

# Exercices-td-5-6

L'objectif de ces exercices est de vous familiariser avec les outils Palladio et RAWGraphs, et de manière générale, de vous familiariser avec la préparation de données à des fins de visualisations.

Vous documenterez chaque étape dans un document Word / Pages / LibreOffice. Pour chaque exercice, vous présenterez les prompts que vous avez employés, ainsi que la réponse obtenues. N'hésitez pas à mettre les différentes versions d'un même prompt, ainsi que les réponses correspondantes

Pour chaque exercice, vous vous assurerez que les résultats sont corrects, et si besoin, vous les corrigerez manuellement. Dans le dossier, vous indiquerez ces résultats incorrects, et présenterez comment vous avez procédé pour les corriger.

Si ChatGPT ne vous permet plus de lire un fichier, n'hésitez pas à copier-coller le contenu du fichier directement dans le prompt. De plus, si ChatGPT ne traite pas tous les documents (par exemple, ne traite que 5 textes sur 10 fournis dans le prompt), n'hésitez pas à réemployer le prompt plusieurs fois, en changeant les textes.

Vous enverrez ce document ainsi que tous les fichiers produits au cours des exercices par mail à [nicolas.gutehrle@univ-fcomte.fr](mailto:nicolas.gutehrle@univ-fcomte.fr).

## Exercice 1

Employer un prompt pour identifier les mentions de lieux mentionnés dans la colonne `content` du fichier `exercice-td-5-6.csv`. Pour chaque lieu identifié, vous obtiendrez les coordonnées ainsi que leur fréquence d'apparition dans le texte. Vous sauvegarderez ces résultats dans un fichier `lieux.csv` qui contiendra une colonne `lieu` pour les lieux identifiées, `coordonnees` pour les coordonnées et `frequence` pour leur fréquence. Les coordonnées doivent être sous la forme Latitude-Longitude, séparé par une virgule.

Vous vous servirez ensuite de l'outil Palladio pour visualiser les mentions de lieux sur une carte. Vous ferez en sorte que le nom du lieu apparaisse lorsqu'on le survole avec une souris. De plus, la taille de chaque lieu sur la carte doit dépendre de sa fréquence d'apparition dans le corpus.

Que pouvez-vous observer ? Comment les données sont-elles réparties géographiquement ? Quels sont les lieux les plus mentionnés ? Quels sont les lieux les moins mentionnés au contraire ? Vous répondrez à ces questions en fournissant des captures d'écran de l'interface.

## Exercice 2

A l'aide d'OpenRefine ou de prompts, calculez la fréquence de chaque mot dans la colonne `content` du fichier `exercice_td5-6.csv`. Vous stockerez les résultats dans un fichier `frequence_mots.csv` dans lequel une colonne `mot` contiendra chaque mot et une colonne `frequence` contiendra la fréquence respective de chaque mot.

Vous vous servirez ensuite de l'outil RAWGraph pour visualiser les fréquences de mots dans le corpus. Que pouvez-vous observer ? Quels sont les mots les plus fréquents ? Au contraire, quels sont les mots les moins fréquents ? A partir de ces fréquences, pouvez-vous identifier des thématiques dans le corpus ? Vous vous appuyerez d'au moins un graphique pour présenter vos résultats. N'hésitez pas à sélectionner un sous-ensemble de données si cela fait trop pour le graphique.