

Résumé de l'article “Enumerating Regular Languages with Bounded Delay”

Cet article a été publié dans la conférence européenne “Symposium on Theoretical Aspects of Computer Science” (STACS) en 2023, qui est une des conférences de référence en informatique théorique, dont les actes sont disponibles en libre accès.

Nous étudions dans cet article un problème abstrait d'énumération sur les langages réguliers : on cherche à produire la séquence infinie des mots du langage, sans doublons, en produisant chaque mot par une petite modification du mot précédent. Par exemple, pour le langage a^* , on peut produire la séquence $\epsilon, a, aa, aaa, \dots$, où chaque mot s'obtient à partir du précédent en rajoutant un a . En revanche, pour le langage $a^* + b^*$, une telle approche ne peut pas se généraliser. On s'intéresse plus précisément à des modifications décrites par des insertions, suppressions et substitutions (suivant la distance de Levenshtein), et on cherche à savoir si un langage donné peut être écrit comme une séquence infinie où la distance d'édition entre deux mots consécutifs est borné par une constante.

Nous caractérisons les langages réguliers pour lesquels une telle séquence existe. Nous démontrons par ailleurs que, dans le cas contraire (par exemple pour $a^* + b^*$), on peut toujours partitionner le langage en un nombre fini de langages énumérables. De plus, nous montrons que, pour les langages énumérables, on peut toujours se limiter à des opérations d'édition où on insère ou supprime à une des extrémités du mot (gauche ou droite), sans modifications au milieu du mot. Nous montrons enfin que, si l'énumération est possible, alors la séquence infinie des mises à jour peut être efficacement calculée, c'est-à-dire que le délai entre deux mots successifs reste bornée lui aussi même lorsque la longueur des mots croît indéfiniment. Nous montrons enfin que notre caractérisation est effective : étant donné un automate déterministe qui définit un langage régulier, alors on peut décider en temps polynomial si l'énumération est possible, ou à défaut on peut déterminer le nombre minimal de langages énumérables en lequel on peut le partitionner.

Ce travail est une collaboration avec Mikaël Monet (Université de Lille), dont j'ai précédemment co-encadré la thèse, et qui est à présent chargé de recherches à Inria Lille. Nous avons défini le problème ensemble avec Mikaël, j'ai trouvé les arguments techniques principaux, et nous les avons développés ensemble et avons rédigé l'article ensemble.