

## Offre de stage de M2 recherche en informatique et ingénierie des connaissances

Laboratoire : Dicen-Université Gustave Eiffel-Campus de Marne la Vallée

Durée : 5-6 mois

contact et encadrement : olivier.champalle@univ-eiffel.fr

mots clefs : IC, EIAH, exerciceur, similarité, ML, compétence, pédagogie via le numérique

Projet de recherche :

Nous nous intéressons à la notion de similarité entre exercice dans le cadre des enseignements algorithmique et de programmation informatique quelque soit le langage.

L'idée est de comprendre si il est possible d'identifier des exercices (énoncé + correction)

« proches » ou « distants » entre-eux et comment caractériser cette distance de manière à créer une classification réutilisable.

Cette classification pourrait prendre en compte :

- les concepts ou notions d'algorithmique et leur maîtrise nécessaire pour répondre correctement à l'exercice
- les compétences propre au langage (python, java, C, ..) et leur niveau de maîtrise

Une telle classification permettrait de recommander et/ou substituer des exercices à d'autres en garantissant que les étudiants devront mobiliser, et donc renforcer, des notions et concepts proches attendus pour la bonne réalisation de l'exercice.

Les applications potentielles d'une telle recherche (sur le long terme), pourraient être déployées dans des EIAH de type exerciceur pour permettre à des étudiants de s'entraîner sur des exercices auto-corrigés, mais aussi d'assister les enseignants dans la réutilisation d'exercice sur étagère.

Un autre terrain d'application pourrait être de faciliter la validation ou l'invalidation d'exercices générés par IA.

Attendu :

- Une analyse de l'état de l'art de la littérature scientifique sur la notion de similarité entre exercices, algorithmes, code, ainsi que les moyens utilisés et leurs terrains d'application:
  - similarité entre algorithme, AST [ZS89], ...
  - ML, TAL [RG19], ...
- Création / adaptation / réutilisation (en fonction de la littérature) d'une ou de plusieurs méthodes de détection de similarité entre exercices
- Validation et calibration sur des données expérimentales

Le ou la candidate pourra s'appuyer sur un travail préliminaire de M1 qui a déjà permis de construire une base de données d'exercices et d'explorer quelques premières méthodes automatique de détection de similarité [A.Ziani25].

Références :

[RG19] Nils Reimers and Iryna Gurevych. Sentence-bert : Sentence embeddings using siamese bert-networks. arXiv preprint arXiv :1908.10084, 2019.

[A.Ziani25] Adel Ziani, Détection automatique d'exercices de programmation similaires, rapport de stage de M1 informatique, 2025

[ZS89] Kaizhong Zhang and Dennis Shasha. Simple fast algorithms for the editing distance between trees and related problems. SIAM journal on computing, 18(6) :1245–1262, 1989.