

MR - Scala

Objectif

Premiers pas en Scala.

Manupuler la scaladoc

Pré-traitement d'un fichier de données sous forme de scripts Scala.

Télécharger et installer Scala (<http://www.scala-lang.org/>)

Vérifier si ce n'est pas déjà installé.

Version actuelle est la 2.11.7

On va essentiellement tester sur le REPL de scala => depuis une console, taper `scala`.

Récupérer le jeu de données sur le site (bdpmFiles.zip)

1. Pré-traitement des fichiers

Ecrire un script permettant de générer les fichiers SQL pour les trois fichiers du zip téléchargé.

Pour cela, vous pouvez utiliser les méthodes de la classe Source (scala.io.Source).

a. Démarche proposée : lire le fichier, récupérer ces lignes et les transformer en une liste. Combien de lignes obtenez-vous sur le fichier CIS_bdpm.txt ?

b. Transformer chacune des lignes de cette liste (avec un map) en tokens (séparateur est un "\t"). A partir des tokens, rédiger les instructions SQL qui vont enrichir vos tables SQL. Quel le type de sortie de votre variable ?

c. Sauvegarder le contenu de cette variable dans un fichier texte. Pour cela, vous allez importer les classes suivantes: scala.collection.JavaConverters._ et import java.nio.file.{Paths, Files}, import java.nio.charset.StandardCharsets

2. Analyser le jeu de données de CIS_bdpm

Récupérer les codes CIS (1^{er} champ) des entrées dont la présentation est 'pommade' (3ème champ du fichier). Combien avez-vous de résultats ?

3. Analyser le fichier CIS_CIP

Comptez le nombre distinct de cis dans le fichier CIS_CIP_bdpm.txt. Pour cela, vous n'avez le droit de n'utiliser que les méthodes map et groupBy de la classe List.

4. Version OOP

Si vous avez le temps, développer une solution OOP (avec class et object) de votre scripts Scala.