A study of help requested online by spreadsheet users

ABSTRACT

The software used by organizations can have a substantial effect on employees' productivity. One of the dominant kinds of software in the workplace today is the spreadsheet, which has proven useful for performing numeric computations as well as for organizing, manipulating, and visualizing data. In this paper, we present an analysis of the problems reported online by users of different spreadsheet editors, in order to guide the selection and enhancement of those editors. We found that most requests for help were related to operations that we refer to as "foundational," such as with configuring the editor and setting up real-world problems in the form of spreadsheets, rather than with using particular formulas or features. We identified a number of key similarities and differences in questions about different spreadsheet editors and versions of Excel, ultimately leading to several challenging opportunities for future research.

Keywords: End-User development, End-user programming, Spreadsheets, Information technology, Information systems
INTRODUCTION

Spreadsheets play an increasingly important role in the operations of companies and other organizations. Projections indicate that by 2012, over 50 million people in American workplaces will create spreadsheets or databases as part of their everyday work (Scaffidi et al., 2005). Worldwide, over 500 million people use spreadsheets (Microsoft, 2010). This popularity stems from spreadsheets' usefulness for a broad range of tasks that include organizing, analyzing, exploring, and visualizing data (Nardi, 1993). In prior studies, we have also found that spreadsheets are becoming a crucial resource for storing and manipulating data that information workers retrieve from web sites (Scaffidi et al., 2007).

With the rising importance of spreadsheets in the organizational context, it is vital to examine the problems reported when people use editors such as Microsoft Excel to create and modify spreadsheets. One reason is that the acceptance and adoption of technology is strongly controlled by its perceived usefulness, as well as its perceived ease of use (Davis, 1989; Ma and Liu, 2004). Therefore, understanding the relative strengths and weaknesses of different spreadsheet editors is likely to be useful for guiding organizations as they select a tool for their use. In addition, identifying common questions can provide researchers with insights about where spreadsheet editors as a whole need improvements, in order to facilitate smoother and more productive adoption.

Moreover, it is important to characterize the problems that people ask for help with because these problems directly affect the support that organizations will need to provide to users. In particular, an assessment of training needs—including an identification of problems commonly reported by users—is a necessary precursor to implementation and testing of training materials (Bostrom et al., 1998; Oinas-Kukkonen et al., 2010). Organizations typically respond
to requests for help through an information technology department or through informal peer support by users who are often called “local developers” or “gardeners” (Nardi, 1993; Morch and Asand, 2006). These IT departments and gardeners invest time in learning how to overcome problems that people commonly ask for help with solving, in order that they can provide support to problems as needed without having to discover solutions during a crisis. Making an informed decision about this investment of time requires knowing what kinds of problems users commonly ask for help with solving.

To date, empirical studies have focused on only a narrow range of problems reported by spreadsheet users. In particular, despite the fact that spreadsheet editors have progressed far beyond simple calculation engines, formula errors are the only spreadsheet user problem that has received extensive investigation. Studies have shown that formula errors are common (Panko, 1998; Powell et al., 2009) and can lead to monetary loss (Caulkins et al., 2007). In response, researchers have proposed a variety of approaches for improving the process by which spreadsheets are created and tested (e.g., Panko, 1999; Rothermel et al., 2002; Panko, 2007). Yet formula errors are only one of many conceivable problems that users might be facing, and it is not at all clear that they remain the most important category of problem from the standpoint of designing training materials and support systems.

In response to this gap in our understanding of common requests for help, we conducted a preliminary study in prior work, where we investigated questions that users posted to an online forum about Microsoft Excel (Chambers and Scaffidi, 2010). We identified seven common categories of problems and noted that only 18% of posted user questions were about formulas.

The current paper presents a follow-up study where we extend our qualitative analysis of the forum for Microsoft Excel and also analyze data from forums for two additional spreadsheet
editors, Apple Numbers and OpenOffice Calc. Broadening our analysis enables us to compare posted problems among these three tools, and it also allows us to compare posted problems related to different versions of Excel. We use this new data set to answer three new questions:

Q1. What spreadsheet editor problems do people commonly ask online for help with solving?
Q2. How do posted problems vary among different spreadsheet editors?
Q3. Have users' posted problems with Excel evolved over time?

Overall, our study has shown that while there were differences among the problems posted to the three forums, there several topical areas were common among the editors. In particular, questions about what we refer to as “foundational” operations were widespread. These problems included installing and configuring the editor, setting up a problem as a spreadsheet (i.e., conceptualizing a real-world problem in terms of a spreadsheet-programming problem), and finding appropriate features in the editor for solving the spreadsheet. We call these “foundational” because users must perform these operations before they can dig into the minutia of how to use specific spreadsheet features (e.g., before determining what specific parameters to pass into a particular function, or how to recolor the bars of a bar chart). In other words, regardless of spreadsheet editor, we found that people asked for help with specific features less often than they asked for help with installing, reasoning with, and finding their way around spreadsheet editors.

We found several differences among tools and versions of Excel. Compared to the other forums, the Excel forum had a relatively high proportion of questions about macros and formulas. The Excel forum also had a high proportion of answers that mentioned macros. In
contrast, the other two forums had a relatively high proportion of questions about tool configuration. The Excel forum not only had a lower proportion of configuration questions, but this proportion has decreased over the past few years.

The remainder of this paper is organized as follows. First, we summarize related work, including the numerous prior studies of errors in spreadsheet formulas. We next describe our classification scheme for user comments, which we developed as part of our prior work. We then present our current study, involving our analyses of forum comments to answer the research questions above. We close with a discussion of how our results might be interpreted and used to drive future research.

RELATED WORK

Prior empirical studies of spreadsheet problems have largely focused on formula errors. These studies have shown that spreadsheets are frequently "built in an informal, iterative manner" that typically omits quality control (Cragg and King, 1993). Consequently, up to 94% of spreadsheets typically contain at least one formula error (Panko, 1998), despite the fact that users are highly confident in the correctness of their spreadsheets (Panko, 2007). Formula errors come in many different varieties (Powell et al., 2009; Rajalingham et al., 2001) that ultimately can lead to monetary losses (Caulkins et al., 2007).

Yet surveys consistently show that organizations have come to depend on spreadsheets for far more than just formula-computation. One survey found that other features such as charts and macros were used just as often as several varieties of functions (financial, database, and goal-seeking), with only statistical functions seeing more use than charts or macros (Chan and Storcy, 1996). A second survey found that 70% of spreadsheets were used for laying out reports,
with only 30% for calculation-intensive activities like modeling or analysis (Hall, 1996). A third survey found that although formulas were used by 90% of information workers or their subordinates, other features were also frequently used, such as charts by 80% and macros by 42% (Scaffidi et al., 2006). Finally, a fourth survey found that charts were used more frequently than every function except the “if” function (Lawson et al., 2009).

Given the importance of other spreadsheet features in addition to formula computation, we recently conducted a preliminary study investigating questions that users posted to an online forum about Microsoft Excel (Chambers and Scaffidi, 2010). We identified seven common categories of user questions and noted that only 18% of posted questions about Excel were about formulas. While this study offered an important insight into problems posted by users, it focused on only one specific spreadsheet editor rather than spreadsheet editors in general. Moreover, it included user comments from a lengthy 5-year period of time, without any analysis of whether users of the most recent versions of Excel still ask for help with the same problems. Our current study addresses these limitations by analyzing more recent data describing posted user problems related to multiple spreadsheet editors and versions of Excel.

ACQUISITION AND CLASSIFICATION OF USER COMMENTS

To obtain data for qualitative analysis, we downloaded user comments and questions from three popular online forums for discussing Excel¹, Apple Numbers², and OpenOffice Calc³.

---

¹ http://www.excelforum.com
² http://discussions.apple.com/category.jspa?categoryId=202
We selected the 200 most-recently-modified discussion threads from each forum (or its subforums), yielding 600 forum threads for analysis. Approximately 97% of these discussions were initiated by users during the period of August-October 2010; the remaining 3% were initiated at an earlier point in time (with the most recent reply arriving during that period). These forums are intended for discussions among users of these editors. Developers of these tools hold discussions in separate venues (either internally to Microsoft and Apple, or on a separate website in the case of Open Office\(^4\)).

We categorized each thread according to the classification scheme created in our prior study (Chambers and Scaffidi, 2010). This classification scheme was originally created by applying a grounded theory approach (Glaser and Strauss, 1977) to analyze 200 randomly-selected threads from the Excel online forum over a five year time period from January 2005 to January 2010. The classification scheme covered seven categories of user problems (Table 1).

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem setup</td>
<td>Often the user will know what they want to do, but wants to know how to set the problem up</td>
</tr>
<tr>
<td>Feature finding</td>
<td>Problems dealing with finding specific functionality through the user interface of Excel</td>
</tr>
<tr>
<td>Configuration</td>
<td>Problems dealing with use of Excel’s settings, such as default colors or startup toolbars</td>
</tr>
<tr>
<td>Formulas</td>
<td>Problems dealing with use of formulas, such as asking about specific formulas or formula debugging.</td>
</tr>
<tr>
<td>Macro</td>
<td>Problems dealing with use of macros, including writing, debugging and invoking</td>
</tr>
<tr>
<td>Integration</td>
<td>Problems dealing with use of features to integrate Excel with other programs, such as data import</td>
</tr>
<tr>
<td>Data visualization</td>
<td>Problems dealing with use of advanced data visualization features, such as pivot tables or charting</td>
</tr>
</tbody>
</table>

\(^3\)http://www.oooforum.org/forum/viewforum.phtml?f=3

\(^4\)http://development.openoffice.org
The *Problem setup* category covers difficulties before users even know what features are needed and, in general, are requests for help designing a spreadsheet to solve a particular problem. For example, in one forum post from our earlier study, a user stated, “I would like to develop a spreadsheet to forecast the evolution of the corresponding accounts.” The *Feature finding* category deals with locating features in the user interface. For instance, one user requested, “How can I keep the first line on the screen when I scroll down?” The *Configuration* category deals with requests to reconfigure the way the editor behaves in general (i.e., not in the context of a specific spreadsheet). In our prior study, these included settings requests such as “How can I change the default fill color for Excel?” The last four categories relate to actually using certain kinds of features for a particular spreadsheet. The *Data visualization* category includes questions about charts and filtering pivot tables. The *Formula* category deals with questions about how to use a specific formula, such as “I want to create an *if* formula that will determine whether an employee is hourly or salary and then compute pay.” Finally, *Macro* questions are concerned with programmatic macro scripts.

Overall, these seven categories can be viewed as representing a general *progression* through a typical programming task. The first category deals with mapping the problem to the conceptual structure afforded by a spreadsheet; once a general plan of attack has been identified, the second category deals with finding the necessary tools (features) within the user interface. Upon trying those tools, or failing to find those tools, the user might need to configure the editor (our third category); due to its global effects, this configuration might also be done prior to the start of a particular spreadsheet problem (i.e., before the set up of a later problem). Finally, within the context of using a particular feature to create or modify a spreadsheet, users have specific and often detailed questions along the way about how to use the particular features that
they have found. These final questions fall into our last four categories. If users cannot progress past problems in the first three categories, then they might not get to the point where they encounter problems in these last four categories. In this sense, we will refer to the first three categories as covering “foundational” spreadsheet operations.

Each discussion thread from our three forums was classified by one of us using the classification scheme above. To verify that this classification was correct and consistent, a second researcher independently coded a random subset of 75 threads (25 from each forum). The overall agreement was 92%, with a Cohen’s Kappa of 90%. These were comparable to the 92% and 89%, respectively, that we had previously obtained on our initial data set when we originally developed this classification scheme in prior work (Chambers and Scaffidi, 2010). Thus, we judged that the classification scheme was reliable and an effective way to qualitatively analyze the questions that users are asking about spreadsheet editors.

RESULTS

Having classified the threads, we analyzed the data to answer our three research questions. Below, we first present our analysis of the threads from all three forums, in order to assess what problems these spreadsheet editors have in common. We then compare across spreadsheet editors and across different versions of the Microsoft Excel editor.

What spreadsheet editor problems do people commonly ask online for help with solving?

Our first analysis combined the data from all three forums. After applying our classification scheme to the 600 threads from the three spreadsheet forums, we found that 69% of questions were related to the three foundational categories of Problem setup, Feature finding, and Configuration (Figure 1). The other four feature-specific categories of Formulas, Macro,
Integration, and Data visualization accounted for the remaining 31% of the discussions. In other words, people asked for help with specific features less often than they asked for help with installing, reasoning with, and finding their way around spreadsheet editors.

Figure 1. Distribution of recent discussions (aggregating all three spreadsheet forums)

The most prolific category, Problem setup, accounted for 38% of the forum discussions (231 threads). After reviewing discussions in this category, we found that it could be further decomposed into a few identifiable subgroups: questions about setting up a spreadsheet structure, questions about modifying sections of an existing spreadsheet structure, and questions about how to implement a specific cell.

Among the 231 Problem setup threads, 25% (58 threads) were initiated by users who were struggling with how to design a spreadsheet structure appropriate for their problem. In most cases, users apparently had not yet made any progress with creating a spreadsheet for their problem. For example, one user explained, “I want to input my occupancy levels from the last
two years so I can judge whether one room in particular is 'pulling it's[sic] weight.' In order to be able to turn the raw data into charts what would be the best way to do this?" The answer to such a question includes not just how to input data and create charts, but also how to structure a spreadsheet to address the problem.

The majority of the Problem setup questions, 64% (147), dealt with situations where a user had already constructed some sort of spreadsheet but was struggling with how to modify or extend large sections of this structure in order to solve a particular problem. We observed several recurring problems: 33 users wanted to perform string transformations or other automatic data conversion; 23 wanted to transfer data sets, or portions of data sets, between spreadsheets or worksheets; 17 wanted to perform conditional operations on data sets, such as computing conditional sums on some columns; 16 wanted to group, sort, or filter data sets; and 12 wanted to use a data set to drive a custom widget-based user interface (e.g., so another user could open the spreadsheet and use checkboxes to control which rows would be used to compute a total).

In the final 11% of Problem setup threads (26 threads), a user had a spreadsheet and was struggling with how to set up one formula to solve a problem within that existing structure. These questions involved coding single cells rather than modifying the overall structure. For example, one user wanted to create a formula for converting a numeric grade to a letter grade (e.g., "A-") based on numeric ranges. While these questions might seem similar to the Formula category (Figure 1), the distinction is that these Problem setup questions were about discovering some formula—any formula—to solve a problem, while the Formula category deals with questions about how to use specific formula functions (such as if or vlookup).
How do posted problems vary among different spreadsheet editors?

Our second analysis examined how the 200 threads from each forum were distributed. Among the forums, we found substantial differences (Figure 2), which were statistically significant ($\chi^2(12) = 103.93, P<0.001$). In particular, we found that the Microsoft Excel forum had the lowest proportion of foundational questions, at 55%. Over 70% of questions in the OpenOffice Calc forum fell into the three foundational categories, and over 80% of questions in the Apple Numbers forum fell into these categories.

Figure 2. Comparison of recent discussions across different forums

The Apple Numbers forum had the highest proportion of Problem setup questions. Many questions in Numbers were related to the specific way in which information is displayed in this editor’s user interface compared to the other two spreadsheet editors. For example, Numbers
allows users to create multiple small tables on the same page, rather than a single large grid in each page (or "sheet") as in the other editors. Consequently, some Numbers users asked how to allocate data and computation among different tables in a spreadsheet.

Another interesting difference between editors was the large percentage of Feature Finding questions about Numbers. We observed that 10 of the 46 questions in this category dealt with trying to find features in Numbers for a task that the user was previously able to do in Excel. In 8 of these 10 requests, the reply effectively was "This can't be done in Numbers." Some examples of these "missing features" are the ability to protect particular cells, the ability to add hyperlinks or URLs to cells, and the ability to print a selected portion of a spreadsheet. Similarly, many Calc users explicitly commented on difficulties with this editor compared to Microsoft Excel. For example, one user stated, "I miss being able to press <CTRL+> and have it copy the data from the cell above into the blank cell. The second thing I miss is being able to press <CTRL + Down Arrow> to jump to the first empty cell in a column, and <CTRL + Right Arrow> to jump the first blank cell in a row... likewise Left Arrow... etc. etc. Can these be created in OO?" This and other Feature Finding comments highlight that some users are expecting the performance and behavior of Calc to be similar to Excel.

We also observed a relatively high number of Configuration questions in Calc. Two particularly common topics were tool configuration to support printing and macros. For example, one poster asked, "I can't record a macro, because the option doesn't appear in the menu. Can someone help?" Other questions dealt with setting default grid styles, turning off auto correct, and configuring other default settings. For example, one user stated "For some reason my spreadsheet files are all opening as 'Read Only'. Every time I open a spreadsheet to make any
changes, I have to save a new file.” Fixing problems like these requires adjusting the software’s settings to correct the default settings.

One major difference across the three editors is that Numbers does not have the ability to include macros, which explains why there were no Macro questions in the Numbers forum. On the other hand, Macro questions were the third most common question type for Excel. In fact, for Excel, macros were given as a solution to 17 of the 62 Problem setup questions, many even without a macro being requested (which we did not see for Calc). Since Numbers does not support macros, we recompiled the statistics of the three forums, excluding all of the threads asking about macros and recomputed the percentages. This did not affect any of the differences discussed above in a substantial way. If anything, this change accentuated the differences, specifically in the Configuration and Formulas groups.

We observed a relatively high number of Integration questions about Calc compared to the other two editors. Very few of these threads involved importing files directly from Excel to Calc, but instead focused on problems such as trying to import other types of files into Calc. Users again seemed to have problems with things that Excel does automatically. For example, one user asked, “How can I import this csv?”, a task that is automatic in Excel (or, depending on the user’s configuration, nearly automatic and requiring only one click of a button in a wizard). One integration question that arose for Calc but not for the other two editors was how to integrate a user’s spreadsheets through code such as C#, of which there were three questions, compared to none in the Excel and Numbers forums.
Have users' posted problems with Excel evolved over time?

Our final analysis compared the data analyzed in our previous work (Chambers and Scaffidi, 2010) with the new results for Excel, in order to reveal how questions have changed over time. Our older data set of 200 threads was randomly sampled from a 5-year period leading up to the beginning of 2010, while the recent set of 200 threads is focused on late 2010. We found an increase of about 10% in the proportions of the Problem setup and Macro questions over time, as well as a decrease in the proportions of both Configuration and Integration questions (Figure 3). The remaining categories had approximately equal proportions in the two data sets. Overall, the two distributions were statistically different ($\chi^2(6) = 22.54, P<0.001$).

Figure 3. Comparison of recent discussions versus earlier discussions in the Excel forum
Between the two data sets, the proportion of Configuration and Integration questions each dropped by over half, indicating decreasing proportions of questions that dealt with setting up Excel and using it to load data. In fact, some of the configuration questions from the old analysis (over the past 5 years) were directly solved by newer versions of Excel. For example, Excel 2007 introduced the ability to create themes, which allowed users to easily change the default colors. In the recent dataset, none of these questions were asked, despite being a frequent topic of discussion previously.

Meanwhile, the proportions of Formulas, Feature finding, and Data visualization questions have held relatively steady over time but shifted in character. For example, the earlier data included questions about sum and subtraction, whereas no questions were asked about these formulas in the more recent data. Instead, we saw many more recent questions touching on more specialized formulas such as sumif and quartile. A similar shift toward specialized features or requirements was apparent in recent Data visualization threads: of the recent 16 Data visualization threads, 5 dealt with ways to create dynamic charts, either by using a scroll bar to change the data, or through the use of formulas. We did not observe such questions in the earlier data set. Another discussion thread asked for help with creating a new type of graph, a 3D scatter plot, which has only become available in the most recent versions of Excel.

Finally, the proportion of Problem setup and Macro questions both increased by close to 10% between the two data sets. In addition to the rising proportion of macro questions being asked, macros also became increasingly common as answers to Problem setup questions. In the data spanning five years, only 7% of the responses to Problem setup questions contained a solution that included a macro. In contrast, 27% of the recent responses to Problem setup questions had a macro solution.
Overall, the results above depict a statistically significant shift away from Configuration and Integration questions toward Problem setup and Macro questions. This shift has occurred contemporaneously with changes within Microsoft Excel. In particular, Excel 2007 introduced the button-and-tab “ribbon” interface (in place of a maze of menus), support for custom themes (color schemes), improved data-import wizards, enhanced sorting, enhanced charting capabilities, support for conditional formatting, and function auto complete.

To investigate whether users’ posted questions demonstrated any differences depending on the version of Excel used, we split our new data set of 200 threads for further qualitative analysis based on what version of Excel each user had. (For every discussion thread, this version information can be explicitly provided by the user in a specific field.) We found that 83 threads were from users who had Excel 2007 or 2010, while 92 were from users who had Excel 2000, Excel 2002, or Excel 2003. (In addition, we found five threads with users of a Mac version of Excel, as well as 20 threads that did not specify a version. We discarded these 25 from further consideration.)

Comparing the 2000-2003 and 2007-2010 subsets of our newer data, we found no statistically significant differences in the two distributions ($\chi^2(6)=5.91$, $P=0.87$). In fact, most of the changes in proportions of user questions had occurred prior to the major tool changes in 2007 (Figure 4). Therefore, we cannot attribute the shift in the overall distribution of users’ questions to the wave of changes in Excel 2007, even if these tool changes might have somewhat affected a few particular categories (e.g., by eliminating some Integration questions through new data import wizards, or by introducing new Data visualization questions through the availability of 3D scatter plots).
Thus, the shift in posted questions is somewhat more accurately described as a shift over time rather than a shift between tool versions. In particular, posted Excel questions have moved away from Configuration and Integration and moved toward Problem setup and Macros.

Figure 4. Comparison of discussions about different versions of Excel

<table>
<thead>
<tr>
<th>Percent of threads in forum</th>
<th>Problem Setup</th>
<th>Feature Finding</th>
<th>Configuration</th>
<th>Formula</th>
<th>Macro</th>
<th>Integration</th>
<th>Data Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Older user data: 5 years of Excel questions</td>
<td>35%</td>
<td>15%</td>
<td>16%</td>
<td>18%</td>
<td>20%</td>
<td>8%</td>
<td>4%</td>
</tr>
<tr>
<td>Recent user questions about Excel 2000-2003</td>
<td>30%</td>
<td>15%</td>
<td>16%</td>
<td>18%</td>
<td>20%</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>Recent user questions about Excel 2007-2010</td>
<td>21%</td>
<td>19%</td>
<td>23%</td>
<td>22%</td>
<td>22%</td>
<td>8%</td>
<td>9%</td>
</tr>
</tbody>
</table>

DISCUSSION

Our analysis of online requests for help with spreadsheet problems has enabled us to answer our research questions as well as to identify opportunities for future research.

Conclusions

Our first research question was What spreadsheet editor problems do people commonly ask online for help with solving? The most common requests were related to foundational
operations rather than with how to use specific features in the spreadsheet editors. These foundational operations included configuring the editor, mapping a real-world problem to the spreadsheet domain, and navigating around the editor to find features (rather than understanding how to use particular features). Overall, the most common kinds of questions had to do with setting up a problem as a spreadsheet.

Our second question was How do posted problems vary among different spreadsheet editors? The Excel forum had a relatively low proportion of foundational questions and a relatively high proportion of questions about macros and formulas. Yet we observed that many of the answers to our first question, above, did not vary across editors: in each forum, foundational threads outnumbered non-foundational threads, and questions about problem setup were the most common.

Our final research question was Have users’ posted problems with Excel evolved over time? The proportions of requests for help with configuration and integration problems have decreased, while the proportions have increased for requests concerning problem setup and macros. These changes are somewhat better described as a shift over time rather than a difference across versions of Excel.

Potential threats to validity

The primary potential threat to validity is that our research is based on data from forum postings, which might reflect all problems that spreadsheet users encounter. In particular, spreadsheet users might encounter certain problems that they never post in online forums, either because they never ask for assistance or because they find adequate assistance through another means. Some kinds of users might be more or less likely to make online requests for help (e.g.,
government workers versus home users versus corporate users). A secondary, related threat to validity is that we only investigated forum posts that were written in English. It is likely that members of different cultures have different attitudes about when it is appropriate to ask for help, and they might have different levels of access to online forums. Further research could investigate whether various groups of spreadsheet users encounter different problems, particularly problems that do not result in online requests for help.

Possible avenues for future work

Future work could be aimed at helping users to solve the kinds of problems uncovered by our analysis of online forum posts (regardless of, or perhaps in addition to, any additional problems that they might be encountering but not asking online for help with solving). Below, we offer a few ideas on some possible avenues for further research.

First, future work could investigate how to help users with setting up problems as spreadsheets, which was the most common kind of request for all three spreadsheet editors. For example, studies could explore the specific types of spreadsheets that users are trying to create and see if any patterns exist that could be exploited to help users to set up their spreadsheets for particular kinds of problems. One possible approach for helping users with Problem setup problems might begin with classifying or grouping similar problems, from which it might be possible to design templates that could be used to answer typical user questions. Such templates might need to be organization-specific, so researchers might aim to provide methods and tools so custom templates could be created and disseminated by “gardeners,” who are users (typically in large organizations) whose job description includes providing programming materials and assistance to other users (Nardi, 1993).
Second, users need more help with Feature finding. We have previously speculated that part of the Feature finding challenge is that users are overwhelmed with too many choices, and we suggested that children might be able to learn spreadsheet programming more easily if some features were temporarily turned off (Chambers and Scaffidi, 2010). However, while this approach might be appropriate for children, turning off features would be untenable in an organizational or business context. The main reason is that by virtue of lacking certain features, Numbers and Calc essentially do have features “turned off,” and yet we found in our analysis above that this caused some forum users to ask for the missing features. The implication is that while the number of features is indeed overwhelming, the solution must be to help business users to cope with the complexity rather than to hide from it.

As the number of spreadsheet features has increased, this problem has actually come to somewhat resemble the challenge faced by professional programmers when they must sift through thousands or tens of thousands of application programming interfaces (APIs) to find the ones appropriate for a particular programming task (Myers et al., 2010). To help these programmers to find the API functions that they need, sophisticated new search tools have been invented by researchers (e.g., Mandelin et al., 2005; Stylus and Myers, 2006; Brandt et al., 2010). Many of these approaches are directed at textual languages (such as Java), and these could be directly applied to spreadsheet formula languages. Research could also aim to adapt these approaches in order to help users with finding features in the graphical user interface (e.g., to help them with finding a particular chart or sorting wizard).

A third potential area of future research is suggested by the fact that thriving online forums exist to support users of all three major spreadsheet editors. Such forums effectively decentralize some gardener responsibilities outside of the organization. The delegation of
responsibility to a large decentralized community is sometimes called “crowd-sourcing,” and researchers are only beginning to understand how organizations can reliably harness such a community for productive ends, particularly in contexts where financial incentives affect user behavior (Howe, 2009; Horton and Chilton, 2010).

Further research could explore how it might be possible to design improved forums or other technology to aid decentralized gardening for end-user programmers. For example, users could form collectives on websites where they share expertise with one another for a fee or for the purpose of establishing a collective competitive advantage against mutual competitors. As inspiration for designing new crowd-sourcing social structures in the area of end-user programming, researchers could look to existing web sites where professional software engineers pay an annual per-person fee for the privilege of sharing and receiving programming advice (e.g., Experts Exchange\textsuperscript{5}). This could help users to obtain answers in the heat of a crisis more quickly than is possible through the “volunteer” forums that we have examined, where questions typically go for days or even weeks before they are answered (if anyone ever responds at all). Therefore, providing more effective social support might enable people to use spreadsheets even more productively in the pursuit of their organizational goals.

\textsuperscript{5} http://www.experts-exchange.com
REFERENCES


