INRAE



Verrouillage technologique et alimentation durable: le cas des légumineuses

▶ La petite porte du code : SQL



- 2016 ANR LISTIC : militantisme online, 38 M de tweets et posts FB
- 2018 projet européen DARE : radicalisation dans les réseaux sociaux 600 000 tweets et posts FB
- 2019 projet européen LEGVALUE, projet région KING : 300 000 produits agro-alimentaires



> Enjeux environnementaux et sociaux des "pulses"

Pourquoi les légumineuses?:

- Limitation des intrants dans les cultures
- Réduction de la consommation de protéines animales
- Un driver important des transitions agroécologiques?

Problèmes:

- Consommation qui demeure faible malgré un consensus en faveur de l'augmentation des légumineuses dans nos systèmes laimentaires : En Europe 4kg/an/hab.
- Une situation de verrouillage technologique freinant le developpement de produits à base de légumineuses





> Terrain - miné -

- * Traquer dans des bases de données alimentaires(Open Food Facts, USDA, Mintel GNPD) les produits(innovants) contenant des légumineuses
 - a) MAIS peu/pas d'études scientifiques globale approfondie des innovations produits sur les marchés :
 - « (...) there is a lack of information in the scientific literature on type of ingredients used in packaged foods » Ahuja and Al., 2021
 - MAIS peu/pas d'outils pré-existant de traitement et l'analyse fine de ces bases de données alimentaires
- * Donc que faire de ça :

"gluten-free oats, organic coconut sugar, freeze dried banana, pea protein, almonds, chickpea powder, organic chia seeds, organic flax seeds, sea salt, organic quinoa flakes, organic puffed amaranth"

... × 300 000



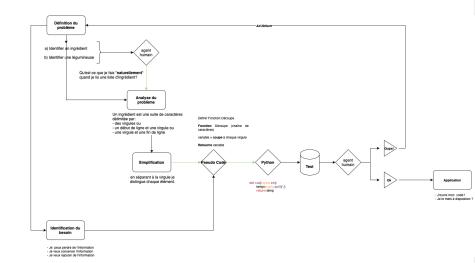
> Objectifs

- → Traduire les listes d'ingrédients dans des structures de données
 - > Passer de listes d'ingrédients à des dictionnaires d'ingrédients
 "tomato* (tomato pulp*, tomato concentrate*), water, rehydrated red kidney
 beans* (11%),... *ingredients from organic farming"
 à :
 {'75_1': {'rawing': 'tomato', 'level': 0, 'comment': ['ingredients from organic
 farming']}, '75_2': {'rawing': 'tomato pulp', 'level': 1, 'comment': ['ingredients
 from organic farming']}, '75_3': {'rawing': 'tomato concentrate', 'level': 1,
 'comment': ['ingredients from organic farming']}, '75_4': {'rawing': 'water',
 'level': 0}, ...}}
- → Pour automatiser des tâches d'analyses :
 - > plongement lexical, pour réduire la synonimie entre ingrédients et les ambiguïtés,
 - > statistiques descriptives, inférentielles
 - > analyse des réseaux d'ingrédients pour déterminer les fréquences d'association...





> Workflow Général







Les "Oups!"

Coder est toujours un processus itératif

Les données ne sont jamais propres

- ▶ 3 grammaires différentes de notation des listes d'ingrédients
- Des variations dans les modalités de notation des adjuvants
- Des variations dans les modalités de notation des proportions
- Des variations dans les modalités de notation des allégations ou des labels



> Organisation du travail

- \rightarrow les fichiers sources à part
- → une règle de nommage fixe (pour_moi_snake_case_20220317.py)
- → une hiérarchie de dossiers fixes :
 - un dossier Active qui ne contient que le fichier de code sur lequel je travailles
 - un dossier History qui contient les versions de fichiers de code dépassées
 - un dossier Raw qui contient une copie des sources de données utilisées pour le code
 - éventuellement un fichier Output pour les sorties (pour stats, visualisations,...)
- → un seul environnement de travail par projet (pour moi Visual Studio Code)
- → utilisation de **chemin relatif** PAS de chemins absolus











> Mes règles de travail

→ Fichiers verbeux... mais explicites

```
Account of the Control of the Contro
```

```
The content of the co
```

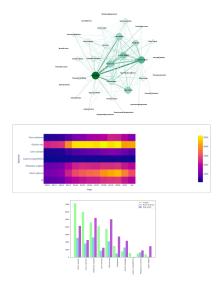
- \rightarrow Si plus de 500/600 lignes de code alors :
 - > refactoriser!
 - > basculer en P.O.O?
 - > Découper?
- → Simplifier, simplifier
- ightarrow Si ça marche sans apprentissage, alors pas d'apprentissage



> So what? - Garfinkel -

Des étapes nécessaires pour :

- → ouvrir les boîtes noires de la production agro-alimentaire
- → la production de résultats affinés,
- ouvrir à d'autres perspectives de recherche : risques sanitaires, pratiques de consommation, habitudes alimentaires, ...
- → ouvrir les données à la communauté des pairs









Merci pour votre attention

Références :

- → GitHub: https://github.com/Pythrix/FOODCOP
- → Pub. HAL: https://cv.archives-ouvertes.fr/tristansalord



