

Offre de Stage de Fin d'Étude de Master - Système de Recommandation Basée sur les Graphes de Connaissances

Laboratoire SAMOVAR - Télécom SudParis & ensIE

1. Introduction et Contexte

Le laboratoire SAMOVAR de Télécom SudParis et de l'ensIE (Évry) recherche un(e) étudiant(e) en fin de cycle Master pour un stage de six mois portant sur le développement d'un système de recommandation basé sur des graphes de connaissances. Ce projet s'inscrit dans le cadre des recherches avancées en systèmes de recommandation, un domaine en pleine évolution qui exploite des données complexes et hétérogènes pour générer des suggestions personnalisées.

L'objectif du stage est de concevoir un modèle de recommandation reposant sur un graphe de connaissances unifié, capable d'intégrer et de fusionner plusieurs sources d'informations. En mobilisant des techniques avancées d'apprentissage profond, notamment les réseaux de neurones pour graphes (Graph Neural Networks, GNN), le projet vise à développer un système robuste, pertinent et performant. Les défis du stage concernent principalement la construction et l'optimisation de sous-graphes, un aspect clé pour garantir l'efficacité et la précision des recommandations.

2. Déroulement du Stage

Le stage se déroulera en trois grandes phases :

Phase 1 : Intégration de sources dans un graphe de connaissances unifié

Dans cette phase, l'étudiant(e) travaillera sur la collecte et l'intégration de plusieurs sources d'informations dans un graphe de connaissances global. Cette étape nécessite la structuration et l'enrichissement des données, ainsi que leur alignement pour assurer une cohérence au sein du graphe unifié. L'objectif sera d'obtenir une base de connaissances qui pourra être exploitée pour les recommandations.

Phase 2 : Extraction de sous-graphes pertinents

Pour réduire le bruit et optimiser les performances du modèle, l'étudiant(e) développera une méthode de sampling intelligente pour sélectionner des sous-graphes à la fois suffisamment riches en informations et suffisamment légers pour être traités rapidement. Cette étape nécessitera l'implémentation de techniques de sélection et de réduction de graphes, en tenant compte des caractéristiques structurelles et du contexte de la recommandation.

Phase 3 : Recommandation via un Graph Neural Network (GNN)

La dernière phase consistera à utiliser un réseau de neurones pour graphes pour générer les recommandations finales. L'étudiant(e) implémentera, entraînera et ajustera un GNN sur les

sous-graphes extraits, afin de maximiser la pertinence des recommandations tout en maintenant une faible latence de traitement.

3. Compétences Attendues

- **Connaissances en apprentissage automatique et réseaux de neurones**, idéalement sur des structures de graphes (GNN).
- **Expérience en manipulation et gestion de données** structurées et non structurées.
- **Compétences en programmation** : Python, et familiarité avec des bibliothèques telles que PyTorch, TensorFlow ou équivalentes pour les modèles de machine learning.
- **Notions de graphes de connaissances** et de bases de données orientées graphes (RDF, Neo4j, etc.).
- **Esprit d'analyse** et capacité à formuler et tester des hypothèses de recherche.
- **Capacité de travail en autonomie et en équipe**, pour interagir avec les membres du laboratoire SAMOVAR.

4. Documents demandés

Les candidat(e)s intéressé(e)s sont invité(e)s à envoyer leur CV et une lettre de motivation à :

- **Julien Romero** (julien.romero@telecom-sudparis.eu)
- **Stefania Dumbrava** (stefania.dumbrava@ensiie.fr)

Ce stage représente une opportunité unique de contribuer à un projet de recherche innovant et à fort impact au sein de Télécom SudParis et de l'ensIIE, dans un environnement de travail stimulant.