

# TP1 :Web scraping et APIs

1. Creez un compte GitHub et *forkez* le repository contenant tout le code pour ce cours : <https://github.com/madewild/tac>

✓ **Done**

2. Mettez en place un site HTML basique avec GitHub Pages sur la base des explications fournies en annexe

✓ <https://amar-ulb.github.io/>

3. Telechargez sur l'UV le sous-ensemble du corpus CAMille et explorez au hasard quelques fichiers pour avoir une premiere impression de la langue utilisee, des themes abordes, de la longueur des textes, de la qualite de l'OCR, etc.

✓ **Done**

4. Imaginez une requete SPARQL sur Wikidata qui puisse etre utile pour les chercheurs en journalisme utilisant le corpus CAMille.

✓ **Done**

5. Implementez-la dans un court script (sur la base de `s2_sparql.ipynb`) et *pushez* le resultat sur votre repo GitHub dans un dossier nomme « tp1 »

✓ **Done**

<https://w.wiki/5rbr>

**Premier ministre belge, francophone et ayant été député européen.**

6. Rendez-vous sur <https://rapidapi.com/blog/most-popular-api/> et choisissez n'importe quelle API qui presente un interet pour enrichir CAMille (vous pouvez aussi selectionner une API ailleurs que sur ce site)

✓ <https://rapidapi.com/contextualwebsearch/api/web-search>

7. Ecrivez un court script, sur la base de `s3_api.ipynb`, qui permette d'interroger cette API et *pushez*-le sur votre repo GitHub dans le dossier tp1

x Noke