Expert and Novice End-User Spreadsheet Debugging: A Comparative Study of Performance and Behaviour

Brian Bishop, Dundalk Institute of Technology, Ireland
Kevin McDaid, Dundalk Institute of Technology, Ireland

ABSTRACT

The reliability of end-user developed spreadsheets is poor. Research studies find that 94% of ‘real-world’ spreadsheets contain errors. Although some research has been conducted in the area of spreadsheet testing, little is known about the behaviour or processes of individuals during the debugging task. In this paper, the authors investigate the performance and behaviour of expert and novice end-users in the debugging of an experimental spreadsheet. To achieve this aim, a spreadsheet debugging experiment was conducted, with professional and student participants requested to debug a spreadsheet seeded with errors. The work utilises a novel approach for acquiring experimental data through the unobtrusive recording of participants’ actions using a custom built VBA tool. Based on findings from the experiment, a debugging tool is developed, and its effects on debugging performance are investigated.

Keywords: Auditing, Debugging, Electronic Spreadsheets, End-User Programming, User-Centered Design

INTRODUCTION

Spreadsheet Prevalence

Commercial off-the-shelf spreadsheet programs are one of the most popular end-user programming environments in use today (Burnett, Cook, & Rothermel, 2004), and have become ubiquitous, and in many cases indispensable software tools throughout industry and within all levels of the business world. Some of the reasons spreadsheets are so popular in end-user computing is that they are easy to develop and modify, they are highly scalable, and they support numeric data analysis without the specific need to use a ‘behind the scenes’ programming environment. Spreadsheet programs are also extremely flexible in terms of size and complexity: the 2007 version of Microsoft Excel can support over 1 million rows and 16 thousand columns per worksheet, with a maximum formula length of 8 thousand characters per cell.

It has been estimated that by 2012 there will be 90 million end-users in the U.S. alone and of these 13 million will be end-user programmers; which is significantly higher than the estimated...
Worryingly, the reliability of end-user developed spreadsheets has been shown to be poor following empirical and anecdotal evidence collected from studies investigating operational (real world) spreadsheet error rates, most of which are detailed in previous studies (Panko, 1998; Panko, 2000; Powell, Baker, & Lawson, 2008). The most recent study that investigated spreadsheet error rates was conducted by Powell, Baker, and Lawson (2007), who reported that of the 50 real-world operational spreadsheets they audited, 94% contained errors. Due to the nature of spreadsheet use, it is the case that when failures do occur the results can be quite significant. Many real-world examples of the impact of errors in spreadsheets are published on the European Spreadsheet Risks Interest Group website (www.eusprig.org). The following is just a small selection of these reported errors:

- A government housing authority had to pay $216,352 to cover expenses incurred as a result of a spreadsheet data-entry error that overpaid landlords.
- A spreadsheet error caused the Nevada City 2006 budget to show a deficit of $5 million for a water fund.
- In 2005, a Kodak employee was mistakenly accrued an $11 million severance. A Kodak spokesman said the severance error was traced to a faulty spreadsheet, and the money was not paid out.
- Stock prices for an online retailer tumbled by 25% in one day, and the Chief Financial Officer resigned due to an error in a spreadsheet formula.
- A finance company was undercharged $25,652 for rent in 1999 due to an error in a spreadsheet formula.
- A Legislative Auditor found a government department’s key performance indicator value overstated by $58 million (45%). The miscalculation was due to a spreadsheet formula error.

The reported and published spreadsheet errors are just the tip of the iceberg. Very few organisations report embarrassing spreadsheet errors and their impacts, and details of these errors only enter the public domain when it becomes unavoidable. From correspondence with professionals that cannot be published, there appears to be many instances of such errors throughout industry that remain behind closed doors.

Spreadsheet Debugging

Relatively little research has been conducted on the debugging and error detection process for spreadsheets. The emphasis of spreadsheet research available has been on the prevention of spreadsheet errors through spreadsheet design and testing methodologies. The notable
An Efficient and Effective Approach to Developing Engineering E-Training Courses
www.igi-global.com/chapter/efficient-effective-approach-developing-engineering/163867?camid=4v1a

A Generic Context Interpreter for Pervasive Context-Aware Systems
www.igi-global.com/chapter/generic-context-interpreter-pervasive-context/73219?camid=4v1a