

TP BeautifulSoup4

Scraper le site du Blog du Modérateur avec BeautifulSoup

Vous devez écrire un script Python pour scraper les pages d'articles du site <https://www.blogdumoderateur.com> en utilisant **BeautifulSoup**.

Le script devra extraire les éléments suivants :

1. **Le titre** de l'article
2. **L'image miniature (thumbnail)** principale
3. **La sous-catégorie** (attention : elle peut être identique à la catégorie principale)
4. **Le résumé** (extrait ou chapô de l'article)
5. **La date de publication**, au format `AAAA-MM-JJ`
6. **L'auteur** de l'article
7. Le contenu de l'article (Veuillez bien formater le contenu de l'article dans un seul bloc de texte propre et bien nettoyé)
8. Un **dictionnaire des images** présentes dans l'article, avec :
 - L'**URL** de chaque image dans l'**ordre d'apparition**
 - Sa **légende** (ou texte alternatif, ou description)
9. **Sauvegarder toutes ces informations dans une base MongoDB** dans une collection appelée `articles`

Créez ensuite un script qui permet de retourner tous les articles d'une catégorie ou d'une sous-catégorie.

Bonus : Création d'un front sur lequel vous pouvez effectuer des recherches d'articles selon les critères suivants : date de publication (début/fin), auteur de l'article, catégorie, sous-catégorie, ainsi qu'une recherche par sous-chaîne dans le titre

