



## TP - SOCIOLOGIE DE L'ALIMENTATION

Formation à destination des URFIST

### 1 Description générale

**Besoin** “Je suis sociologue et je débute une recherche sur l'évolution des motivations des personnes véganes et végétariennes au fil du temps (sociologie de l'alimentation).”

**Objectif** Détection des études sur le véganisme et le végétarisme. Analyse avec Lodex.

**Outils de TDM envisagés par la suite** Lodex et les web services de TDM.

**Contraintes imposées par les outils** La plupart des web services ne traitent qu'une langue à la fois (conserver la langue majoritaire).

### 2 Étapes

#### 2.1 Interrogation de la base Istex

**Étape 1** Se rendre sur Istex Search <https://search.istex.fr/>.

**Étape 2** Rechercher les formes *végane*, *végétarien*, *végétarisme*, *vegan*, *vegetarian* et *vegetarism*.

**Étapes 3** Pour limiter le silence, rechercher les dérivés et variantes morphosyntaxiques (féminin et pluriel).

**Étape 4** Pour limiter le bruit, on peut préciser les champs les plus à même de renvoyer des résultats pertinents. La liste des champs est accessible dans la recherche assistée. L'objectif est de rechercher les mots-clés dans le **titre**, le **résumé** et les **mots-clés d'auteurs**.

**Étape 5** En affinant ma problématique, je découvre que le terme *vegan* a été créé en 1944 par le cofondateur de la Vegan Society, Donald Watson. Je souhaite vérifier si cette revendication a eu un impact social et s'il transparaît dans la littérature scientifique. [Peut-on limiter les dates de publication du corpus à des documents publiés après 1944? Quel champ interroger?](#)

**Étape 6** Je souhaite disposer d'un résumé pour chacun des documents pour m'assurer de la pertinence des articles avant une lecture complète. [Comment m'assurer de la présence de résumés dans les documents?](#)

**Étape 7** [Quelle est la langue majoritaire?](#) Conditionner la requête pour obtenir un corpus monolingue.

#### 2.2 Téléchargement du corpus

**Étape 8** Télécharger les documents au format approprié.







## 2.3 Chargement dans Lodex

**Étape 9** Se rendre sur votre instance Lodex, se connecter et charger les données. Le lien, votre identifiant et votre mot de passe sont disponibles depuis le site contenant les supports <https://urfist-paris-1591.formation.lodex.fr/instance/supports-formation>.

- Aller dans l'interface administrateur en cliquant sur *Voir plus > Admin* en bas à droite.
- Importer le corpus en glissant le fichier .zip téléchargé **sans décompression préalable**.
- Choisir un loader<sup>1</sup>.
- Cliquer sur *Importer les données*. [Comment vérifier que votre importation est complète ?](#)

**Étape 10** Créer votre site.

- Importer le modèle<sup>2</sup> fourni (fichier *URFIST-modele-vegan.tar*) : menu en haut à droite : *> Modèle > + Importer un modèle*.
- Publier le site pour une première visualisation en cliquant sur *Publier* en haut à droite.
- Cliquer sur l'icône en forme d'œil  pour voir le résultat.
- Explorer les différents graphiques à partir de l'onglet  *Graphiques* en bas.
- Consulter quelques ressources grâce à l'onglet  *Recherche*.
- Modifier un champ du modèle : *la date de création*. Aller sur l'interface administrateur en bas à droite *Voir plus > Admin > Affichage (onglet en haut à droite de Données) > Page d'accueil*.
- Cliquer sur le champ à modifier (sur la roue crantée ) et aller modifier le texte dans *Valeur arbitraire* puis sauvegarder.

## 2.4 Premiers pas en TDM

**Étape 11** Extraire les mots-clés des résumés via le web service *Teeft*.


- Aller dans *Données > Enrichissements* et cliquer sur *+ Ajouter*.
- Donner le nom *Mots-clés (WS)*, aller chercher l'url du web service *Teeft* dans le catalogue en cliquant sur le bouton vert à droite du champ *URL du web service*. [Selon votre corpus, quelle version de Teeft faut-il sélectionner ? Choisissez le nombre de termes à extraire.](#)
- Choisir *Résumé* dans la *Colonne de la source*, cliquer sur *Sauvegarder*. Cliquer sur *Lancer*. [Observer la sortie du web service dans le tableau des données. Que constatez-vous ?](#)

**Dans Lodex, avant de faire un graphique, il est nécessaire de déclarer la colonne comme une ressource (c.-à-d. créer un champ qui contient les données utilisées dans le graphique, dans ce cas : le résultat de Teeft).**

1. Un loader est un script d'adaptation du fichier à Lodex. Il dépend du format de fichier fourni en entrée.

2. Pour rappel, un modèle est un fichier .tar qui permet de mettre en forme le site créé avec Lodex.



- d) Aller dans *Affichage > Ressource principale*, cliquer sur *+ Nouveau champ*. Pour paramétrer ce nouveau champ : dans *Étiquette* nommer le champ *Mots-clés*, sélectionner *Colonne(s) existante(s)* et aller chercher la colonne nommée *Mots-clés (WS)* (le contenu de la colonne s'affiche à droite).
- e) Après avoir sélectionné la colonne, cliquer sur *Ajouter une opération* pour appliquer l'opération de transformation *GET*, renseigner *term* dans le *path*. **À quoi sert cette dernière manipulation ?** Dans *Affichage* choisir le format *Texte-Liste de valeurs*. *Confirmer* puis *Sauvegarder*.
- f) Pour créer le graphique, aller dans *Affichage > Graphiques* cliquer sur *+ Nouveau champ*, nommer le graphique en renseignant *Mots-clés les plus spécifiques* dans le champ *Étiquette*. Choisir la routine<sup>3</sup> *distinct-by* puis choisir le champ sur lequel la routine va s'appliquer grâce au menu déroulant (soit le champ *Mots-clés*). Dans *Affichage*, choisir le format *Graphe à bulles* en filtrant les résultats : dans *Paramètres des Données* mettre valeur minimum à afficher à 2 et choisir le tri *Descendant*. Cliquer sur *Confirmer* puis *Sauvegarder*.
- g) Pour créer un filtre avec ces nouvelles valeurs, se rendre dans *Affichage > Recherche et facettes* et cocher la case *Mots-clés* dans les *Facettes*.
- h) Enfin, aller consulter les résultats de votre premier traitement de TDM . **Combien de documents contiennent le mot-clé *meat consumption* ?**

## 2.5 Pour aller plus loin...

**Étape 12** Extraire les entités nommées des résumés pour déterminer les noms de lieux les plus fréquents.

- a) Utiliser le web service *entityTag - Extraction de différentes entités nommées d'un texte en anglais* sur les résumés, le nommer *entityTag* (au besoin s'aider des détails de l'étape 11). **Observer la sortie du web service dans le tableau des données. Que constatez-vous ?**
- b) Pour créer une colonne contenant uniquement les noms de lieux, aller dans *Données > Enrichissements* et cliquer sur *+ Ajouter*. Donner le nom *ENgeographie*, cliquer sur *Mode avancé*<sup>4</sup> et coller le code ci-après, cliquer sur *Sauvegarder*. Cliquer enfin sur *Lancer*. Pour des raisons de temps de calcul (et temps de TP) ce code ne conserve que le premier nom de lieu du document grâce à l'instruction `.slice(0,1)`.

3. Les routines de Lodex sont des scripts, fournis par Lodex ou stockables n'importe où sur le web, qui peuvent effectuer des agrégations, des calculs, des reformatages et renvoyer des données sous une forme utilisable (souvent un format de type graphique).

4. Le mode avancé offre une meilleure flexibilité pour transformer les données. Il s'agit de code en Lodash, une librairie Javascript (cf. tutorial). Il n'est pas toujours nécessaire de comprendre le code pour appliquer des modifications : les transformations les plus usuelles sont disponibles via ce lien.



```
[assign]
```

```
path = value
```

```
value = get("value.entityTag.LOC").slice(0,1)
```

- c) Faire tourner le web service *Associer un terme au vocabulaire Pays et Subdivisions* sur la nouvelle colonne *ENgeographie* en créant de nouveau un enrichissement nommé *Lieux*. Ce web service permet d'obtenir un code qui permet de créer une cartographie.
- d) Aller dans *Affichage > Ressource principale*, cliquer sur *+ Nouveau champ*. Pour paramétrer ce nouveau champ : dans *Étiquette* nommer le champ, sélectionner *Colonne(s) existante(s)* et aller chercher la colonne nommée *Lieux*. Après avoir sélectionné la colonne, cliquer sur *Ajouter une opération* pour appliquer une transformation *GET*, renseigner *cartographyCode* dans le *path*. Cliquer sur *Sauvegarder*.
- e) Pour créer le graphique : aller dans *Affichage > graphiques* cliquer sur *+ Nouveau champ*, nommer le graphique en renseignant *Lieux cités* dans le champ *Étiquette*. Choisir la routine *distinct-by* puis choisir le champ sur lequel la routine va s'appliquer grâce au menu déroulant. Enfin, dans *Affichage*, choisir le format *Graphique* puis *Cartographie* et paramétrer la carte (nombre maximum de champs : 10000 ; choisir une couleur ; afficher l'info-bulle) et sauvegarder. **Quels sont les pays les plus cités ? Pourquoi ?**

Ce site contient la correction du TP : <https://urfist-paris-1591.formation.lodex.fr/instance/correction-tp-vegan>.