

Sujet de PSTL : Génération et réparation d'instances pour JSON Schema

Encadrants : Mohamed-Amine Baazizi (mohamed-amine.baazizi@lip6.fr),
Lyes Attouche <lyes.attouche@dauphine.psl.eu>

Objectifs

JSON Schema [1] s'est imposé comme standard pour décrire et valider les données JSON.

De nombreux outils permettent de générer des instances à partir d'un schéma mais ne garantissent pas un résultat correct. Il devient alors intéressant de réparer les instances ainsi générées en opérant une modification minimale mais suffisante pour les rendre conformes au schéma de départ.

L'objectif de ce projet est de mettre en œuvre une plateforme d'expérimentations visant à évaluer différentes approches de réparation en terme de cout des opérations nécessaires pour produire une instance valide à partir d'une instance non-valide.

Tâches à réaliser

- 1- Prendre en main trois générateurs d'instances préalablement identifiés [2-4] pour générer des instances et analyser leur conformité au moyen d'un validateur standard.
- 2- Étudier l'approche [6] de génération de distance d'édition entre documents JSON.
- 3- Générer, pour les instances non conformes obtenues en (1), les distances d'édicions calculée avec une instance valide qui servira de témoin.
- 4- Étudier le lien entre les erreurs de validation et la distance d'édition.
- 5- Mettre en place une package expérimental reproductible utilisant Docker.

Prérequis

- Maîtrise de Java, Javascript et C# langages utilisés par les trois librairies identifiées
- Sens analytique et lecture d'articles scientifiques

Références

- [1] <https://json-schema.org>
- [2] Jsongenerator. <https://github.com/jimblackler/jsongenerator>
- [3] Json schema faker. <https://github.com/json-schema-faker/json-schema-fakeryes>
- [4] JSON everything. <https://github.com/gregsdennis/json-everything>
- [5] Lyes Attouche, Mohamed Amine Baazizi, Dario Colazzo, Francesco Falleni, Giorgio Ghelli, Cristiano Landi, Carlo Sartiani, Stefanie Scherzinger: A Tool for JSON Schema Witness Generation. EDBT 2021: 694-697
- [6] Hütter, T., Augsten, N., Kirsch, C. M., Carey, M. J., & Li, C. (2022, June). JEDI: These aren't the JSON documents you're looking for... In *Proceedings of the 2022 International Conference on Management of Data*