Spreadsheet Based Software Engineering

Awais Azam[1] Khubaib Amjad Alam[2]

Abstract

Spreadsheets play a vital role in data processing and reporting procedures of any organization. That is why spreadsheet programming is the most successful end-user programming. Keeping in mind the similarity between spreadsheet programs and traditional programs the techniques that can be applied to traditional programs can also be applied in context of spreadsheets. The main objectives of this research are: (1) to classify spreadsheet research papers according to three criteria: techniques used, datasets used, publication channels and trends; and (2) to analyze these studies from four perspectives: study objectives, methods, method accuracy and limitations of the study. We perform a systematic mapping study on spreadsheet studies published in the period 2013-2018, collected from automated four electronic databases. We identified a total of 44 studies published between 2013 and 2018 and classified them on predefined classification criteria. Based on the findings of this research, it is concluded that Smell Detection is the technique that is applied in most of the cases on spreadsheets. The year 2016 receives the highest number of publications on spreadsheets. The most used dataset is from the corporation called EUSES.

Keyword: Software Engineering, Systematic Mapping, Spreadsheets, Spreadsheet Programming

1 Introduction

Research in software maintenance has shown that many programs contain a significant amount of duplicated (cloned) code. Such cloned code is considered harmful for two reasons: (1) multiple, possibly unnecessary, duplicates of code increase maintenance costs and, (2) inconsistent changes to cloned code can create faults, hence, lead to incorrect program behavior [1].

Spreadsheets have been widely used for various business tasks, including data management, decision support, financial reporting and so on. It is estimated that 90% of desktops have Excel installed [2] and there were over 55 million users in the United State working with spreadsheets in 2012 [3]. It is also believed that the number of spreadsheet programmers is bigger than that of software programmers [3]. In most of the cases, the people who are responsible for developing and maintaining spreadsheets are end users and they are not familiar with software development practices, so errors are easily induced into spreadsheets during maintenance and updates [4].

There is no mapping study applied on spreadsheets previously that demands from the research community to provide an overview of the trends and techniques that are being followed by researchers and are suitable to be applied on spreadsheets. Another problem in the previous studies is that each study focuses on some specific method or technique and does not cover the entire domain. The main goals and purpose of this research is to introduce a broader

[1] *National University of Computer & Emerging Sciences, Islamabad | awaisazam08@gmail.com*
[2] *National University of Computer & Emerging Sciences, Islamabad | khubaib.amjad@nu.edu.pk*

and precise overview of almost all the most commonly used latest techniques on spreadsheets in the form of a Systematic Mapping. We followed the guidelines of Petersen [4].

The organization of the paper is as follows: The detailed steps and the structured strategy of Systematic Mapping is described in section 2. Section 3 contains the presentation and discussion of results. Finally, the conclusion and future directions are given in section 4. Section 5 contains the references.

2 Systematic Mapping

A systematic mapping was conducted by following the guidelines of Petersen [4] and the collected data is analyzed in an unbiased and structured fashion. The first and the basic step to start the process of systematic mapping was the formulation of protocol that was designed and structured by Awais Azam and reviewed by Dr. Khubaib Amjad Alam. Now the steps performed in systematic mapping are described in the next sections.

A Research Questions

The research questions that were formed to escort mapping study are shown in Table 1. The table shows the research question and the motivation behind these research questions in order to clarify what was the main reason to include that research question in this research process.

Table 1: Research Question

RQ #	Research Question	Motivation
RQ 1	What Software Engineering methods have been applied on spreadsheets?	To identify the areas which are under consideration by the research community in case of spreadsheets.
RQ 2	What are the datasets and sources from where these datasets are collected?	The main focus of this question is to identify the datasets and their sources which will depict the overall trend in selecting the datasets for spreadsheets.
RQ 3	What is overall research productivity in the field of software engineering in the context of spreadsheets?	The purpose of this question is to give an idea of the overall research that is done and currently going on in this field.

B Search Strategy

An automated search was performed that consists of the following steps.

Data Sources: In order to answer the research questions, an automated search was performed on the previously constructed terms on the four electronic databases. Table 2 shows the database and the online link to that database.

Table 2: Electronic Databases

Database Name	Link
IEEE Xplore Digital Library	http://ieeexplore.ieee.org/
ACM Digital Library	http://dl.acm.org/
Science Direct	http://sciencedirect.com/
Springer	http://link.springer.com/

The studies that were the part of this research activity were from a time span of 2013 to 2018. The digital libraries that were considered are IEEE, ACM, Science Direct, Springer and Google Scholar databases based on title, abstract, and keywords.

Search Process: In order to make sure that we were not leaving any related study, a two-stage search process was adopted.

Initial search stage: Here, we used the proposed search terms to search for primary candidate studies in the four electronic databases. The retrieved papers were grouped together to form a set of candidate papers.

Secondary search stage: In this step, we reviewed all the studies retrieved after title-based search where we read the abstracts of the remaining studies and based on the abstract the studies which were not relevant were excluded and the studies that passed this search qualified for the full-text reading.

C Study Selection Procedure

This step was designed to get the most relevant studies which were retrieved from four electronic databases in order to answer the research questions. The selection procedure consists of the following basic steps:

- Initial records
- Title based records
- Abstract based records
- Full article-based records

As we progressed by following these steps the irrelevant studies kept on excluding in each step. Table 3 gives an overview of the number of studies that were selected at different stages from different databases.

Table 3: Study Selection Procedure

Database Name	Initial Records	Title Based	Abstract Based	Full Text
ACM	174	71	29	24
IEEE Explore	321	73	27	16
Science Direct	5	5	1	1
Springer	20	4	3	3

The study selection process can be visualized in detail and is represented in Table 3 that contains all the details of the number of studies at every stage and how much studies were excluded. The study selection process can be visualized in detail that is represented in Figure 1 that contains all the details of the number of studies at every stage and how much studies were excluded and the reason for exclusion and other related information can be found in Figure 1 that is constructed by following the PRISMA guidelines for systematic review.

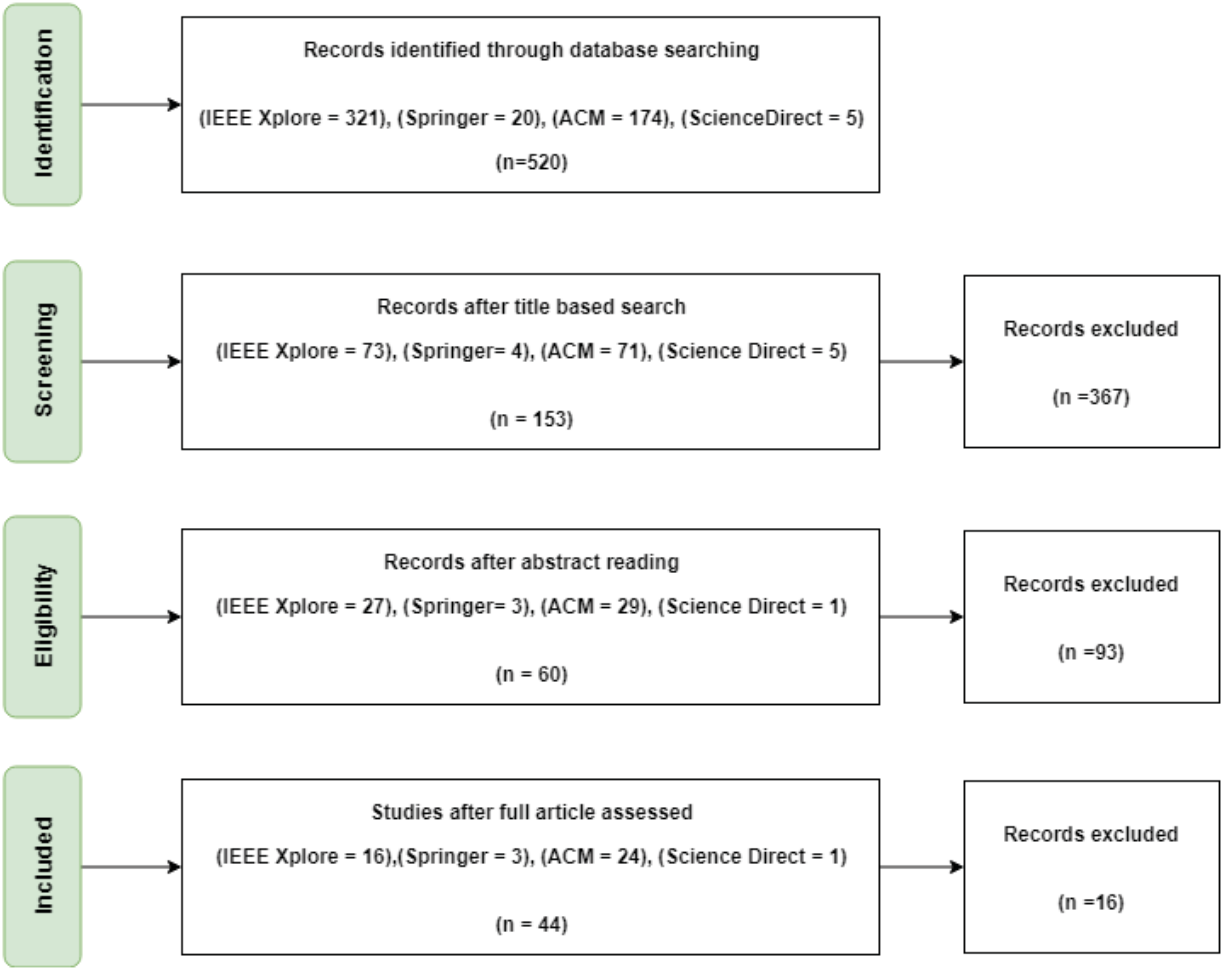

Figure 1: PRISMA Flow Diagram

Data Extraction: The full text of all the qualifying studies was analyzed and the relevant information was extracted to already defined data extraction form shown in Figure 2.

Data Extraction form

Study ID *(surname of first author and year first full report of study was published e.g. Smith 2001)*

Notes:

Year of publication					Notes
Title		Journal	Conference Proceeding	Other	
Publication Venue		☐	☐	☐	
Process of SE					
		Eligible	Ineligible	Unclear	
Conformance to Inclusion Criteria		☐	☐	☐	
		Fully	Partially	No	
Quality Ranking	QC1: Are study objectives clearly defined?	☐	☐	☐	
	QC2: Are applied methods well defined?	☐	☐	☐	
	QC3: Is accuracy of applied method measured and reported?	☐	☐	☐	
	QC4: Are limitations of study explicitly stated?	☐	☐	☐	

Dataset used:					
Metrics (If provided)					
Contribution	Method/ Technique	Tool	Framework	Model	Comparison
	☐	☐	☐	☐	☐
Research Approach	Solution	Validation	Evaluation		
	☐	☐	☐		
	INCLUDE ☐	EXCLUDE ☐			
Reason for exclusion					

Figure 2: Data Extraction Form

Inclusion and Exclusion Criteria: In this step, an inclusion and exclusion criteria were developed in order to further select the most related studies to carry out the research process. The inclusion and exclusion criteria are defined in Table 4.

Table 4: Inclusion and Exclusion Criteria

	Inclusion Criteria
IC1	Studies seeking convergence of software engineering in spreadsheets
IC2	Studies published in peer-reviewed conferences or journals
IC3	Studies published in or after 2013
IC4	Studies in English
	Exclusion Criteria
EC1	Studies with no validation of the proposed technique
EC2	Editorials, short papers, posters, technical reports, patents and reviews

Based on the criteria above, if the study meets the inclusion criteria and none of the exclusion criteria is met then such a study is further moved to the next stage that is quality assessment criteria.

Quality Assessment Criteria: This step was designed to ensure the quality of the studies that were finally going to be the part of the research process. There were total of four questions to estimate the quality of a study and each study is assessed against these four questions based on a 3-point scale. If the study answers the question it is indicated by Y (1 point) and if the study fails to answer the question it is indicated by N (0 points) and if it partially satisfies the answer then P (0.5 points) is given. The overall score required to include the study is 3 out of 4 to maintain high-quality standards. Table 5 contains the questions that were part of the quality assessment criteria.

Table 5: Quality Assessment Criteria

Question #	Criteria	Score
QC1	Are study objectives clearly defined?	Y|N|P
QC2	Are applied methods well defined?	Y|N|P
QC3	Is the accuracy of the applied method measured and reported?	Y|N|P
QC4	Are limitations of study explicitly stated?	Y|N|P

3 RESULTS

This section contains the results and discussion related to the research questions presented in Table. 1.

RQ1: Convergence of Software Engineering Methods with Spreadsheets

The graphical representation of the techniques and their usage in percentage is shown in Figure 3.

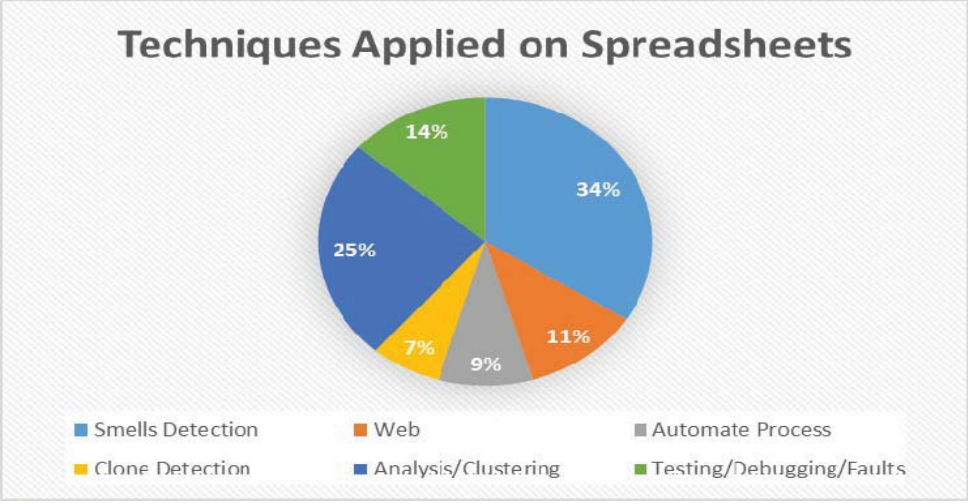

Figure 3: Graph of methods applied to Spreadsheets

It can be clearly seen in Figure 3 the highest percentage 34% is from smell detection which shows that the research community focused more on this area. The second place is acquired by Analysis/Clustering techniques which are 25%. It represents that this area is also the center of attention. Testing/Faults cover 14% of the overall studies. The area that is neglected by researchers is clone detection in spreadsheets that only occupies 7% of the overall research. Year wise distribution of studies with respect to areas on which they focused is shown in Figure 4. The studies are analyzed based on four areas. First is the Web, which means studies that are somehow related to the Web are covered in this area. Second is Automate Process, the studies that are involved in any kind of process that is automated using spreadsheets are included in this area. Third, is Testing/Debugging/Faults, the studies that are related to maintenance of spreadsheets are added in this area. The final and fourth area is Analysis/Comparison, in this area the studies in which any kind of analysis or comparison is done in the context of spreadsheets are included.

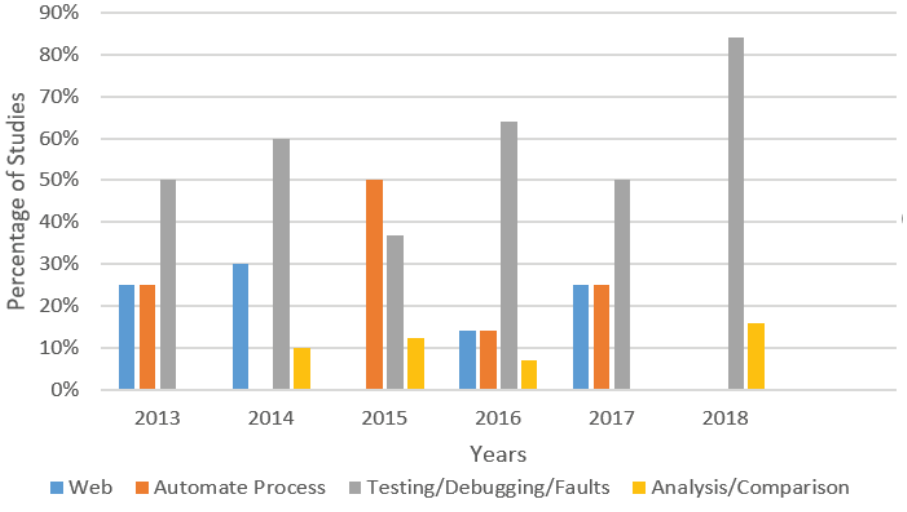

Figure 4: Year Wise Distribution of Studies

In case of year 2013, we can see that Web and Automate Process contains 25% of studies each and 50 % of studies in this year are focused on the research that is related to maintenance. In the year 2014, 30% of research is dedicated to the area of Web and a small portion of 10% is covered by Analysis/Comparison but the significant portion of this year is grabbed by the maintenance. The trend shows that Automate Process contains the highest number of studies in 2015 that is 50% and 37% of studies are included in the area of maintenance and around 12% of studies focused on Analysis/Comparison. It can be visualized from the chart that in the year 2016, all the four areas are covered by studies. Web and Automate Process covers 15% each. Maintenance got the highest percentage of 65% and Analysis/Comparison contributed by 7%. The year 2017 also represents the same trend that was observed in the year 2013. In 2018 85% of studies focused on maintenance and 15% of studies are for Analysis/Comparison. We can conclude from the above distribution of studies that the most focused area in the context of spreadsheets is maintenance from 2013 to 2018. However, it is also advocated that in every single year the highest inclination of research is towards the maintenance of spreadsheets.

RQ2: Data Sources and Datasets

The overall trend of the datasets that are currently preferred for spreadsheets can be visualized in Table 6.

Table 6: Data Sources

Data Source	No. of Studies
MS corpus	1
Enron	5
EUSES	11
Venron	1
DBpedia	1
Wiki Tables	1
ClueWeb09 Web crawl	1
F1F9	1
INFO1	1
Online Sources	29

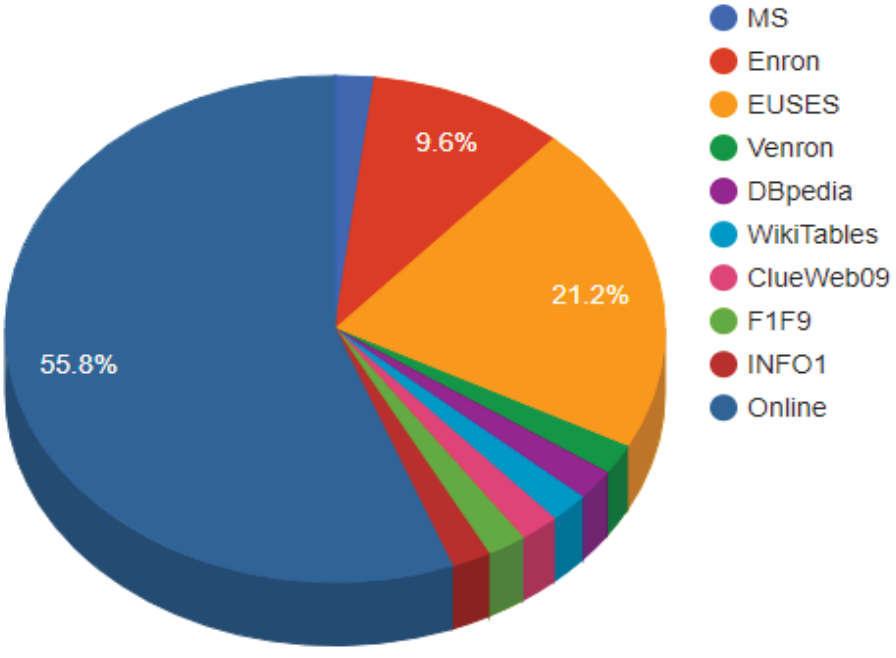

Figure 5: Data Sources

Figure 5 gives an overview in terms of the number of studies against the data sources. The highest number of studies suggests that most studies acquire data from online sources but it do not mean that such datasets do not belong to any of the data sources mentioned above, it means that the data is acquired from online and not from the official source. EUSES is also easily visible and is placed at the second position. Third position is for Enron corps. It can be concluded that the data sets that are mostly used to apply different methods on spreadsheets are from EUSES corps and Enron corps.

RQ3: Overall Research Productivity

Figure 6 shows the overall trend of the publications from the year 2013 to 2018. It can be observed that most studies are published in 2016. From the year 2013 to 2016, there is an increase in the number of studies and in 2017 and 2018, shows identical results.

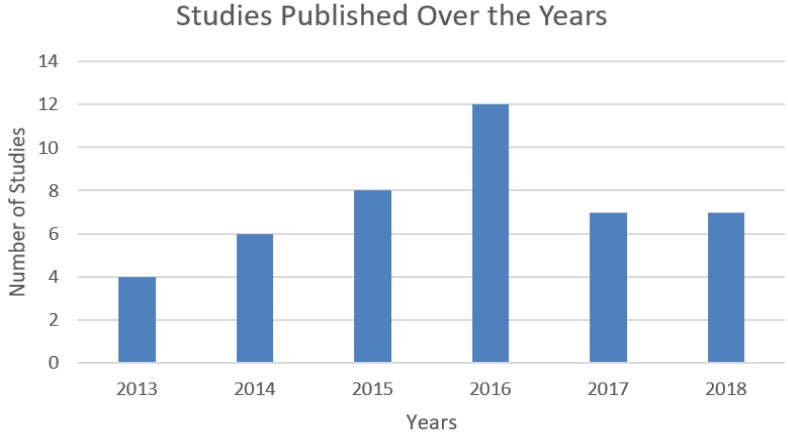

Figure 6: Studies Published over the Years

Table 7: Quality Levels of Selected Studies

Quality Level	Number of Studies	Percentage (%)
Very High (score = 4)	11	25 %
High (score = 3.5)	13	29.54 %
Medium (score = 3)	20	45.45 %
Total	44	100.00 %

Table 7 reveals that 25% of the studies lies in the highest quality span. While more than 29% of studies managed to qualify for the high score and the most number of studies are from medium quality which is around 45% according to our predefined quality assessment criteria. The research approaches that are considered when it comes to spreadsheets are of three types, solution, validation, evaluation.

Table 8: Research Approaches

Research Approach	Number of Studies	Percentage (%)
Solution	34	77.27 %
Validation	6	13.63 %
Evaluation	4	9.09 %
Total	44	100.00 %

Table 8 presents the trend of the approaches followed by studies to perform different methods on spreadsheets. Solution approach is most adopted by the researchers and covers about 77% of the overall studies.

Table 9 gives a detailed overview of the venues in which the research papers about spreadsheets are published. International Conference on Software Engineering is the most visible venue as most number of studies are published in it.

Table 9: Publication Venues

Venue	Number of Studies
International Conference on Software Engineering	6
Symposium on User interface software and technology	2
International Conference on Knowledge Discovery and Data Mining	2
International Conference on Software Analysis, Evolution and Re-engineering	2
Symposium on Visual Languages and Human-Centric Computing	2
International Conference on Software Maintenance and Evolution	2
IEEE TRANSACTIONS ON SOFTWARE ENGINEERING	3
International Conference on Inventive Computation Technologies	2

Automated Software Engineering	2
International Conference on Knowledge Capture	1
Software Engineering Conference and Symposium on the Foundations of Software Engineering	1
Proceedings of the ACM on Programming Languages	1
SIGPLAN Notices	1
International Symposium on Software Testing and Analysis	1
Brazilian Symposium on Systematic and Automated Software Testing	1
International Workshop on Semantic Search over the Web	1
Conference on Information and Knowledge Management	1
International Database Engineering & Applications Symposium	1
International Symposium on Foundations of Software Engineering	1
International Conference on Software Engineering Companion	1
Conference on Human Factors in Computing Systems	1
International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software	1
Conference of the Center for Advanced Studies on Collaborative Research	1
Conference on Programming Language Design and Implementation	1
International Conference on Mining Software Repositories	1
International Symposium on Software Reliability Engineering Workshops	1
IEEE Transactions on Reliability	1
International Workshop on Document Analysis Systems	1
Science of Computer Programming	1
Empirical Software Engineering	1

The contribution of studies that is published by the researchers depends on the nature of the research and the intention of the publisher. There are several types of contribution that is made to the research community. Some researchers do a comparison among different studies, some prefer to propose a new model, framework, tool or some kind of method/technique. In case of spreadsheets the detailed overview of the contribution can be found in Figure 7.

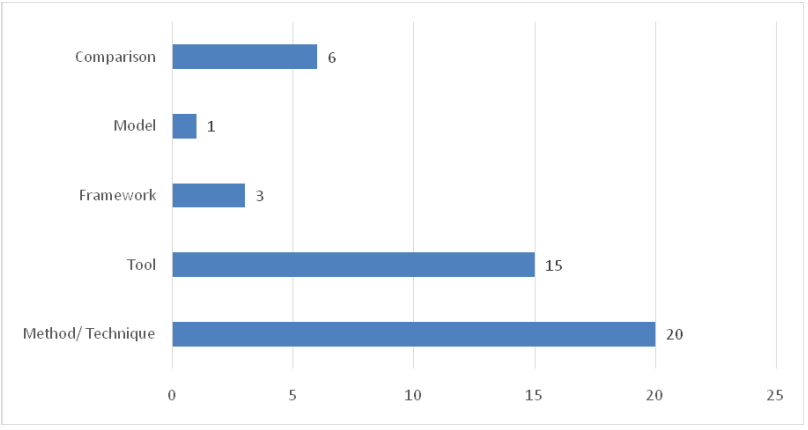

Figure 7: Contribution Facets

As the Figure 7 depicts the trend that 20 of the total studies published focuses on the presentation of some new technique/method. After that, 15 studies proposed a tool. Then 6 studies give a comparison of already existing techniques or tools. New framework was given in 3 publications. And only one study proposed a model.

Another interesting finding is that which countries are actively participating in the spreadsheet research and this can be found out by having a look at the Figure 8, which briefly shows the country name and the number of studies published in that country.

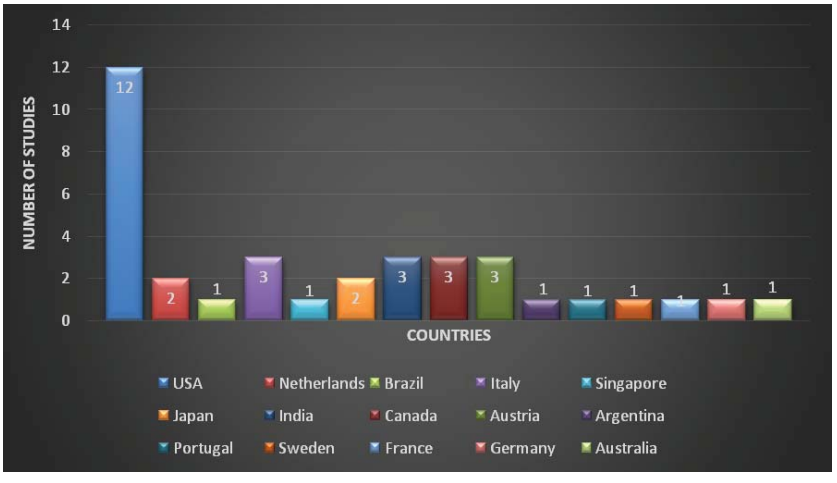

Figure 8: Contribution Facets

There are 12 studies that were published in USA, so USA has the biggest contribution towards spreadsheet research. After that second place is claimed by India, Canada and Austria containing 3 studies each. At third place there are two countries Netherlands and Japan with 2 studies published in each country.

4 CONCLUSIONS

This systematic mapping study summarizes the existing studies with their focus on the spreadsheets and the methods or processes that are applied on spreadsheets. The paper

presents a range of papers on spreadsheets and classifies them according to different criteria like techniques, approaches, datasets and quality. The primary search fetched 520 studies. Then title-based search was performed to get the most relevant papers and 153 papers were extracted. After that, abstract reading was performed to further extend the filtration process and 60 papers managed that filter. Then these papers were passed through the predefined research questions and quality assessment criteria test and finally 44 studies were selected to carry out this systematic mapping study. The main findings of this research are as follows:

- Clone detection is the most neglected area in the context of spreadsheets.
- In most of the studies, researchers gave new techniques or they were able to improve already existing techniques which show solution research is the approach that gained more importance.
- The most used dataset for spreadsheets is from EUSES and Enron.

Software engineering methods application on spreadsheets is gaining importance with time because it is becoming more and more important for the research community because there is a high potential in this area for researchers. The main objective of research in spreadsheets is to get the most out of it because almost every institution, company or business is utilizing spreadsheets in one way or the other.

Acknowledgments

The authors would like to thank Usman Ali (PhD Scholar at Institute of Business Administration) for his valuable feedback on this paper.

References

[1] Juergens, E., Deissenboeck, F., Hummel, B., & Wagner, S. (2009, May). Do code clones matter?. In Software Engineering, 2009. ICSE 2009. IEEE 31st International Conference on (pp. 485-495). IEEE.

[2] Bradley, L., & McDaid, K. (2009, May). Using Bayesian statistical methods to determine the level of error in large spreadsheets. In Software Engineering-Companion Volume, 2009. ICSE-Companion 2009. 31st International Conference on (pp. 351-354). IEEE.

[3] Scaffidi, C., Shaw, M., & Myers, B. (2005, September). Estimating the numbers of end users and end user programmers. In Visual Languages and Human-Centric Computing, 2005 IEEE Symposium on (pp. 207-214). IEEE.

[4] Petersen, K., Vakkalanka, S., &Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. Information and Software Technology, 64, 1-18.

[5] de Vos, M., Wielemaker, J., Schreiber, G., Wielinga, B., & Top, J. (2015, October). A methodology for constructing the calculation model of scientific spreadsheets. In Proceedings of the 8th International Conference on Knowledge Capture (p. 2). ACM.

[6] Zhang, J., Han, S., Hao, D., Zhang, L., & Zhang, D. (2018, October). Automated refactoring of nested-if formulae in spreadsheets. In Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 833-838). ACM.

[7] Inala, J. P., & Singh, R. (2017). WebRelate: integrating web data with spreadsheets using examples. Proceedings of the ACM on Programming Languages, 2(POPL), 2.

[8] Almasi, M. M., Hemmati, H., Fraser, G., McMinn, P., &Benefelds, J. (2018, July). Search-based detection of deviation failures in the migration of legacy spreadsheet applications. In Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis(pp. 266-275). ACM.

[9] Almeida, L., Cirilo, E., & Barbosa, E. A. (2016, September). SS-BDD: Automated Acceptance Testing for Spreadsheets. In Proceedings of the 1st Brazilian Symposium on Systematic and Automated Software Testing (p. 5). ACM.

[10] Benson, E., Zhang, A. X., & Karger, D. R. (2014, October). Spreadsheet driven web applications. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 97-106). ACM.

[11] Chang, K. S. P., & Myers, B. A. (2014, October). Creating interactive web data applications with spreadsheets. In Proceedings of the 27th annual ACM symposium on User interface software and technology (pp. 87-96). ACM.

[12] Barowy, D. W., Gochev, D., & Berger, E. D. (2014, October). CheckCell: data debugging for spreadsheets. In ACM SIGPLAN Notices (Vol. 49, No. 10, pp. 507-523). ACM.

[13] Chen, Z., &Cafarella, M. (2013, August). Automatic web spreadsheet data extraction. In Proceedings of the 3rd International Workshop on Semantic Search over the Web (p. 1). ACM.

[14] Chen, Z., Dadiomov, S., Wesley, R., Xiao, G., Cory, D., Cafarella, M., & Mackinlay, J. (2017, November). Spreadsheet property detection with rule-assisted active learning. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 999-1008). ACM.

[15] Cheung, S. C., Chen, W., Liu, Y., & Xu, C. (2016, May). CUSTODES: automatic spreadsheet cell clustering and smell detection using strong and weak features. In Proceedings of the 38th International Conference on Software Engineering(pp. 464-475). ACM.

[16] Compton, M., Ratcliffe, D., Shu, Y., & Squire, G. (2015, July). Declarative Data Exchange: Spreadsheets to RDF/OWL. In Proceedings of the 19th International Database Engineering & Applications Symposium (pp. 96-105). ACM.

[17] Dou, W., Cheung, S. C., & Wei, J. (2014, May). Is spreadsheet ambiguity harmful? detecting and repairing spreadsheet smells due to ambiguous computation. In Proceedings of the 36th International Conference on Software Engineering (pp. 848-858). ACM.

[18] Dou, W., Cheung, S. C., Gao, C., Xu, C., Xu, L., & Wei, J. (2016, November). Detecting table clones and smells in spreadsheets. In Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering (pp. 787-798). ACM.

[19] Dou, W., Xu, L., Cheung, S. C., Gao, C., Wei, J., & Huang, T. (2016, May). VEnron: a versioned spreadsheet corpus and related evolution analysis. In Software Engineering Companion (ICSE-C), IEEE/ACM International Conference on(pp. 162- 171). IEEE.

[20] Chang, K. S. P., & Myers, B. A. (2016, May). Using and exploring hierarchical data in spreadsheets. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 2497-2507). ACM.

[21] Hermans, F., Sedee, B., Pinzger, M., & van Deursen, A. (2013, May). Data clone detection and visualization in spreadsheets. In Software Engineering (ICSE), 2013 35th International Conference on (pp. 292-301). IEEE.

[22] Hermans, F., & Murphy-Hill, E. (2015, May). Enron's spreadsheets and related emails: A dataset and analysis. In Proceedings of the 37th International Conference on Software Engineering-Volume 2 (pp. 7-16). IEEE Press.

[23] Li, K., He, Y., &Ganjam, K. (2017, August). Discovering enterprise concepts using spreadsheet tables. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1873-1882). ACM.

[24] Koch, P., Schekotihin, K., Jannach, D., Hofer, B., Wotawa, F., & Schmitz, T. (2018, May). Combining spreadsheet smells for improved fault prediction. In Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results (pp. 25-28). ACM.

[25] Kovalenko, O., Serral, E., &Biffl, S. (2013, September). Towards evaluation and comparison of tools for ontology population from spreadsheet data. In Proceedings of the 9th International Conference on Semantic Systems (pp. 57-64). ACM.

[26] McCutchen, M., Itzhaky, S., & Jackson, D. (2016, October). Object spreadsheets: a new computational model for end-user development of data-centric web applications. In Proceedings of the 2016 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software (pp. 112-127). ACM.

[27] Hermans, F. (2013, November). Improving spreadsheet test practices. In Proceedings of the 2013 Conference of the Center for Advanced Studies on Collaborative Research (pp. 56-69). IBM Corp..

[28] Barowy, D. W., Gulwani, S., Hart, T., & Zorn, B. (2015, June). FlashRelate: extracting relational data from semi-structured spreadsheets using examples. In ACM SIGPLAN Notices (Vol. 50, No. 6, pp. 218-228). ACM.

[29] Xu, L., Dou, W., Gao, C., Wang, J., Wei, J., Zhong, H., & Huang, T. (2017, May). SpreadCluster: recovering versioned spreadsheets through similarity-based clustering. In Proceedings of the 14th International Conference on Mining Software Repositories (pp. 158-169). IEEE Press.

[30] Ayalew, Y., Clermont, M., &Mittermeir, R. T. (2008). Detecting errors in spreadsheets. arXiv preprint arXiv:0805.1740.

[31] Koch, P., &Schekotihin, K. (2018, October). Fritz: A Tool for Spreadsheet Quality Assurance. In 2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)(pp. 285-286). IEEE.

[32] Pandita, R., Parnin, C., Hermans, F., & Murphy-Hill, E. (2018, October). No half-measures: A study of manual and tool-assisted end-user programming tasks in Excel. In 2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 95-103). IEEE.

[33] Koch, P., Schekotihin, K., Jannach, D., Hofer, B., Wotawa, F., & Schmitz, T. (2018, May). Combining spreadsheet smells for improved fault prediction. In Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results (pp. 25-28). ACM.

[34] Abreu, R., Cunha, J., Fernandes, J. P., Martins, P., Perez, A., & Saraiva, J. (2014, September). Faultysheet detective: When smells meet fault localization. In Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on(pp. 625-628). IEEE.

[35] Abreu, R., Cunha, J., Fernandes, J. P., Martins, P., Perez, A., & Saraiva, J. (2014, September). Smelling faults in spreadsheets. In Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on (pp. 111-120). IEEE.

[36] Cunha, J., Fernandes, J. P., Mendes, J., & Saraiva, J. (2015). Embedding, evolution, and validation of model-driven spreadsheets. IEEE Transactions on Software Engineering, 41(3), 241-263.

[37] Dou, W., Xu, C., Cheung, S. C., & Wei, J. (2017). CACheck: Detecting and Repairing Cell Arrays in Spreadsheets. IEEE Trans. Software Eng., 43(3), 226-251.

[38] Hermans, F., & Murphy-Hill, E. (2015, May). Enron's spreadsheets and related emails: A dataset and analysis. In Proceedings of the 37th International Conference on Software Engineering-Volume 2 (pp. 7-16). IEEE Press.

[39] Hermans, F., Jansen, B., Roy, S., Aivaloglou, E., Swidan, A., &Hoepelman, D. (2016, March). Spreadsheets are code: An overview of software engineering approaches applied to spreadsheets. In Software Analysis, Evolution, and Reengineering (SANER), 2016 IEEE 23rd International Conference on (Vol. 5, pp. 56-65). IEEE.

[40] Hofer, B. (2017, October). Removing Coincidental Correctness in Spectrum-Based Fault Localization for Circuit and Spreadsheet Debugging. In Software Reliability Engineering Workshops (ISSREW), 2017 IEEE International Symposium on (pp. 199-206). IEEE.

[41] Hofer, B., Höfler, A., &Wotawa, F. (2017). Combining models for improved fault localization in spreadsheets. IEEE Transactions on Reliability, 66(1), 38-53.

[42] Jansen, B., &Hermans, F. (2015, September). Code smells in spreadsheet formulas revisited on an industrial dataset. In Software Maintenance and Evolution (ICSME), 2015 IEEE International Conference on (pp. 372-380). IEEE.

[43] Koci, E., Thiele, M., Lehner, W., & Romero, O. (2018, April). Table Recognition in Spreadsheets via a Graph Representation. In 2018 13th IAPR International Workshop on Document Analysis Systems (DAS) (pp. 139-144). IEEE.

[44] Macedo, N., Pacheco, H., Sousa, N. R., & Cunha, A. (2014, July). Bidirectional spreadsheet formulas. In Visual Languages and Human-Centric Computing (VL/HCC), 2014 IEEE Symposium on (pp. 161-168). IEEE.

[45] Rajdev, U., & Kaur, A. (2016, August). Automatic detection of bad smells from excel sheets and refactor for performance improvement. In Inventive Computation Technologies (ICICT), International Conference on (Vol. 2, pp. 1-8). IEEE.

[46] Rajdev, U., & Kaur, A. (2016, August). Bad smell refactor removable detection of audit spreadsheet. In Inventive Computation Technologies (ICICT), International Conference on (Vol. 1, pp. 1-7). IEEE.

[47] Cunha, J., Mendes, J., Saraiva, J., & Visser, J. (2014). Model-based programming environments for spreadsheets. Science of Computer Programming, 96, 254-275.

[48] Hermans, F., Pinzger, M., & van Deursen, A. (2015). Detecting and refactoring code smells in spreadsheet formulas. Empirical Software Engineering, 20(2), 549-575.

[49] Jannach, D., & Schmitz, T. (2016). Model-based diagnosis of spreadsheet programs: a constraint-based debugging approach. Automated Software Engineering, 23(1), 105-144.

[50] Schmitz, T., Hofer, B., Jannach, D., &Wotawa, F. (2016, July). Fragment-based diagnosis of spreadsheets. In Federation of International Conferences on Software.