Dear Editors,

Thank you for the opportunity to publish our work on the spreadsheets formula grammar to the JSEP special issue. The paper that we are submitting is enriched in comparison to the SCAM paper in all aspects: grammar, used datasets, evaluation and analyses.

Since the SCAM publication, two factors enabled further enrichment of our grammar: first, presenting the grammar to a wider audience and making its parser open source enabled users to test it with their own datasets and discover new parse errors. Second, a new spreadsheets dataset became available, FUSE, and we further accumulated a fourth one by crawling WikiLeaks. On this extended version of the paper we performed grammar analysis on those two new datasets too, moving from analyzing 20K spreadsheets with 1M unique formulas in the SCAM paper to analyzing 360K spreadsheets with 8.5M unique formulas in this new version.

Summarizing the enrichments made to submitted paper, in comparison to the one presented in SCAM, we have:

1. Evaluated the grammar against and analyzed 2 new datasets, using with eight times more formulas than used in the previous paper (information about the spreadsheets, cells and formulas that each dataset contains is summarized in Table 3)
2. Updated the grammar, added support for structured references and rare cases found in the 2 new datasets
3. Enriched the grammar analysis section (4.3) with discussion on the structured references and the newly supported rare cases
4. Added section 4.2 Parse trees, presenting statistics on four complexity indicators for the dataset formulas: the parse tree depth, the formula depth, the operator depth and the conditional depth
5. Added table 5 and discussion on the frequency of the most common functions and operators found in the unique formulas of the four datasets
6. Added table 6 and discussion on the frequency of the most common constants found in the unique formulas of the four datasets
7. Added figures 5 and 6 and discussion on the frequency of functions, constants, references and prefixes in the unique formulas of the four datasets
8. Improved the description of the intersection operation in the smelly grammar constants
9. Added Section 3.1 Design process explaining our approach for designing and improving the grammar

The comments that we received from the reviewers of the SCAM submission proved very helpful in improving the paper:

- We have added an example parse tree for our grammar as well as for the official Excel grammar for comparison (Figure 1)

- We enriched our motivation description, explained that the grammar can support research on spreadsheet formula codebases and can enhance the understanding and usability of research results and we summarize what our grammar can be used for (third paragraph of Introduction)
- We shortened come discussions in the background section
- We added a discussion on parse tree correctness in a separate Section 5.4
- We added a discussion on dialects and Excel bias in a separate Section 5.1
- We added Section 3.1 Design process describing how the grammar evolved before and after the SCAM publication
- We have added more analyses of the formulas, their syntactical features and complexity, and we have identified more rare cases of formulas in Sections 4.2 and 4.3

Looking forward to your feedback on our paper,

Efthimia Aivaloglou