

OBJECTIF

Dans ce TP, vous allez créer un prototype d'analyse automatisée, basé sur le logiciel que vous avez créé pour TP1. Puis vous allez l'utiliser pour étudier la corrélation entre deux métriques.

TÂCHE 1 (30%)

Créez le programme *proto*. Il prend comme entrée le URL d'un dépôt Git et produit comme sortie un fichier CSV (« *comma separated values* », valeurs séparées par des virgules) avec les colonnes : `id_version`, `n_classes`

- **id_version** : identificateur hexadécimal d'une version. Votre programme peut faire appel à la commande `git rev-list master` pour trouver l'ensemble des identificateurs des versions dans l'historique de la dernière version du branchement principal.
- **n_classes** : nombre de classes (pour nos besoins : nombre de fichiers *.java) dans une version. Votre programme peut faire appel à la commande `git reset --hard <HEX>` pour amener le dépôt git local à la version avec l'identificateur <HEX>.

TÂCHE 2 (30%)

Ajoutez fonctionnalité à *proto* pour ajouter à sa sortie la colonne `m_c_BC`

- **m_c_BC** : *médiane* de la métrique `classe_BC`. La métrique `classe_BC` est définie dans TP1. Utilisez votre implémentation du TP1 pour calculer les valeurs `classe_BC` pour les classes de chaque version. Idéalement, *proto* devrait faire appel à votre programme du TP1 comme commande externe.

(Ici, nous n'évaluerons pas la correction de votre implémentation du TP1. Mais votre code de TP1 doit au minimum réussir à générer les fichiers avec les données requises. Au pire cas, utilisez un *mock* pour une pénalité de 15%.)

TÂCHE 3 (40%)

Appliquez *proto* au code du <https://github.com/jfree/jfreechart>. (Dès décembre 3 2020, jfreechart comptait 4044 commit. Si vous le souhaitez, vous pouvez choisir de vous concentrer sur un échantillon aléatoire de 10% du total.) Soumettez dans votre ZIP le fichier CSV généré par *proto*. Après, évaluez l'hypothèse selon laquelle pour les versions du jfreechart il existe une **corrélation** entre les métriques `n_classes` et `m_c_BC`. Basez votre analyse sur l'hypothèse que leurs valeurs **ne sont pas** normalement distribuées.

BONUS (5%)

Montrez visuellement l'évolution des deux métriques à travers l'histoire de jfreechart et essayez d'expliquer de façon qualitative ce que vous observez.

Décrivez brièvement votre travail dans votre rapport.

Expliquez toute hypothèse et décrivez comment exécuter *proto*.

CONDITIONS DE RÉALISATION

- Mêmes groupes que TP3.
- Travail à remettre le **vendredi 18 décembre 23h59** via StudiUM.
- Le membre de l'équipe qui ne soumet pas le fichier Zip principal de la soumission doit soumettre un fichier contenant les noms de l'équipe, comme dans les autres TP.
- Vous pouvez utiliser votre IDE et langage préféré (python, groovy, bash, java, ...). Vous êtes encouragés à utiliser les outils standards de UNIX (ls, shuf, echo, grep, wc, awk, ...) et vous pouvez mélanger des langages (ex : un script bash qui fait appel à un script R pour des calculs numériques). Cependant, nous n'offrons pas de conseils sur l'utilisation des IDE ou sur le scripting.
- Mettez le code/script du *proto*, le fichier CSV et votre rapport en format PDF dans le Zip de votre soumission.
- Pour T3 vous pouvez utiliser n'importe quel logiciel qui appuie les opérations statistiques requises (Excel, R, Octave, SPSS, Python, Java, ...). Cependant, nous n'offrons pas de conseils sur l'utilisation des logiciels statistiques. **Ajoutez votre code/script/fichiers dans le Zip de votre soumission.**
- Rapports de plus de 10 pages (12p, simple espace) vont être éliminés d'office.
- Assurez-vous de communiquer clairement, la qualité de votre rapport est **très importante**.