Big Data: Analyse des grands réseaux d'intéraction

Kadia Touré et Symphoris Tsague Nguegang

16/12/2020

Contents

1	Introduction	1
2	Extraction des publications scientifiques	1
	2.1 Import des données	2
	2.2 Préparation des données	2

1 Introduction

Ce projet à pour but d'étudier l'évolution des réseaux d'intéraction de la communauté scientifique du domaine des neurotransmetteurs entre 2015 et 2019. Ces données sont issus du site PubMed, un moteur de recherche spécialisé dans les publications scientifiques.

2 Extraction des publications scientifiques

Pour cette étude nous allons rechercher les publications scientifiques ayant comme titre "neurotransmitters" entre 2015 et 2019. Pour cela, nous avons réalisé une requête le 16/12/2020 et nous avons obtenu un résultat de 4,939. (cf figure 1) publications.

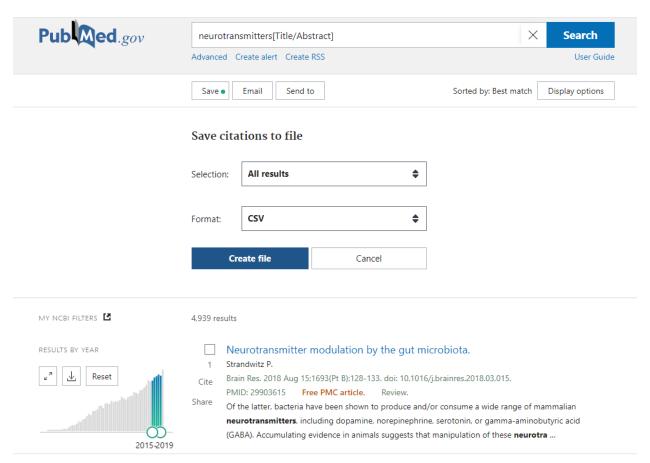


Figure 1:Téléchargement des données sur pubMed de 2015 à 2019

2.1 Import des données

```
## Warning: package 'readxl' was built under R version 3.6.3
#imports des données
pubmed <- read.csv(file = '~/R/projects/BigDataProject/data/excel/csv-neurotrans-set-2.csv')</pre>
```

2.2 Préparation des données

Nous allons récupérer les colonnes "Authors" et "Publication Year et créer 5 vecteurs correspondants à une année".

```
authors <- pubmed[,c("Authors", "PublicationYear")]

#authors
authors2015 <- authors[authors$PublicationYear == "2015",]
authors2016 <- authors[authors$PublicationYear == "2016",]
authors2017 <- authors[authors$PublicationYear == "2017",]
authors2018 <- authors[authors$PublicationYear == "2018",]
authors2019 <- authors[authors$PublicationYear == "2019",]</pre>
```

Nous allons maintenant sauvegarder les auteurs de ces vectors dans un fichier csv (cf figure 2)

```
#On ne garde que les auteurs
authors2015 <- authors2015["Authors"]
authors2016 <- authors2016["Authors"]</pre>
```

```
authors2017 <- authors2017["Authors"]</pre>
authors2018 <- authors2018["Authors"]</pre>
authors2019 <- authors2019["Authors"]</pre>
#sauvegarde en fichier csv
write.csv (authors2015, file = "export/authors2015.csv", row.names=FALSE)
write.csv (authors2016, file = "export/authors2016.csv", row.names=FALSE)
write.csv (authors2017, file = "export/authors2017.csv", row.names=FALSE)
write.csv (authors2018, file = "export/authors2018.csv", row.names=FALSE)
write.csv (authors2019, file = "export/authors2019.csv", row.names=FALSE)
"Authors"
"Mittal R, Debs LH, Patel AP, Nguyen D, Patel K, O'Connor G, Grati M, Mittal J, Yan D, Eshraghi AA, Deo SK, Daunert S, Liu XZ."
"Kandimalla R, Reddy PH."
"Oishi Y, Lazarus M."
"Reddy PH."
"Ascenzi M, Bony G."
"Snyder SH."
"Bazargani N, Attwell D."
"Shea-Donohue T, Urban JF Jr."
"Dumitrescu E, Andreescu S."
"Kondziella D."
"Grässel S, Muschter D."
"Żakowski W."
"Lussier F, Brulé T, Bourque MJ, Ducrot C, Trudeau LÉ, Masson JF."
"Greene DL, Hoshi N."
"Angelovski G, Tóth É."
"Kljakic O, Janickova H, Prado VF, Prado MAM."
"Inoue T, Tanaka S, Okusa MD."
"Shen C, Xue M, Qiu H, Guo W."
"Jiang J, Wang J, Li C.
"Nakamichi N, Kato Y."
"Anderson H, Yang Y."
"Mann FA, Herrmann N, Meyer D, Kruss S."
"Widiapradja A, Chunduri P, Levick SP."
"Gümürdü A, Yildiz R, Eren E, Karakülah G, Ünver T, Genç Ş, Park Y."
```

Figure 2: extrait des premières lignes concernant les plucations de 2018