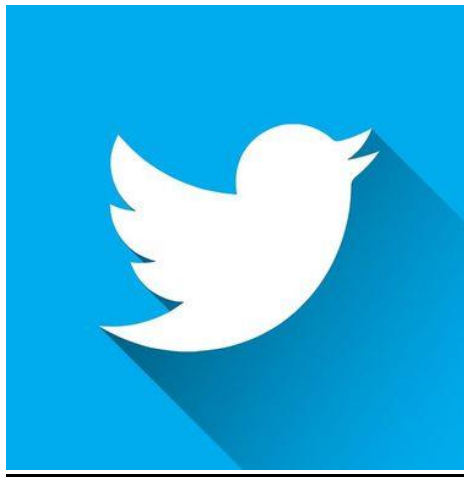


# **PROJET TWITTERBUZZ**

## **LOGICIEL DE DETECTION DES HOT TOPICS SUR TWITTER**

Le traitement automatique du langage naturel est une discipline de l'intelligence artificielle permettant de traiter des données au format textuel. Elle rassemble de nombreuses techniques. Parmi elles, certaines permettent de détecter les *mots saillants* depuis un corpus de texte.

L'objectif de ce projet est d'utiliser ces techniques pour identifier les hot topics de Twitter et de les utiliser pour des analyses plus approfondies.



### **CONTEXTE ET DESCRIPTION DU SUJET**

Twitter est un réseau social très connu permettant aux usagers du monde entier de s'exprimer sur les sujets qui les intéressent. Cet espace d'expression est très riche pour connaître l'opinion publique de la population, et pour connaître les sujets importants à une date donnée.

Cette connaissance permet également d'influencer les décisions de manière plus démocratique.

### **OBJECTIF DU PROJET**

L'objectif du projet est de mettre en place un logiciel de suivi des hot topics sur une période donnée à partir des tweets publiés. Ce logiciel pourra ensuite être appliqué à un cas d'usage, comme celui de la pandémie mondiale de COVID-19 afin de connaître le ressenti des populations face aux différentes décisions politiques à l'international.

## **TRAVAIL A REALISER**

### ***Etapas préalables :***

Le groupe devra se former aux techniques de traitement automatique de langage naturel permettant l'analyse de corpus et la détection de mots saillants. Ces techniques seront utilisées sur des corpus de texte de littérature classique faciles à exploiter.

### ***Réalisation :***

- Le groupe devra tout d'abord une base de données permettant de stocker les données permettant d'exploiter les données utiles à la détection des hot topics
- Il faudra ensuite aspirer des tweets à partir d'une bibliothèque Python et extraire les informations pour alimenter la base de données
- Enfin, il faudra construire les indicateurs de suivi en portant une attention particulière à la possibilité de travailler sur des échelles temporelles différentes et/ou des groupes d'utilisateurs différents.

### ***Pour aller plus loin :***

Les indicateurs conçus seront utilisés pour analyser les Tweets, détecter les influenceurs, et extraire la polarité (opinion positive ou négative) sur les différents sujets d'actualité.