

Spreadsheets: 'spread' as in 'middle-age ...'

Spreadsheets are interesting applications. When the spreadsheet Supercalc appeared around 25 years ago, its effect was nothing short of astonishing. Here was an application which unlocked the computational secrets of the then fledgling computers in a way which anybody could use. No strange programming languages, no knowledge of arcane operating system interfaces. No nothing in fact. All you did to add a column of figures was put a little formula in a box somewhere =SUM(..) and this box in more or less real-time would tell you what the sum was even when you changed individual elements.

Within a very short period, people were beginning to do all kinds of exotic things with them and in every way, they seemed to be a killer application area every bit as important as the word processor.

The reason I am writing about them is that in many ways, this killer application is two-edged and I now look on spreadsheets as being a considerable threat. Let me give you some reasons.

First of all, as I have previously described in this column, people make mistakes with them. In fact according to some sources, people make lots of mistakes with them, with academics like Ray Panko of the University of Hawaii reporting that 90% of all spreadsheets submitted to them for analysis contained errors which affected the results by more than 5%. So this is my first beef. They are being programmed by people who don't understand programming principles. Checking them for errors is punishingly difficult with little tool support and it is common to find in a software company that the quality control lovingly applied to the software is completely bypassed by the spreadsheeters. Spreadsheets grow like weeds and seem to exist completely outside any formal engineering methodology.

Second, the mathematics of spreadsheets is decidedly dodgy. Dr. Alan Stevens kindly sent me some notes on this. Amongst many little gotchas, try typing in a couple of adjacent cells, $-x^2+1$ and $1-x^2$, where x^2 means x times x and x is some cell reference. Now put any number you like in that cell and you will find that $-x^2+1$ is not equal to $1-x^2$. Oh goody. Not only that but of the four spreadsheets I tried with $x=23$, I got the following pairs (530, -528), (true, true), (-22,-528) and (530,-528). Somebody used to a programming language would expect them to be the same.

Third, and now this is why I really don't like them. People store important data in them. I have just been involved in an exercise in my university to gather together data for our preparations for the upcoming country-wide Research Assessment Exercise about which the less said the better. For this we need data on various aspects of research. In most institutions like ours, the data is spread all over the place usually in spreadsheets with different formats, columns, incorrect formulae and so on. Extracting it is a nightmare. The problem of course is that people very widely use spreadsheets as databases but without any of the normal intellectual effort spent in data normalisation and so on. The result is utter data chaos. You might as well store your data in a black hole.

lesh@leshatton.org