EuSpRIG 2011 Conference Summary

Breviz: Visualizing Spreadsheets using Dataflow Diagrams

Felienne Hermans, Martin Pinzger, Arie van Deursen

This paper won the EuSpRIG 2011 David Chadwick Student Prize. I had always discounted previous research into DFDs because of the spaghetti diagram from real-sized spreadsheets, but Felienne has developed it further into a layered diagram of semantically related blocks. Early days, but promising.

Spreadsheets in Financial Departments: An automated analysis

Dr. Kevin McDaid, Dr Ronan MacRuairi, Mr. Neil Clynch, Mr. Kevin Logue, Mr. Cian Clancy, Mr. Shane Hayes

All commercial spreadsheet discovery and monitoring applications analyse repositories. What is unique about the DKIT approach is an innovative system which can at last give companies a handle on the amount of time people are spending with Excel, on which project, doing what operations in Excel. This at last holds out the possibility of spreadsheet productivity improvements, which is the Holy Grail of researchers like Tom Grossmann and practitioners like us. Up to now, we had to rely on over-the-shoulder observation and on-the-job training and consulting along the lines of "here's an easier way in Excel to do this", "There's a better solution to this problem", "This could be automated completely if you adopted x technology" etc. Now we still need Excel/VBA experts to assess the findings but they will have a better view of the overall process and daily use of spreadsheets.

Requirements for Automated Assessment of Spreadsheet Maintainability

José Pedro Correia, Miguel A. Ferreira

This starts from existing software quality assessment techniques to map metrics of static analysis to ratings. As we know, all the main commercial assessment frameworks already collect spreadsheet metrics. Their contribution is to group these into a five-point scale of maintainability.They have made a start but need more than the commonly available teaching or example spreadsheets to work with. They offer to curate a repository of serious spreadsheets. Send them some big ones, anonymised! I have a list of about 100 big ones I've given them.

From Good Practices to Effective Policies for Preventing Errors in Spreadsheets

Daniel Kulesz

Unlike the sig.eu paper, they combine human ratings (which sig,eu avoid) with metrics from static analysis. They intend to allow the rules engine to be expanded using a Java API. This, again, is very similar to what the commercial spreadsheet assessment platforms do.

Controls over Spreadsheets for Financial Reporting in Practice

Nancy Coster, Linda Leon, Lawrence Kalbers, and Dolphy Abraham

25 of the 38 companies surveyed have assets between $1Bn and $100Bn. Change management was the no.1 self-identified area of concern. Although over half say they have a review process with a checklist, 30% report the reviewers have no domain experience. 11 indicated that internal reviews have discovered lapses in controls. Astonishingly most say they plan to implement Excel Track Changes, apparently without realising the problems that brings in its wake.

Leveraging User Profile and Behavior to Design Practical Spreadsheet Controls for the Finance Function

Nancy Wu

The most implementable controls are those that do not require advanced Excel knowledge and enhance existing paper review activities – validation, labelling and separating assumptions.

Towards Evaluating the Quality of a Spreadsheet: The Case of the Analytical Spreadsheet Model

Thomas A. Grossman, Vijay Mehrotra, Johncharles Sander

They select six quality dimensions: Suitable, Readable, Transferable, Accurate, Reusable, Modifiable. For 'analytical' spreadsheets, ie those that take some inputs, perform some non-obvious calculations, and present reports to decision makers, they propose these design principles: Modular, Structured, Input-Output, Information Flow, Separate Inputs & Computations & Reports.

In Search of a Taxonomy for Classifying Qualitative Spreadsheet Errors

Zbigniew Przasnyski, Linda Leon, and Kala Chand Seal

This consists of a list of what I call 'bad smells' in spreadsheet structures. One, 'Poor layout for model extension' is a common problem addressed by the FormulaDataSleuth presentation later. Some of these can be identified by automated tools, some only by expert inspection.

Workbook Structure Analysis – "Coping with the Imperfect"

Bill Bekenn and Ray Hooper

This is particularly suited to spreadsheets designed by people with a 'product' focus where there is a block of similar calculations repeated for a large number of products; as distinct from the data-oriented tabular approach or the calculation-oriented modelling approach. The tool uses algorithms to identify stripes, areas, and blocks, and facilitates safer copying of these blocks with the correct relative and absolute addresses.

Effect of Range Naming Conventions on Reliability and Development Time for Simple Spreadsheet Formulas

Ruth McKeever & Kevin McDaid

This experiment asked people to create formulas using only names and prevented them from checking what cells the names referred to. So, the less meaningful the name, the worse the user's guesses were. It shows that you can't expect automatic improvements by simply applying commonly recommended practices without the implicit context that goes with them. This kind of research shows the need for practitioners to more carefully define the implicit 'terms and conditions' of their best practice guidelines.

An Empirical Study on End-users Productivity Using Model-based Spreadsheets

Laura Beckwith, Jácome Cunha, João Paulo Fernandes, João Saraiva

They apply database normalisation techniques to refactor spreadsheets with data tables, then add automatic data validation based on inferences from the table relationships. The tradeoff is that the refactored model is harder to understand for the first time.

Spreadsheet on Cloud – Framework for Learning and Health Management System (SocialCalc)

K.S. Preeti, Vijit Singh, Sushant Bhatia , Ekansh Preet Singh, Manu Sheel Gupta

This describes the SocialCalc project, a shared spreadsheet for the XO Sugar operating system of the One Laptop Per Child (OLPC) project. The case study is a school administration system.

A Platform for Spreadsheet Composition (DISCOM)

Pierpaolo Baglietto, Martino Fornasa, Simone Mangiante, Massimo Maresca, Andrea Parodi, Michele Stecca

This consists of a client plug-in which manages the automatic export and import of data from a shared platform DISCOM. They describe it using a car dealership example. I would welcome a more detail comparison with the MS Sharepoint system.

Beyond The Desktop Spreadsheet

Gordon Guthrie, Stephen McCrory

They provide a web-based platform with a hierarchical view of data. This can be contrasted with the approaches of the other two presentations of SocialCalc and DISCOM.

Spreadsheets on the Move: An Evaluation of Mobile Spreadsheets

Derek Flood, Rachel Harrison, Kevin McDaid

The first limitation is the screen, the second the input method; it may take up to 4 keystrokes to enter a single digit or equals sign.

An Insight into Spreadsheet User Behaviour through an Analysis of EuSpRIG Website Statistics

Grenville J. Croll

In 2010 there were about 35,000 unique visitors to eusprig.org. The traffic trends to the site mirror other trends before and after a financial collapse. A few thousand people worldwide have a qualified interest in the integrity and quality of spreadsheets.

Living with spreadsheets

Dean Buckner

Spreadsheets are the tool of choice for linking separated corporate systems. (Think Tower of Babel, Stovepipe solutions, Islands of Automation, EAI). The FSA found one 'awesome' system with 9,000 worksheets. Date, source, and type of data is completely opaque, the nature and location of transformations completely unclear. The solution is left as an exercise for the reader.

Estimating Spreadsheet Review Workload in Project Finance

David Colver

He analysed a large number of projects and concluded that the ease of reviewing spreadsheets, and re-reviewing the corrections the client makes, is related to the skill and cooperative attitude of the client rather than to any obvious static spreadsheet metrics.

This is relevant to a number of researchers who are trying to measure quality in spreadsheets by static (structural) inspection techniques. Presumably this is so that some indication can be derived of the workload needed to remediate them. But David's point is that there is no relationship – it depends on the ability of the author to do what the professional concludes has to be done. This relationship will be familiar to professional software testers as the defect injection rate versus the detection and removal efficiency.

This was a rare opportunity to hear a long-experienced practitioner of both model building and auditing open the kimono on their in-house activity data on work on significantly sized spreadsheet projects. His slides are not publicly available and if we do webcast EuSpRIG conferences in future, this may be the kind of one for which we turn off the camera. There was a 2004 paper from Mercer consultants on the same topic "Financial Modelling of Project Financing Transactions," which is now removed from the web, email me for my archive copy if you're interested.

Concluding remarks

I'm pleased with the level of debate and discussion this year. Our task is to make all these good ideas more widely known. I wish researchers would make more of an effort to contact members of the EuSpRIG committee before embarking on a year's work, which could be accelerated by pointing them to research already done and giving an opinion on how realistic their proposed approach is in the real world. Not that that is definitive – it's always satisfying in research to confound the 'experts'!

Pat O'Beirne July 2011