

---

**Projet de fin de module Master**  
**Analyse des Publications Scientifiques avec l'API Scopus**  
**(A rendre avant 30/06/2024)**

## **Scopus**

Scopus (<https://www.scopus.com/home.uri>) est une grande base de données bibliographique contenant des résumés et des citations pour des articles de revues académiques. Il est géré par Elsevier, un éditeur majeur de revues scientifiques et de littérature universitaire. Scopus couvre une vaste gamme de disciplines scientifiques, techniques, médicales et sociales, offrant des outils pour la recherche et l'analyse des publications scientifiques.

## **Objectif du projet**

L'objectif de ce projet est de récupérer des informations sur les publications scientifiques à partir de **Scopus** en utilisant leur API, puis d'analyser et de visualiser ces données en utilisant les bibliothèques Python NumPy, Pandas et Matplotlib.

Ce projet vous permet de vous familiariser avec l'utilisation des APIs pour récupérer des données réelles, de les analyser avec des outils puissants comme Pandas et NumPy, et de les visualiser avec Matplotlib. Vous pouvez étendre ce projet en ajoutant des analyses plus complexes ou en récupérant des ensembles de données plus larges.

## **Étapes du Projet**

- 1. Inscription et Configuration de l'API Scopus**
  - Inscrivez-vous sur le site des développeurs Elsevier pour obtenir une clé API.
  - Familiarisez-vous avec la documentation de l'API Scopus pour comprendre comment faire des requêtes et quelles données sont disponibles.
- 2. Récupération des Données de Scopus**
  - Écrivez un script Python pour faire des requêtes à l'API Scopus en utilisant la clé API.
  - Récupérez des informations sur des publications spécifiques en utilisant leur DOI (Digital Object Identifier).
- 3. Analyse des Données**
  - Utilisez Pandas pour manipuler et analyser les données récupérées.
  - Par exemple, calculez le nombre total de citations pour chaque publication, analysez la répartition des publications par année, etc.
- 4. Visualisation des Données**

- Utilisez Matplotlib pour créer des graphiques qui illustrent vos analyses.
- Par exemple, créez un graphique à barres pour montrer le nombre de citations par publication ou un graphique linéaire pour montrer la tendance des publications au fil du temps.

## **Livrables**

- Un script Python documenté qui inclut :
  - La récupération des données depuis l'API Scopus.
  - L'analyse des données récupérées.
  - La visualisation des données analysées sous forme de graphiques.
- Un rapport (d'une dizaine de pages) expliquant le processus, les analyses effectuées et les résultats obtenus, ainsi que les instructions pour exécuter le script.

## **Annexe**

Voici quelques exemples de statistiques que vous pouvez effectuer sur les données récupérées de Scopus en utilisant une ontologie et des outils comme NumPy, Pandas, et Matplotlib :

- 1. Nombre Total de Publications :** Calculez le nombre total de publications dans votre dataset.
- 2. Nombre Total de Citations :** Calculez le nombre total de citations de toutes les publications.
- 3. Citations Moyennes par Publication :** Calculez la moyenne des citations par publication.
- 4. Publications avec le Plus de Citations :** Identifiez les publications avec le plus de citations.
- 5. Distribution des Citations :** Visualisez la distribution des citations par publication.
- 6. Citations par Année :** Si vous avez les dates de publication, vous pouvez analyser les citations par année.
- 7. Répartition des Publications par Auteur :** Si vous avez les informations sur les auteurs, vous pouvez analyser la répartition des publications par auteur.
- 8. Corrélation entre le Nombre de Citations et les Années de Publication :** Calculez la corrélation entre le nombre de citations et les années de publication pour voir s'il y a une tendance.

## **En option**

En intégrant RDF et SPARQL dans votre projet, vous obtenez une structure de données flexible et interopérable qui facilite les requêtes complexes et l'intégration avec d'autres sources de données. Voici les étapes générales suivies dans ce projet :

1. **Récupération des données** à partir de l'API Scopus.
2. **Conversion en RDF** pour créer un graphe RDF structuré.
3. **Stockage et gestion** des données RDF.
4. **Exécution de requêtes SPARQL** pour extraire et analyser les données.