



# Projet 3h : Scraper intelligent de flux RSS

---



## Objectif

Concevoir un programme Python capable de **scanner des centaines de flux RSS** à la recherche de mots-clés donnés, afin de repérer les actualités pertinentes pour un utilisateur.

Le programme fonctionnera en ligne de commande et affichera la **liste des articles correspondants** aux mots-clés recherchés.

---



## Données fournies

- Une liste de **1910 flux RSS** (fichier `rss_list.txt`)
  - Un **notebook Python de démarrage** pour vous guider pas à pas
- 



## Fonctionnalités attendues

Votre application devra :

- Charger les **flux RSS depuis un fichier texte** (`rss_list.txt`)
  - Récupérer **les titres et descriptions** des articles
  - Filtrer les articles en fonction d'une liste de minimum 3 mots-clés que vous choisirez
  - Afficher les informations
  - Enregistrer ensuite dans un fichier `resultat.txt` les informations suivantes
    - Le titre
    - La date de publication
    - L'URL de l'article
    - Le mot-clé correspondant
  - Améliorer grandement les performances en lançant plusieurs requêtes en asynchrone.
  - Poster le résultat sur un repository public et renvoyer le lien du repo
- 



## Contraintes techniques

- Langage : Python
  - Bibliothèques recommandées :
    - `feedparser`
    - `concurrent.futures` (ou `asyncio`)
    - `time`, `logging`
  - Entrée : liste de mots-clés passés via l'appel à un fichier `mots_cles.txt`
  - Sortie : affichage console clair
- 



## Étapes suggérées

Temps	Étapes
0h00 - 0h15	Présentation + test d'un flux RSS avec <code>feedparser</code>
0h15 - 1h00	Charger et parser plusieurs flux manuellement
1h00 - 2h00	Ajout des mots-clés + filtrage
2h00 - 2h30	Ajout du parallélisme et de la tolérance aux erreurs
2h30 - 3h00	Nettoyage du code + rendu <code>README.md</code> + bonus

## Bonus possibles

- Gérer les **erreurs de réseau ou de parsing** sans planter

## Fichiers à rendre

- `README.md` avec description du fonctionnement et instructions d'exécution
- `notebook_scraper.py` avec le code Python complet

## Pour démarrer

1. Installez `feedparser` :

```
pip install feedparser
```

2. Lancez le fichier `notebook_scraper.py` pour voir un exemple minimal