

Text-mining project documentation

Installer les packages nécessaires et préparer l'environnement

```
library(httr)
library(jsonlite)
#exemple de chargement de base de données
dataset="/Users/user/Desktop/text-mining/VariableCibles.csv"
data <- read.csv(file = dataset,sep=";")
#transform csv to dataframe
df <- data.frame(data)

url_correction <- "http://127.0.0.1:7300/api/correction"
url_categorization <- "http://127.0.0.1:7300/api/categorization"
url_model <- "http://127.0.0.1:7300/api/model"
```

Correction d'une base de données

```
result_correction = content(POST(url_correction,
                                body = list(file = toJSON(df),target="profession"),
                                add_headers(c("Content-Type" = "multipart/form-data"))))
#Ne vous inquiétez pas de la cohérence des résultats car il essaie de ne garder que les mots
#pour faciliter la catégorisation
new_data <- as.data.frame(do.call(cbind, result_correction$data))
new_data$profession <- unlist(new_data$profession, use.names = FALSE)
new_data$Corrected <- unlist(new_data$Corrected, use.names = FALSE)
```

Catégorisation de la colonne profession

```
list_professions <- paste("menagere","enseignement","administration","commerce","medical","eleveurs")
#soyez le plus exhaustif possible en renseigner la liste des catégories
result_categorization = content(POST(url_categorization,
                                     body = list(data = toJSON(new_data),target="profession",elements=list_professions),
                                     add_headers(c("Content-Type" = "multipart/form-data"))))

new_data2 <- as.data.frame(do.call(cbind, result_categorization$data))
new_data2$profession <- unlist(new_data2$profession, use.names = FALSE)
new_data2$Corrected <- unlist(new_data2$Corrected, use.names = FALSE)
new_data2$Categorie <- unlist(new_data2$Categorie, use.names = FALSE)
```

Prédiction avec le modèle de machine learning

```
prediction_categories <- content(POST(url_model,
                                     body = list(data = toJSON(new_data2))))

new_data3 <- as.data.frame(do.call(cbind, prediction_categories$predicted_dataset))
```

```
new_data3$profession <- unlist(new_data3$profession, use.names = FALSE)
new_data3$Corrected <- unlist(new_data3$Corrected, use.names = FALSE)
new_data3$prediction <- unlist(new_data3$prediction, use.names = FALSE)
```