



# Présentation Projet tutorée : Extraction de données à partir de Twitter

DIGNIEL Valentin FARRUGGIA Thomas NOPRE Benjamin VANCOILLE Victor

# Sommaire

- I. Description
- II. Problématique
- III. Analyse
- IV. Besoins
- V. Etude de l'existant
- VI. Solutions
  - A. Conception/Méthode
- VII. Selenium



# Description

Une application qui permettra de récupérer toute les infos d'un compte twitter ciblé et les ajoutées à une base de données, un des but est de pouvoir récupérer les infos d'un compte sans connexions internet mais elle permettra aussi de récupérer des infos sur les followers de la cible si demandée ( si le compte est déjà renseigné dans la base de données).



# Problématique

Le système de recherche de Twitter n'est pas absolu, en effet la base de données limite les résultats.

Donc on peut faire plusieurs fois une même requête et obtenir des résultats différents.

Et les recherches complexes sont impossible.



# Analyse

Nous avons plusieurs solutions qui s'offrent à nous :

Utiliser l'API de Twitter, mais celui-ci nous contraint de :

- Requête par seconde: 10
- Requête par minute: 30
- Requête incomplète
- La plupart des librairies sont pour Python



# Analyse

Nous nous tournerons donc vers une base de données, et d'un robot nommé Selenium, que nous décrirons plus tard.



# Besoins

Il nous faut une base de données relationnelles, et donc un compte Oracle, du robot Selenium, de plusieurs comptes Twitter (que nous créerons, donc un total de 5 en comptant notre tuteur de projet), du langage Java (avec les librairies permettant de manipuler des base de données : JDBC).



# Etude de l'existant



Twitter search



Topsy - Trie

Twinitor- Chronologie





# Solutions

## Conception/Méthode



Selenium est un outil permettant de simuler un humain automatiquement, c'est-à-dire qu'on le configure pour qu'il ouvre twitter, qu'il se connecte, qu'il fasse des recherches et les stocke dans notre base de données.



# Sélénium


- ★ Les possibilités sont infinies, il permet d'automatiser des tâches récurrentes.
- ★ On configure l'outil, se connecter, cliquer, émuler une touche clavier..



# Dans notre cas

- ★ Récupère les différentes données que nous voulons
- ★ Moyen pour être plus rapide : définir un plus petit nombre de données à récupérer, le nombre de saut que l'on fait à partir du compte visé.



A large, dark blue silhouette of a Twitter bird is centered on a light blue background. The bird is facing right, with its wings curved upwards and its tail feathers visible. A horizontal white bar with a thin blue border is positioned across the middle of the image, partially overlapping the bird's body.

Merci de nous avoir écouté