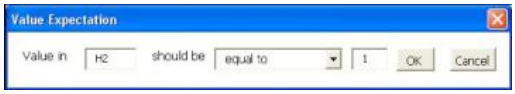
Roland Piquepaille's Technology Trends
How new technologies are modifying our way of life

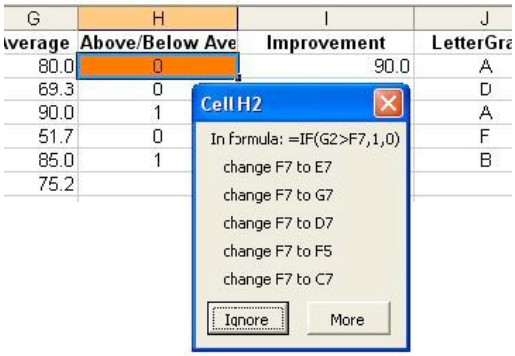
mardi 29 mai 2007

Fixing spreadsheet errors

Admit it: when you look at a spreadsheet, you trust the results. You don't question the totals or the percentages you see. But data, and formulas used to compute what appear in the spreadsheet cells, are both prone to errors. According to computer scientists from Oregon State University (OSU), 90% of the 100 million spreadsheets produced each year in the U.S. alone, contain non-trivial errors. This is why they've developed a semi-automatic debugger for spreadsheet systems. They say that this new approach could save billions of dollars annually. As they've licensed their software to a new company in Oregon, this technology might be one day included in the spreadsheet tools we use.



Their debugger is called GoalDebug -- short for "Goal Directed Debugging of Spreadsheets." With GoalDebug, a user can express expectations about cell values using a simple interface, as shown in the figure above.



As writes one of the researchers, "the system uses this information to compare the expected data with the data computed by formulas in the spreadsheet and creates a ranked list of suggestions for how to change cell formulas." (Credit for images and captions: Martin Erwig, OSU)

This semi-automatic debugger for spreadsheet systems has been developed by Martin Erwig, an associate professor of computer science in the OSU College of Engineering. Here are two links to the GoalDebug project and to a 2005 technical paper about it, "Goal-Directed Debugging of Spreadsheets" (PDF format, 8 pages, 1.25 MB). You'll find more explanations about the above illustrations in this paper.

Now, let's return to the OSU news release to discover the extent of the problems of spreadsheet errors.

> Spreadsheets, a standard bookkeeping and accounting tool used by businesses to track everything from payroll to accounts receivable, are one of the most common of all computer software programs. In the United States it has been estimated that 11 million people create about 100 million spreadsheets a year, which in turn might be managed by up to 60 million users. But they are notoriously prone to errors, experts say.

"Most users of spreadsheets are overconfident, they believe that the data is correct," said Erwig. "But it has

been observed that up to 90 percent of the spreadsheets being used have non-trivial errors in them. In fact, one auditor has said he never inspected a single spreadsheet during his entire career that was completely accurate."

So what to do about these mistakes? Erwig's answer was to build a semi-automatic debugger. But how does it perform?

> What the new OSU systems do is try to identify the ways that humans commonly make mistakes and then suggest what the correct approach might have been. For instance, if someone sees a figure in a spreadsheet that seems suspicious or is clearly incorrect, they can plug in the correct number, and the OSU system can suggest several programming mistakes that might have created the error – which the user can then sort through and use to identify the problem. A study performed by Robin Abraham, a recent OSU doctoral graduate, has shown that in 80 percent of the cases, the needed change is among the top five suggestions produced by the system, and in 72 percent of the cases it is among the top two suggestions.

The latest work on GoalDebug was presented at the 29th ACM/IEEE International Conference on Software Engineering (ICSE 2007) which took place in Minneapolis, MN, from May 19 to May 27, 2007. Robin Abraham and Martin Erwig presented a paper named "GoalDebug: A Spreadsheet Debugger for End Users." Here are two links to the abstract, from IEEE (not always available), and from OSU.

A question remains: will this technology be bought by Microsoft or be used by OpenOffice developers?

Sources: Oregon State University news release, May 24, 2007; and various websites

You'll find related stories by following the links below.

- Computers
- Economy
- IT
- Software

6:31:02 PM Permalink

Last update: 01/06/2007; 18:15:02.