

Approche par graphes temporels pour l'étude de la dynamique et convergence des Recherches sur les Polluants Organiques Persistants et l'endométriase.

Lieu : Toulouse (INRAE ToxAlim, 180 chemin de Tournefeuille, 31207 Toulouse)

Laboratoire d'accueil : Unité TOXALIM (Toxicologie Alimentaire) UMR 1331

Équipe MeX (Métabolisme et Xénobiotiques)

Supervision : Meije Mathé et Clément Frainay

Durée : 6 mois

Période de stage : à partir de Janvier 2025

Niveau : stage de M2

Contexte

Le projet dans lequel s'inscrit ce stage vise à explorer le lien entre les Polluants Organiques Persistants (POP), les modulations métaboliques et les pathologies associées, notamment l'endométriase. Pour cela, il s'appuie sur des graphes de connaissances et des bases de données comme FORUM [1], contenant des millions d'associations entre composés chimiques et concepts biomédicaux. L'objectif de ce projet est de développer des méthodes informatiques pour aider à analyser et interpréter ces données massives, en facilitant la collaboration humains - machines en toxicologie.

L'exploration des liens entre les POP et l'endométriase est un domaine de recherche émergent, et le projet de stage se propose de cartographier l'évolution de ces recherches à travers l'analyse automatique des publications scientifiques et des concepts biomédicaux associés. L'objectif est de comprendre l'évolution de l'intérêt scientifique pour ces sujets, en représentant ces données sous forme de graphes temporels, pour visualiser [2] et analyser les dynamiques des recherches [3].

L'analyse de la littérature scientifique sur les POP et l'endométriase soulève plusieurs problématiques comme l'identification des concepts émergents qui ont récemment été associés à ces sujets et comment leur importance relative a évolué au cours du temps. Il est aussi essentiel d'identifier les publications et découvertes majeures qui ont marqué un tournant dans ces domaines, ainsi que la façon dont elles ont redéfini les directions futures de la recherche. La détermination des indicateurs permettant de détecter les phases critiques et les tendances dans la production scientifique permettra de mieux comprendre l'évolution des connaissances sur les POP et l'endométriase.



Missions

Modélisation :

- Formuler des requêtes SPARQL pour récupérer des données sur la littérature scientifique.
- Modifier et enrichir un graphe de connaissance déjà existant pour y ajouter des informations temporelles.

Visualisation :

- Utiliser Neo4j et KeyLines pour créer des visualisations dynamiques et interactives de graphes temporels.
- Ajuster les visualisations pour représenter l'information de manière pertinente.

Analyse :

- Explorer le graphe en utilisant des métriques propres aux graphes temporels pour identifier des tendances, des phases critiques et des moments clés de l'évolution de la littérature scientifique.
- Concevoir une bibliothèque de requêtes pour extraire des publications selon des critères temporels ou conceptuels (requêtes SPARQL et CYPHER).



Compétences requises

- Connaissances de base en programmation orientée objet.
- Connaissances de base en bases de données et web sémantique.
- Maîtrise de Git et du système Unix.
- Bonne communication orale et écrite.
- Savoir travailler en autonomie et au sein d'une équipe.



Compétences complémentaires

- Savoir programmer en JAVA.
- Connaissances en théorie des graphes.
- Expérience avec des BDD bioinfo (PubMed, PubChem, ChEBI).
- Intérêt pour la recherche en santé humaine et environnementale.



Encadrement et conditions d'accueil

Le stage sera réalisé au sein d'une équipe de bioinformatique de ToxAlim. Il sera encadré par Meije Mathé, étudiante en thèse dans l'équipe, et Clément Frainay, chercheur dans l'équipe.



Candidature

Si vous êtes intéressé.e par cette offre de stage, envoyez votre CV ainsi qu'une lettre de motivation exposant votre intérêt pour le projet à meije.mathe@inrae.fr et clement.frainay@inrae.fr avant le 31/10/2024.

[1] M. Delmas et al. FORUM: building a Knowledge Graph from public databases and scientific literature to extract associations between chemicals and diseases, *Bioinformatics*, Volume 37, Issue 21, November 2021, Pages 3896–3904, <https://doi.org/10.1093/bioinformatics/btab627>

[2] A. Pereira et al.. Semantic Data Visualisation for Biomedical Database Catalogues. *Healthcare*, November 2022, <https://doi.org/10.3390/healthcare10112287>

[3] JC. Delvenne et al. Stability of graph communities across time scales. *PNAS*, June 2010, <https://doi.org/10.1073/pnas.0903215107>