13 mai 2022

S8 – Projet SISN

# **Analyse des données textuelles avec R et Python**

Une intelligence artificielle éthique ?

Exploration à partir d’une analyse textuelle automatisée

*Tutoriel pour mettre en œuvre le RcmdrPlugin.temis*

Tout d’abord, il faut installer le logiciel R sur votre ordinateur. C’est possible de le faire en accédant sur le site du CRAN (<https://cran.r-project.org/>). Dans ce cas-ci, on utilise R 3.6.3 sur Windows que vous pouvez installer avec le lien suivant (<https://cran.r-project.org/bin/windows/base/old/3.6.3/>) ou si vous êtes sur MAC (<https://cran.r-project.org/bin/macosx/>).

Ensuite, il faut chercher sur votre ordinateur l’exécutable « Rgui.exe » qui sera sur l’adresse C:\Program Files\R\R-3.6.3\bin\x64. Ainsi, lancez l’exécutable et tapez la commande suivante :

* *install.packages("RcmdrPlugin.temis")*

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Figure 1: Commande d'installation de paquet

De cette façon, le logiciel cherchera tous les paquets qu’il a besoin pour bien installer « RcmdrPlugin.temis ». Dans le cas d’erreur de paquets, il faut télécharger la version la plus récente de chacun sur le site du CRAN (souvent une recherche du nom du paquet sur google suffit) et les installer manuellement avec l’option ci-dessous :

Interface gráfica do usuário, Texto, Aplicativo, Word

Descrição gerada automaticamente

Figure 2: Installation manuelle des paquets

L’étape suivante sera lancer le « R Commander » avec la commande suivante :

* *library ("RcmdrPlugin.temis")*

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Figure 3: chargement du paquet

En tapant « enter », le paquet sera lancé et vous verrez le menu suivant :

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Figure 4: R Commander

Tous les options pour l’analyse textuelle se trouvent dans l’onglet « text mining » :

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Figure 5: Options d'analyse textuelle

Maintenant, on peut charger le corpus dans le logiciel avec l’option « Import Corpus ». Pourtant, le R Commander n’accepte pas des textes en format PDF, donc on a utilisé le convertisseur en ligne (<https://pdftotext.com>) pour les mettre en format TXT.

Alors, avec tous les fichiers TXT dans un dossier, on peut les charger sur R Commander avec ces options-là :

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Figure 6: Importation du corpus

Après le chargement du corpus, on peut lancer le clustering automatisé des documents :

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Figure 7: Clustering des textes

Le clustering permet de filtrer des mots rares et aussi de choisir le nombre de clusters à être formées :

Interface gráfica do usuário, Texto, Aplicativo, Word

Descrição gerada automaticamente

Figure 8: Options de clustering 1

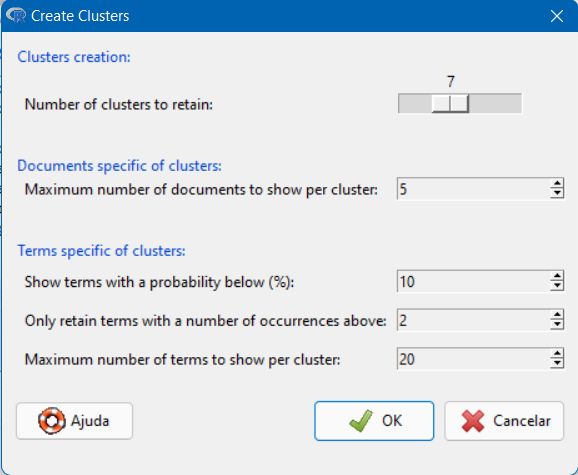


Figure 9: Options de clustering 2

Avant de cliquer sur « OK », on retourne au « R GUI » pour trouver le dendrogramme entier de l’ensemble du corpus. En faisant un clic droit sur l’image du dendrogramme, on peut le sauvegarder en format bitmap ou même l’imprimer en format PDF.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Figure 10: Dendrogramme Entière

Maintenant, on peut retourner à l’écran de « R Commander » pour lancer la clusterisation. Le résultat final sera un rapport avec les documents qui forment chaque cluster et les mots les plus fréquents de chacun. Et sur le « R GUI » le logiciel affiche le dendrogramme final.

Tabela

Descrição gerada automaticamente

Figure 11: Résultat clusterisation

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Figure 12: Dendrogramme final