

# Proposition de TER pour la formation Master Informatique Paris Descartes

## Information concernant l'encadrant

**Encadrant(s) :** Séverine Affeldt et Lazhar Labiod, MCF, Université de Paris  
**Email :** [severine.affeldt@u-paris.fr](mailto:severine.affeldt@u-paris.fr), [lazhar.labiod@u-paris.fr](mailto:lazhar.labiod@u-paris.fr)

## Description générale du projet

### Intitulé du projet :

- Fouille de texte pour l'exploration des maladies similaires au COVID-19 -

### Contexte:

Un nouveau coronavirus responsable de la maladie à coronavirus 2019 (COVID-19; Fig.1), a provoqué une épidémie de pneumonies graves en Chine et dans d'autres pays à partir de décembre 2019. Actuellement, le nombre de patients atteints de la COVID-19 a augmenté rapidement, mais il est encore difficile de définir un traitement efficace contre la maladie ou qui permet d'éviter les formes graves.

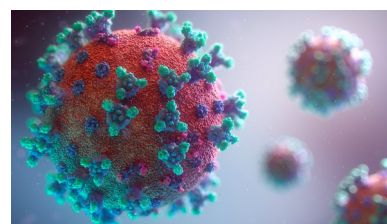


Figure 1: Représentation du coronavirus SARS-CoV-2 © Fusion Medical Animation on Unsplash

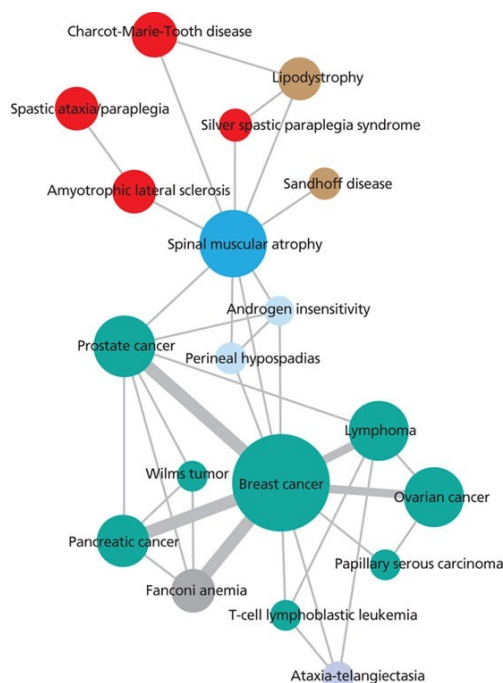


Figure 2: Exemple de réseau de similarités entre les maladies

La mise en évidence de similarités entre certaines maladies aujourd'hui bien connues et la COVID-19 offrirait des pistes intéressantes pour le développement de nouveaux traitements efficaces (voir un exemple Fig.2). Certaines prises en charges médicamenteuses exploitées dans le cadre d'autres infections pourraient également s'appliquer aux patients atteints de formes graves afin de réduire leurs symptômes.

### Objectifs:

Dans ce projet, nous souhaitons explorer les similarités entre des maladies déjà connues et la COVID-19 afin de proposer des traitements innovants susceptibles de réduire la sévérité de la maladie. Nous proposons pour cela d'exploiter la littérature scientifique disponible en ligne.

A partir d'un corpus biomédical portant sur la COVID19, et à l'aide d'approches de fouilles de texte avancées -- telles que le clustering simultané de document et de mots (co-clustering), il serait possible d'identifier les maladies dont les caractéristiques se rapprochent le plus de ceux de la COVID-19. D'autres outils issus du Natural Language Processing (NLP) -- tels que les vecteurs de mots --, pourraient également nous permettre d'explorer les relations entre les symptômes et les traitements.

**Réalisations:**

(1) Ce TER a pour premier objectif de créer un corpus biomédical autour de la COVID19 et des maladies similaires.

(2) Le second objectif de ce projet est l'exploitation d'approche de fouille de texte et de NLP pour l'identification de nouveaux traitements.