Approaches to Transparent Program Synthesis in Excel

Justin A. Middleton
North Carolina State University, Raleigh, USA
jamiddl2@ncsu.edu

April 27, 2016

Abstract

Program synthesis in Excel grants spreadsheet users a powerful way of transforming input without much work. However, users have no insight into what programs are being made and, therefore, they can't apply automatically generated solutions to situations or learn programs on their own. My work here is to meet the program synthesizing capabilities of current spreadsheets but make the resulting program accessible to the user, both in visibility and in comprehensibility. In the process, I evaluate a number of approaches to the this problem of transparent synthesis, ranging from the algorithms of FlashFill itself to principles of planning.

1 Introduction

Millions of programmers rely on spreadsheet programs, like Microsoft Excel, to make sense of their data, and, as part of some of my other ongoing research outside of class, Im looking for ways to make their lives easier. Already shipped with Excel is a feature called Flash Fill given a set of inputs and a few sample outputs, it tries to create a program which captures the pattern of transformation which maps the set of inputs to outputs and then applies it to the rest of the incomplete values. However, it does this opaquely; users dont have insight into the program created, which impairs deep comprehension and reusability of the program. To be clearer, an example from the paper of FlashFill [1] asks about creating a program which captures the following two columns:

I took these words straight from my 582 pitch – no fears, will change.

Input	Output
John DOE 3 Data [TS]865-000-0000 - - 453442-00 06-23-2009	865-000-0000
A FF MARILYN 30S 865-000-0030 4535871-00 07-07-2009	865-000-0030
A GEDA-MARY 100MG 865-001-0020 - - 5941-00 06-23-2009	865-001-0020

What needs to happen is for a program to discover the correct patterns and manipulations to take the string in the input column and change it to that in the output. In this case, there are many acceptable ways to do this: finding three sets of hyphen-separated numbers with digits in quantities of 3-3-4; finding some hyphen-connected numbers beginning with 865; etc.

References

[1] Sumit Gulwani. Automating string processing in spreadsheets using input-output examples. In *ACM SIGPLAN Notices*, volume 46, pages 317–330. ACM, 2011.