

Extraction de données relatives aux produits
alimentaires à partir de documents non structurés

Pierre MASSÉ

Juin 2020

Résumé

La gestion de l'information produit est devenu un enjeu de société majeur ces dernières années. Les scandales sanitaires récents ont déclenché une prise de conscience collective des consommateurs, en parallèle de réglementations de plus en plus contraignantes pour l'ensemble des acteurs de la filière[1][2].

Table des matières

| | | |
|-----------|---|----------|
| I | Contexte métier | 4 |
| 1 | Description du Groupe | 5 |
| 1.1 | Le métier du Groupe Pomona | 5 |
| 1.2 | Les deux niveaux de décentralisation | 5 |
| 1.3 | Les branches | 5 |
| 1.3.1 | Les branches RHD | 5 |
| 1.3.2 | Les branches spécialistes | 5 |
| 1.3.3 | L'étranger | 5 |
| 2 | La gestion de l'information produit | 6 |
| 2.1 | L'information produit | 6 |
| 2.2 | Le processus associé | 6 |
| 2.3 | Le PIM (Product Information Management) | 6 |
| II | Les données disponibles | 7 |
| 1 | Le périmètre produit | 8 |
| 2 | Les données utilisables | 9 |
| 2.1 | Données structurées | 9 |
| 2.2 | Données non structurées | 9 |
| 2.3 | Pièces jointes | 9 |
| 2.3.1 | Fiches techniques fournisseur | 9 |

| | | |
|------------|---|-----------|
| 2.3.2 | Étiquettes produit | 9 |
| 2.3.3 | Fiches logistiques fournisseur | 9 |
| 2.3.4 | Fiches techniques et argumentaires Pomona | 9 |
| 2.4 | Un mot sur la qualité des données | 9 |
| 2.5 | Les données « manuellement étiquetées » | 9 |
| III | Les objectifs de ce projet | 10 |
| 1 | Les cas d’usage | 11 |
| 1.1 | Objectifs : Qualité et productivité | 11 |
| 1.2 | La préalimentation d’information | 11 |
| 1.3 | Le contrôle à la saisie fournisseur | 11 |
| 1.4 | L’aide aux vérifications Pomona | 11 |
| 1.5 | Les contrôles en masse asynchrones | 11 |
| 2 | Le type de données à récupérer | 12 |
| 2.1 | La composition produit | 12 |
| 2.2 | Les données nutritionnelles | 12 |
| 2.3 | Les données logistiques | 12 |
| IV | Figures et tableaux | 13 |
| V | Bibliographie | 16 |
| VI | Exemple de documents fournisseur | 18 |
| A | Fiches techniques | 19 |
| B | Étiquettes produit | 20 |

| | | |
|------------|--|-----------|
| VII | Le code utilisé | 21 |
| A | Extraction de données du PIM | 22 |
| B | Conversion des pièces jointes en textes | 23 |
| C | Identification des listes d'ingrédients | 24 |

Première partie

Contexte métier

Chapitre 1

Description du Groupe

1.1 Le métier du Groupe Pomona

1.2 Les deux niveaux de décentralisation

1.3 Les branches

1.3.1 Les branches RHD

1.3.2 Les branches spécialistes

1.3.3 L'étranger

Chapitre 2

La gestion de l'information produit

2.1 L'information produit

2.2 Le processus associé

2.3 Le PIM (Product Information Management)

Deuxième partie

Les données disponibles

Chapitre 1

Le périmètre produit

Chapitre 2

Les données utilisables

2.1 Données structurées

2.2 Données non structurées

2.3 Pièces jointes

2.3.1 Fiches techniques fournisseur

2.3.2 Étiquettes produit

2.3.3 Fiches logistiques fournisseur

2.3.4 Fiches techniques et argumentaires Pomona

2.4 Un mot sur la qualité des données

2.5 Les données « manuellement étiquetées »

Troisième partie

Les objectifs de ce projet

Chapitre 1

Les cas d'usage

- 1.1 Objectifs : Qualité et productivité
- 1.2 La préalimentation d'information
- 1.3 Le contrôle à la saisie fournisseur
- 1.4 L'aide aux vérifications Pomona
- 1.5 Les contrôles en masse asynchrones

Chapitre 2

Le type de données à récupérer

2.1 La composition produit

2.2 Les données nutritionnelles

2.3 Les données logistiques

Quatrième partie

Figures et tableaux

Liste des tableaux

Table des figures

Cinquième partie

Bibliographie

Bibliographie

- [1] Conseil de l'Union Européenne. Règlement n°1169/2011 dit inco, nov 2011.
https://www.senat.fr/europe/textes_europeens/ue0120.pdf.
- [2] Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes. Étiquetage des denrées alimentaires : nouvelles règles européennes, jan 2015. <https://www.economie.gouv.fr/dgccrf/etiquetage-des-denrees-alimentaires-nouvelles-regles-europeennes>.

Sixième partie

**Exemple de documents
fournisseur**

Annexe A

Fiches techniques

Annexe B

Étiquettes produit

Septième partie

Le code utilisé

Annexe A

Extraction de données du PIM

Annexe B

Conversion des pièces jointes en textes

Annexe C

Identification des listes d'ingrédients