

QuickFill : travailler plus pour générer moins en synthèse de programmes

Vanessa Fokou*, Peggy Cellier**
Maurice Tchuente*, Alexandre Termier**

*Université de Yaoundé I, Faculté des Sciences, Département d'Informatique
BP 812 Yaoundé Cameroun

IRD, Sorbonne Université, UMMISCO, F-93143, Bondy, France
vanfokou@gmail.com, maurice.tchuente@gmail.com

**Univ Rennes, Inria, INSA Rennes, CNRS, IRISA - UMR 6074, France
prenom.nom@irisa.fr

Résumé. Afin de faciliter l'exécution de tâches répétitives, des approches de synthèse de programmes ont été développées. Elles consistent à inférer automatiquement des programmes qui satisfont l'intention d'un utilisateur. L'approche la plus connue est *FlashFill* qui est intégrée au tableur Excel. Dans *FlashFill* l'intention de l'utilisateur est représentée par des exemples, i.e. des couples (*entrée, sortie*). *FlashFill* explore un très grand espace de recherche des programmes et peut donc nécessiter un temps d'exécution important et inférer beaucoup de programmes dont certains fonctionnent sur les exemples donnés mais ne capturent pas l'intention de l'utilisateur. Dans cet article nous présentons *Quickfill*, une approche qui réduit l'espace des programmes à explorer en enrichissant les spécifications fournies par l'utilisateur. Nous montrons qu'avec cette approche, il est souvent possible de donner moins d'exemples qu'avec l'algorithme *FlashFill* pour une plus grande proportion de programmes corrects.

1 Introduction

Le pré-traitement des données est l'une des tâches les plus chronophages, et sans doute les moins intéressantes, d'un processus d'analyse de données. Cette tâche demande souvent de faire des transformations non triviales de formats, qui sont trop complexes pour être effectuées par les outils existants, mais trop simples pour mériter l'attention soutenue d'un analyste (e.g., extraire les initiales du nom complet d'une entité afin d'obtenir son acronyme, cf figure 1). En pratique, la solution classique pour effectuer les tâches de pré-traitement de données est l'écriture de petits scripts dans des langages de programmation généralistes (e.g., Python). La difficulté est que de nombreux utilisateurs ne maîtrisent pas les bases de la programmation, et n'ont donc pas la possibilité d'écrire de tels scripts.

Pour pallier ce problème, une solution très prometteuse est l'utilisation de techniques de "synthèse de programmes" (*program synthesis* en anglais) (Gulwani et al., 2017). La synthèse de programmes consiste à inférer automatiquement des programmes qui satisfont l'intention