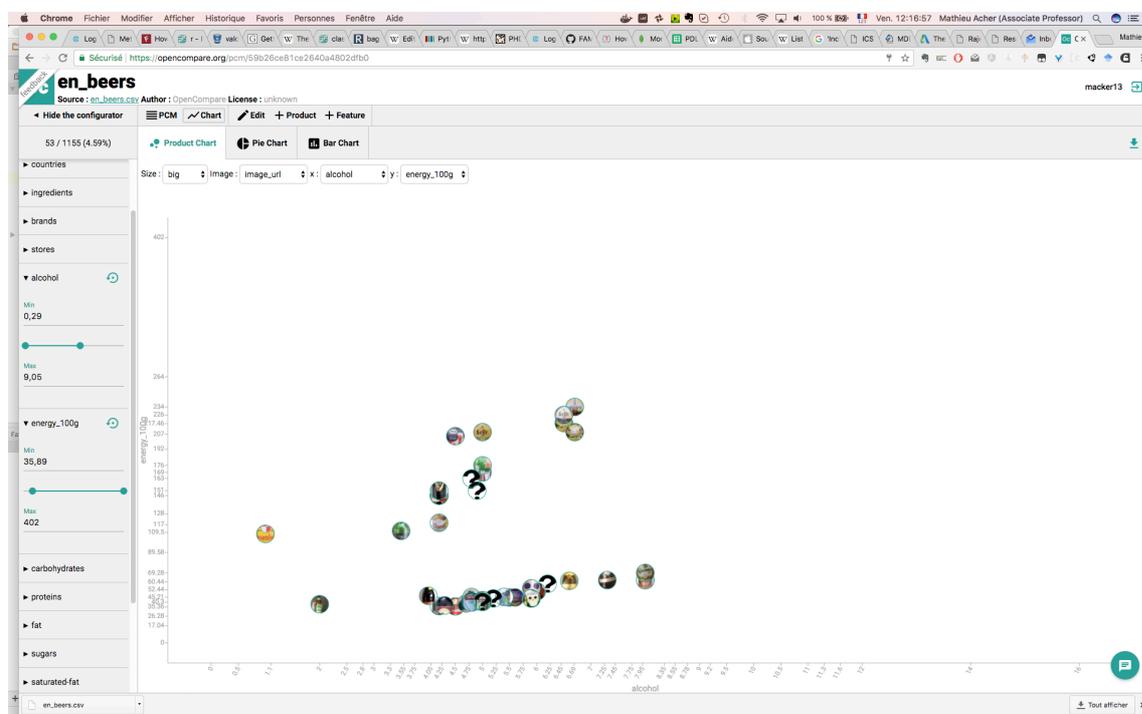


L'objectif de ce projet de PDL est de créer des matrices de comparaisons (au format CSV) à partir d'un dump OFF. Ces matrices seront ensuite visualisées et exploitées sur opencompare qui supporte l'import CSV¹.

Evidemment, il n'y a pas de sens à comparer des produits alimentaires qui n'ont pas de points communs (e.g., comparer des sauces tomates avec du lait). Aussi nous ferons l'hypothèse qu'un utilisateur fournit en entrée de votre procédure quelques critères pour ne garder que les produits qui l'intéressent. Typiquement, on pourra renseigner la catégorie du produit visé (e.g., céréales), voire des sous-catégories, ou n'importe quel filtre sur les données.

Si on prend l'exemple des « bières » on souhaite pouvoir exploiter les données de OFF via CSV et pouvoir jouer avec sur opencompare



¹ Si le format CSV vous apparaît limité (le typage des données est absent par exemple), il sera intéressant d'en discuter.

The screenshot shows a web browser displaying the 'en_beers' application. The interface includes a sidebar with filters for categories like 'countries', 'ingredients', 'brands', 'stores', 'alcohol', 'energy-100g', 'carbohydrates', 'proteins', 'fat', 'sugars', and 'saturated-fat'. The main content area displays a table of beer products with columns for 'id', 'product_name', 'countries', and 'ingredients'.

id	product_name	countries	ingredients
1801828	King of beers	United States	hops, rice, barley malt
7199044	Coors Light	United States	
20460341	Pilsner	Denmark	Water, barley malt, hops extract
50213245	Guinness Draught	France	Contient de forge
54004207	Special export	France, Irlande	eau, malt d'orge, orge torréfié, extrait de houblon, levure
54018013	Lefle Blonde	Belgique	Malt d'orge, gulse et d'autres ingrédients dont de l'eau
90288104	Natur Radler	Deutschland	
200000021470	Bière blonde	France	malt d'orge, malt de blé et blé
200000023758	Premium light lager		
3080210000547	Kronenbourg X6	France	malt d'orge, blé
3080210000707	Bière blonde Kronenbourg	France	Malt d'orge, malt de blé et blé
3080210001476	Kronenbourg	France	
3080210002398	Kronenbourg	France	
3080216003306	Ambre	France	malt d'orge
3080216009094	Bière blonde 1664 (pack de 20 x 25 cl) Kronenbourg	France	contient mg M, O, ORG, DE 9+
3080216024103	Bière blonde	France	Bière blonde
3080216033013	Original bière blonde	France	Contient du malt d'orge et du blé
3080216044309	Bière blanche Fruits rouges	France	Eau, malt d'orge, bière, sucralose, jus de glucose, jus de sureau, bière

En plus d'une solution technologique (en Java) pour produire des matrices de comparaison à partir de OFF, l'objectif de ce projet est aussi d'explorer la pertinence de l'idée : obtient-on des matrices de qualité ? est-ce utile de fabriquer des matrices de comparaison à partir de OFF ?

Vous devez adresser ces deux questions en utilisant votre solution sur plusieurs exemples concrets. Vous devez ainsi démontrer les aspects positifs et la plus value de votre solution mais aussi les limites actuelles, qu'elles soient liées à la qualité des données de OFF, à la difficulté de regrouper des produits similaires, ou à la qualité de votre solution. L'évaluation de votre approche comptera pour 50% de la note : ce n'est donc pas qu'un travail d'implémentation. Par contre, vous avez besoin d'implémenter une solution pour l'évaluer.

Il est tout à fait possible d'évaluer votre solution avec des gens extérieurs au projet.

Comment commencer ?

Etudier OFF. Télécharger un dump OFF, appréhender MongoDB ou CSV et les APIs pour accéder aux données OFF. Prototyper une première transformation « simple ». Itérer.

Tester. Evaluer la pertinence de votre solution. Itérer, tester.

En parallèle de vos expérimentations, écrire le document de spécification.