

Analyse de l'implantations du hachage dans les langages de programmation courants 2022 – 2023

Tuteurs : Cyril Nicaud et Pablo Rotondo (LIGM, équipe BAAM) [prenom.nom@u-pem.fr]

Partenaire : Conrado Martínez Parra (UPC, Barcelone, Espagne) [conrado@cs.upc.edu]

Filière ESIEE : Informatique

Présentation générale : l'équipe BAAM est une équipe d'algorithmique dont une des spécialités est d'analyser comment les différents langages de programmation ont implantés leurs algorithmes standard (tris, hachage, pattern matching, ...). Nous avons par exemple découvert un bug [1] dans le tri de JAVA. Plus récemment, nous avons remis en cause [2] les choix d'implantation du hachage dans LUA. En regardant au cœur des langages comme PHP, JAVA, RUST, ... il s'agira dans ce stage d'étudier la façon dont le hachage est implanté, de trouver et prouver d'éventuels dysfonctionnements, et de proposer des solutions plus efficaces.

Résultats attendus : On peut résumer les étapes d'une analyse de la façon suivante :

- étudier attentivement le code source des langages considérés (le plus souvent en C) afin d'en extraire précisément l'algorithme utilisé (cela peut demander d'enrichir le code du langage pour pouvoir accéder à la représentation interne des structures à l'exécution) ;
- si l'algorithme s'éloigne des algorithmes standard pour faire du hachage, ce qui arrive fréquemment, analyser expérimentalement ses performances en se basant sur des scénarios plausibles d'utilisation ;
- si des dysfonctionnements sont repérés, en général il s'agit de performances moins bonnes que prévu, mener une analyse mathématique des scénarios les mettant en évidence ;
- proposer des améliorations pertinentes.

L'élève suivant ce stage devra avoir un goût pour l'algorithmique et la programmation. Il n'est pas nécessaire de connaître beaucoup de langages. Les mathématiques nécessaires pour l'analyse théoriques reposent en général sur des probabilités élémentaires. Il arrive que des choix d'implantation soient guidés par l'architecture des ordinateurs : des connaissances dans ce domaine sont les bienvenues, mais pas nécessaires.

Remarques : Conrado Martínez Parra parle couramment français. Conrado Martínez Parra et Pablo Rotondo parlent couramment espagnol.

References

- [1] Nicolas Auger, Vincent Jugé, Cyril Nicaud, Carine Pivoteau: On the Worst-Case Complexity of TimSort. ESA 2018. <https://arxiv.org/abs/1805.08612>.
- [2] Conrado Martínez, Cyril Nicaud, Pablo Rotondo: A Probabilistic Model Revealing Shortcomings in Lua's Hybrid Tables. *Accepted at COCOON 2022*. <https://arxiv.org/abs/2208.13602>