

Powerful new software plug-in detects bugs in spreadsheets

October 24 2014



Doctoral student Dan Barowy's CheckCell plug-in automatically finds data errors in spreadsheets. He, Dimitar Gochev and their advisor Emery Berger at UMass Amherst developed it as a plugin for Microsoft's Excel program. Credit: UMass Amherst

An effective new data-debugging software tool dubbed "CheckCell" was

released to the public this week in a presentation by University of Massachusetts Amherst computer science doctoral student Daniel Barowy. He spoke at the premier international computer programming language design conference known as OOPSLA, in Portland, Ore.

CheckCell, which automatically finds data [errors](#) in spreadsheets, was developed as a plugin for Microsoft's popular Excel program. Its release at the highly respected Object-Oriented Programming, Systems, Languages and Applications (OOPSLA) conference this week signals that it is now freely available to anyone who wants to use it.

Spreadsheet data errors can be consequential, Barowy says. "Consider the case of a paper written by Harvard economists Carmen Reinhart and Kenneth Rogoff a couple of years ago. The paper was influential, lending credibility to government austerity measures in Europe and the United States. But in 2013, UMass Amherst economist Thomas Herndon and colleagues found, in combing through the data by hand, that methodological errors undermined Reinhart and Rogoff's argument. In particular, Reinhart and Rogoff exaggerated the impact of key data values in a spreadsheet."

The CheckCell group wondered whether software might be developed to find these kinds of errors automatically. The answer is a definite yes, says UMass Amherst School of Computer Science professor Emery Berger, Barowy's advisor. CheckCell successfully found a number of the same errors as had Herndon.

Berger explains, "Our work for the first time combines data analysis and program analysis. Poor quality data costs everyone money. CheckCell helps users avoid costly mistakes."

He adds, "Basically, CheckCell identifies data points that have a big impact on the final result, even if the impact is super subtle and difficult

to detect. CheckCell immediately flags data points that are very suspicious, the ones that deserve a second look. It's like having a helper who says, 'pay attention to these cells, they really matter.'"

For example, if a teacher has an "A" student who would be expected to get a 94 on a test and the spreadsheet says that student got a 49, CheckCell will flag it, the computer scientist says. "It tells you that you need to make sure this value is correct."

To develop CheckCell, Berger and graduate students Barowy and Dimitar Gochev used a combination of statistical analysis and data flow analysis to flag inputs that have an unusual impact on the program's output. They evaluated the procedure against a collection of real-world spreadsheets such as budgets and student grades. They introduced common errors into the spreadsheets, then asked the plug-in tool to find them.

The technique uses what Berger calls "a threshold of unusualness." CheckCell marks hidden, high-impact data points in red and asks the [spreadsheet](#) designer to check them. If they are indeed correct, they turn green and will not be flagged in subsequent analyses, he notes.

In the future, his team, working with UMass Amherst [computer science](#) colleague Alexandra Meliou, plan to extend CheckCell's use to large-scale data sets.

More information: Download the plug-in free at: www.CheckCell.org

Provided by University of Massachusetts Amherst

Citation: Powerful new software plug-in detects bugs in spreadsheets (2014, October 24)

retrieved 23 April 2024 from

<https://phys.org/news/2014-10-powerful-software-plug-in-bugs-spreadsheets.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.