



**Sujet du stage :** étude comparative d'algorithmes d'apprentissage de représentations de programmes pour la fouille de données d'éducation

**Description du sujet.** Améliorer l'efficacité pédagogique des plateformes d'entraînement à la programmation est une problématique en pleine effervescence qui nécessite de construire des représentations fines et exploitables des programmes d'apprenants. Nous nous intéresserons dans ce stage à l'apprentissage de représentations (ou *embeddings*) de programmes à des fins pédagogiques.

L'apprentissage d'*embeddings* de programmes consiste à apprendre une représentation riche et condensée d'un programmes informatique sous forme d'un vecteur numérique de faible dimension. Il existe différentes stratégies d'apprentissage d'*embeddings* de programmes selon que l'on considère d'une part des méthodes supervisées ou non-supervisées et d'autre part selon la nature de la (ou des) représentation(s) primaire(s) exploitée(s) (séquence textuelle du code, arbre syntaxique abstrait, traces de variables, traces d'exécution, etc.).

L'objet de ce stage sera de réaliser une étude comparative des approches supervisées et non-supervisées parmi ces méthodes. En particulier nous nous intéresserons d'une part à l'approche *code2vec* (Alon et al., 2019) qui exploite de manière supervisée les arbres syntaxiques abstraits (AST) des programmes et d'autre part à l'approche *code2aes2vec* (Cleuziou&Flouvat, 2021) qui génère de façon non-supervisée des *embeddings* de programmes à partir de l'analyse conjointe des AST et des traces d'exécution des programmes.

**Mots-clés.** Intelligence Artificielle, Fouille de Données d'Education, Apprentissage Automatique, Apprentissage de Représentations, Réseaux Neuronaux.

**Candidatures.** Les candidatures (CV, notes et lettre de motivation) doivent être adressées **avant le 10 janvier 2022** à Guillaume Cleuziou ([guillaume.cleuziou@univ-orleans.fr](mailto:guillaume.cleuziou@univ-orleans.fr)).

**Profil attendu.** Etudiant en master et/ou en école d'ingénieur en Mathématiques/Informatique. Une expérience en Machine Learning et un intérêt pour l'analyse de données d'éducation seraient un plus.

**Encadrement du stage.** Le stage se déroulera au LIFO (Laboratoire d'Informatique Fondamentale d'Orléans) de l'Université d'Orléans. Il sera co-encadré par :

LIFO : Guillaume Cleuziou, Matthieu Exbrayat

LIFAT : Nicolas Labroche

ISEA : Frédéric Flouvat

Structures de recherche impliquées :

- Equipe CA (Contraintes et Apprentissage) - LIFO - Université d'Orléans (Orléans)
- Equipe BdTln (Bases de Données et Traitement du Langage Naturel) - LIFAT - Université de Tours (Blois)
- Laboratoire ISEA - Université de la Nouvelle-Calédonie (Nouméa)

**Durée du stage :** 5 mois (à partir de janvier/février 2022)

**Financement du stage.** Le stage sera financé par la fédération ICVL (Informatique en Centre-Val de Loire) regroupant les laboratoires d'informatique d'Orléans (LIFO) et de Tours (LIFAT). Indemnité mensuelle réglementaire (~600€/mois).

### Références.

- Alon, U., M. Zilberstein, O. Levy & E. Yahav (2019). code2vec : Learning distributed representations of code. Proceedings of the ACM on Programming Languages 3(POPL), 1-29.
- Cleuziou, G. & F. Flouvat (2021). Learning student program embeddings using abstract execution traces, in 14th International Conference on Educational Data Mining. Paris, France, Pp. 252-262. 06. 2021.
- Wang, K., R. Singh, and Z. Su (2018). Dynamic neural program embeddings for program repair. In International Conference on Learning Representations (ICLR'2018).