

# Étude de l'évolution des code clones dans Linux

## Contexte

Les code clones sont fréquents dans les systèmes logiciels. Ils apparaissent en général à la suite de copier-coller de fonctionnalités entre plusieurs fichiers sources. Malgré les possibilités de factorisation de code offert par les langages de programmation, les code clones ne peuvent totalement être éliminés et peuvent même être bénéfiques. Les code clones posent cependant un problème majeur de maintenance. En effet, si un bug est corrigé dans une partie de code qui a été clonée, le correction doit être effectuée au sein de tous les clones. Il est donc très important de pouvoir identifier tous les clones d'un code source. Pour cela, il existe des outils nommés détecteurs de code clones.

## Travail

Ce stage de master recherche à pour problématique principale d'étudier l'évolution des code clones, ainsi que l'efficacité des outils de détection. L'idée est de se servir de l'historique de développement des logiciels, généralement accessibles via un système de gestion de version, pour voir comment les code clones évoluent et à partir de quand les détecteurs de code clones n'arrivent plus à les détecter. L'objectif est d'arriver à comprendre les types de changements qui empêchent la détection d'un code clone qui était pourtant détecté à la version précédente. Pour cela, l'étudiant devra mettre en place une procédure expérimentale qui utilise comme jeu de données le code source de Linux, via le système de gestion de versions Git, et qui vise à comparer les codes clones détectés à l'aide de plusieurs outils, sur plusieurs versions. En s'appuyant sur ces résultats, l'étudiant devra ensuite analyser les types de changements à une version donnée qui ont empêchés les détections de code clones.

## Encadrants

- [Laurent Réveillère](#) : Maître de Conférences à l'ENSEIRB-MATMECA
- [Jean-Rémy Falleri](#) : Maître de Conférences à l'ENSEIRB-MATMECA

## Références

- Cory Kapser, Michael W. Godfrey: "Cloning considered harmful" considered harmful: patterns of cloning in software. Empirical Software Engineering (ESE) 13(6):645-692 (2008)
- Lingxiao Jiang, Ghassan Misherghi, Zhendong Su, Stéphane Glondu: DECKARD: Scalable and Accurate Tree-Based Detection of Code Clones. ICSE 2007:96-105
- Nicolas Palix, Julia L. Lawall, Gilles Muller: Tracking code patterns over multiple software versions with Herodotos. AOSD 2010: 169-180