A Tool for Visualizing Patterns of Spreadsheet Function Combinations

The Protracted and Generally Unpleasant Death of Justin A. Middleton
A Methodological Obituary

Justin A. Middleton (posthumous)
North Carolina State University
Raleigh, North Carolina
(Please address all flowers and review notes to cemetery lot 126, grave 2)

***Abstract*—Spreadsheet environments often come equipped with a plethora of functions to manipulate and calculate data, but it can be difficult to understand how end-users combine these functions in practice. Without this knowledge, both researchers and practitioners lack information about how end users construct sophisticated programs from these basic elements. We developed a tool that visualizes patterns of how functions are combined into formulae within Excel spreadsheets. Using the Enron spreadsheet dataset as an example, this paper shows how the tool can display both common and anomalous formulas and their respective contexts in an actual workbook.**

I. Introduction

The world of business owes a considerable debt to spreadsheets, the table-based interface which enables organization and manipulation of massive amounts of data. Their allure is in their versatility: while the end user does not need sophisticated programming skills to fill out a spreadsheet, the expert user has at their disposal numerous built-in operations, or functions, to expedite their work. As such, it should be of little surprise when Scaffidi and colleagues estimated that by 2012, over 50 million workers could be using them, including the 25 million who would be writing programs out of the functions included (citation and such).

Considering this ubiquity, it's crucial to get spreadsheets right. Something, something.

Fortunately, the vanguards of spreadsheet research have assembled, organized, and released a number of spreadsheet corpora to inform work on how people actually use these tools in different contexts. Collections like EUSES[yada], FUSE[yada], and the Enron corpus[yada] have already enabled fruitful work across the field: [WORK PENDING MORE RESEARCH]. A number of these studies focus on discovering which built-in functions are the most common (enabling something).

This paper presents a tool, informed by such variegated sources, that visualizes not how people employ individual Excel functions but how they combine to make more sophisticated shreadsheet formulas.

II. Related Work

Spreadsheets, being a core component of business that they are, have prompted the development of many tools to evaluate them and their many qualities. [RESEARCH ON NON-VISUAL ANALYSIS TOOLS]

Furthermore, this tool comes from a line of spreadsheet visualization tools before it, each with a different focus. For example, Breviz, developed by Hermans and colleagues, uses visualization to illustrate workbook dataflow to address the need for a clearly communicable representation of spreadsheets as they move from person to person.

Other studies focus on API and built-in function use outside the domain of spreadsheets. [BUT FIRST I HAVE TO FIND THEM]

III. Approach

In visualizing the spreadsheet data, I aimed to accomplish the following goals:

- The data should come from a varied enough sources to avoid having too narrow or specialized a scope.
- The tool should accommodate dataset sizes from the single spreadsheet to corpora of thousands.
- The tool should visualize as much data as possible in order to capture (and clearly indicate) both the most typical and the most anomalous cases...
- ...while being interactive enough to let the user choose which cases they see, as to not sacrifice readability.
- The tool should facilitate quantitative understanding by showing which functions and which combinations are used most frequently...
- ...and it should facilitate qualitative understanding by incorporating actual examples into the visualization from the datasets it studies, and it should help the user find these examples in their original context.
- The tool, by the same method as above, should thus guide a user from the abstract patterns of function usage into the specific and applied techniques of functions in their natural habitat.

Likewise, there were a few goals that I was specifically not trying to accomplish with this:

- The tool should not be, by itself, responsible for generating functions from the data it collects. It is purely analytical.

- The tool will not visualize isolated Excel functions, linking the specific instance to the abstract pattern. It works only the other way.
- Likewise, the tool will make no effort to explain what any given set of functions in combination actually does; the function of any given example is up to the user to infer.

IV. Case Study

The patient exhibits strong cadaverous tendencies.

V. Conclusion

This paper is the end-all, be-all. There is no future work simply because the concept of future work cannot exist with reference to this paper. The work is done. I perfected it. Go home, kiss your love, kick back, call it a day. This is not just a good thing, but the best thing. Your job here is done, and I'm the one who did it.

Acknowledgment

This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.