISATION D’UN TRAITEMENT DE FICHIER

OBJECTIF

- En entreprise ce sont des opérations qu’on réalise souvent de nuit et qui existent sur la plupart des applications.
- L’objectif de ce TP est de vous apprendre à réaliser ce genre de traitement.

CONTEXTE

Le fichier que je vais vous demander d’intégrer en base de données est issu de la base de données **Open Food Facts** qui est une base de données open source à l’échelle mondiale. Afin de conserver un volume de données raisonnable j’ai extrait **les produits alimentaires fabriqués en France uniquement**.

Le fichier fourni comporte 13 432 références avec 30 informations associées par produit :

- catégorie
- marque
- nom
- score nutritionnel : A (excellent) à F (Faible)
- liste des ingrédients
- énergie pour 100g (en joules)
- quantité de graisse pour 100g
- liste des allergènes
- liste des additifs
- présence d’huile de palme
- toutes les vitamines,
- calcium, magnesium, etc..

Au total chaque référence de produit a 30 attributs.

Compte tenu du volume de données à traiter (environ 5 Mo) il est préférable de le traiter en local ou prendre une base de données PostgreSQL qui est gratuite jusqu’à 100Mo.

Lors du cours Java EE nous développerons une application type moteur de recherche permettant de naviguer dans les données.

TYPE DE FICHIER

Le fichier fourni est un fichier CSV.

Un fichier CSV est un fichier « plat » dans lesquelles les informations sont séparées par des séparateurs. Ici le séparateur utilisé est le | :

- la première ligne fait la liste des attributs séparés par des caractères |
- chaque ligne suivante correspond à un article et les valeurs des attributs sont séparées par des |
 - la liste des ingrédients est séparée par des virgules
 - la liste des additifs et des allergènes également

INSTRUCTIONS

- Créez un nouveau projet nommé **traitement-fichier**
- Le traitement devra utiliser des **requêtes préparées** beaucoup plus performantes pour réaliser des opérations d'insertions/updates en masse.
- Le traitement que vous allez écrire doit a minima :
 - gérer la table des catégories
 - gérer la table des marques
 - gérer la table des produits
 - gérer la table des ingrédients rattachés aux produits. Exemple : si un produit a 10 ingrédients, alors il y aura 10 lignes dans la table des ingrédients pour ce produit.
- Précautions à prendre :
 - il n'est pas question d'avoir des doublons dans les tables des **catégories** et **marques**.
 - Votre programme doit être rejouable, c'est-à-dire qu'il est capable d'intégrer un produit seulement s'il n'existe pas en base.
 - La liste des ingrédients doit être gérée de la même manière.
 - *Le petit plus : l'application gère l'ajout d'un ingrédient ou la suppression d'un ingrédient.*
 - Les ingrédients est évidemment lié à un produit bien particulier
- Réfléchissez aux classes que vous allez mettre en place avant de vous lancer. Vous pouvez y réfléchir en groupe si vous le souhaitez et utilisez le tableau.

COMMENT LIRE UN FICHIER CSV ?

Avec la classe **FileUtils** de la librairie **APACHE commons-io**. Cette classe possède une méthode qui retourne le contenu d'un fichier sous la forme d'une List de String.

1 seule ligne de code pour avoir le contenu d'un fichier !!

A noter que cette méthode convient pour un fichier de petite taille mais pas pour un fichier de plusieurs centaines de Mo qu'il vaut mieux lire ligne par ligne.

COMMENT SPLITTER UNE CHAÎNE DE CARACTÈRES ?

La classe String possède une méthode **split** très pratique pour ça. Une seule ligne de code suffit pour splitter une chaîne de caractères.

COMMITEZ