Decision Support for Selecting Tools for Software Test Automation

Päivi Raulamo-Jurvanen
M3S (M-Group)
ITEE
University of Oulu, Finland
+358 46 922 9669
paivi.raulamo-jurvanen@oulu.fi

DOI: 10.1145/3011286.3011304
http://doi.acm.org/10.1145/3011286.3011304

ABSTRACT

Context: Test automation is an investment having a high initial economic impact on software development. Utilization of test automation may positively affect the costs (e.g. by speeding up development iterations by providing repeatable tests and regression testing) and the quality of software or system, in large scale. Approaches to test automation may not always be appropriate or successful. The trade-off between manual and automated testing and the tools to be used have to be identified and justified. The task to decide which tools to use, to maximize the benefits is not a trivial one. There are numerous software testing or software test automation tools available, both commercial and open source and unique, multifaceted goals in every development environment (context). The exact number of tools is unknown and chances or resources to try out different choices are very limited. ***Objective:*** Contextual factors are acknowledged as an issue and well known and common to both practitioners in the field and consultation service providers. Selecting and utilizing the most effective and efficient tool(s) for specific purpose(s) in a specific context is essential for the success of business. The goal of the research is to define a systematic, empirically validated decision support system (DSS) for selecting a tool for software test automation.

Categories and Subject Descriptors

D.2.5 [**Software Engineering**]: Testing and Debugging – *testing tools.* H.4.2 [**Information Systems Applications**]: Types of Systems – *decision support.* H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – s*earch process.* D.2.4 [**Software Engineering**]: Software/Program Verification – *validation*.

General Terms

Management, Human Factors, Theory, Verification.

Keywords

Software testing, Software Test Automation, Tool Selection, Contextual factors, Surveys, Web-scraping

1. INTRODUCTION

Testing and quality assurance account up to half of the budgets in development of software systems [1-8]. Those activities are many times complex, take time, require experience and skilled personnel. Therefore, testing and quality assurance have been described as bottlenecks for the software development.

Studies claim that very few companies have realized the value of software test automation [1, 2, 9-17]. However, recent industrial surveys show increased interest in software test automation. For example, general adoption of test automation is claimed to be 72% in organizations and average percentage of test case automation is claimed to have increased within a year (since 2014 to 2015) from 28% to 45%. [18, 19].

Investments for test automation apply to development environments as a whole, not just to processes, practices and tools but also to employees and their skills. Lack of right tools for testing activities and lack of skilled resources have been mentioned as obstacles to test automation in the industrial surveys [18, 20]. Thus, it is not a surprise that the findings of ISTQB report of 2015-2016 [19] ranked test automation and test tool and automation consultation as the main area of improvement opportunities (in testing activities) and the service most needed from external consultants, respectively.

Finding a suitable software test automation tool is not a trivial task. There are numerous tools available, both commercial and open source, for different purposes and with varying functionality. Past work on software testing tool selection, in general, by Poston and Sexton [21] claimed systematic data collection method, preferably with forms or checklists to be the secret for selecting appropriate testing tools. There has been a number of surveys of conducted focusing on software testing e.g. [1, 22-24], but those do not typically cover the actual tools used. Some studies have focused on some few specific tools e.g. [25-28] but such experimental studies conducted by researchers (in some experimental contexts) are not generalizable as such.

There are two essential weaknesses claimed to be common to researches in the area of software engineering [29-33]. First is the inconsistent formulation of specific research questions. The problem statement can be incoherent if not contextualized sufficiently. Research goals should be given an explicit context, a perspective to explore and explain the contextual impact - what, who, where, when & why - to define the context in relation to specific practice situation. Secondly, derived from the first issue, lessons learned from one project (or some results from a specific research) are most likely not applicable to accumulate knowledge in another project or research, respectively. A context cannot be explained as a set of distinct variables but rather as a specific practice

situation having a reflexive relationship with practice, both shaping one another as situations vary and evolve.

That in mind, the past peer-reviewed literature is missing surveys focusing on software test automation tools by naming them. Such tools get more visibility in the industry. Professionals in the field have invested in choosing and using those tools. Thus, they possess knowledge and understanding about important criteria for comparison and selection of tools and they possess knowledge about those tools, too.

2. OBJECTIVES

In the software industry, there is a need and will to automate software testing [1, 2, 9-16, 18-20, 34-37] but no empirically grounded, systematically assessed decision support on how to select appropriate tools to use for software test automation purposes. The goal of this dissertation is to develop such decision support. This requires studying the factors affecting the decision-making for and adoption of test automation and advantageous test automation tools. The idea is to analyze criteria and dependent factors to uncover context and raise awareness of its importance in test automation.

It is expected that general requirements, excluding localization issues, for software testing tools (e.g. costs, possible licensing model or developer support, to mention a few) are rather similar despite the physical location. Software development is not climate dependent like agriculture or livestock farming, where environmental characteristics, in general, may have an impact or clear restrictions on products or livestock possible to farm (at least without special arrangements). To create an accurate and reliable picture of the tools in the markets, capabilities and assets of those, the best experts for such information would be those who have used or experienced the tools in real life, software professionals in the industry.

The academic community and professionals in the industry have expressed mostly guidelines and heuristics for the adoption of test automation [9, 34, 38-40]. The selection criteria leading to the decision of adoption of test automation do not seem to be systematically used. That may originate from either lack of relevant empirical researches or non-contextual results, or both. That in mind, the importance of contextual factors is evident.

The purpose of the research is to define a systematic decision support system (DSS), an empirically validated DSS for selecting a tool for software test automation. The DSS will be based on the concept of "Wisdom of the crowd" which embodies the idea of collective opinion (or intelligence), that claims that under the right circumstances a group can be smarter than a single individual [41]. The DSS will also be based on the research work conducted in the faculty of ITEE of the University of Oulu [42]. (Please note, that paper will be published in the summer of 2016, thus not available at the time of writing this paper). The benefit of the system is to facilitate perceiving of pitfalls specific to a given purpose or context, i.e. problems endangering the actions of improvement or the business as a whole.

3. RESEARCH QUESTIONS

Test automation consultants Graham and Fewster [9] emphasize that "there is no such thing as the perfect tool, but there are many tools that would be adequate for a given situation". It is the current knowledge or preferences that drive the decisions. They claim that the tool must be appropriate for a job, thus, a tool may be inadequate in some context but suitable for another [9]. Also, some tool may be used unconventionally to a totally different purpose than originally created for. Thus, it seems it is the properties of a tool that drive the selection of a software test automation tool, and the context where software automation tool is applied to defines the purpose.

The research will focus on the factors affecting the context and decision-making how adoption of test automation would be appropriate and advantageous. The research will evaluate some known decision support methods (possibly Repertory Grid & Crowd-based decision support system) and the applicability of those to the given context.

The goal of the research is to systematically review and classify the state-of-the-art & state-of-practice of selecting a software test automation tool for developing an empirically validated decision-support. The DSS is intended to provide insight into the following problems, see Table 1, Table 2 and Table 3…

Table 1. Main Research Problem

Main Research Problem
What matters in the process of selecting software test automation tools?

Based on the above goal, the following research questions are raised…

Table 2. Research Question #1

Research Question #1
What are the major contextual factors affecting adoption of software test automation?

The importance of software test automation seems to be widely recognized and accepted, but there seems to be obstacles like issues with commitment, tools and understanding about the assimilation of software test automation into processes. This research question is intended to find out the major contextual factors affecting the adoption of software test automation, in general.

Table 3. Research Question #2

Research Question #2
What are the important criteria affecting the selection of a software test automation tool?

Although the basic criteria for the selection of a software test automation tool are expected to be rather similar it seems the industry is struggling with issues like finding a suitable tool. This research question is intended to find out common important criteria that should be addressed, in general, in the selection of a software test automation tool, in order to find the most suitable tool for the purpose.

4. RESEARCH METHODS

The basis of the research is a literature review, see Table 4 (1 & 7). Relevant literature will be reviewed to identify and to classify selection criteria, or strategy, used for test automation. The idea is to try to find out how knowledge has been applied in such strategy.

The preliminary questionnaire, Table 4 (2), was designed to find out practical knowledge from the field, criteria for tool selection and an initial list of software automation tools used by the respondents in the industry. The questionnaire was sent to selected groups of (mainly Finnish) software professionals, Table 4 (3a). Web-scraping was utilized as an additional quantitative and qualitative support for analysis of the survey, Table 4 (3b). The results of the Survey 1 were embedded in an industrial paper which was submitted and accepted to PROFES2016 industry track, Table 4 (4). The findings and adopted data acquisition methods can be applied to the following studies as well.

The test phase of the DSS will be set up for collecting the core data, Table 4 (5 & 6a). The respondents will be asked to answer the questions and evaluate the selected tools based on given criteria. The collected data will be analyzed and evaluated (Table 4 (6b), to reflect the DSS, Table 4 (8a).

Table 4. Research work & Schedule

Research Step	Activity	Schedule
1. Literature review **Status**: ongoing activity throughout the research	Literature Review / Mapping Study	**Ongoing**
2. Questionnaire for Survey 1	Design of questionnaire for Survey 1 Targeted to selected groups of Finnish software professionals in the industry.	**Completed 01/2016**
3. Survey 1 Find out criteria for tool selection and an initial list of software automation tools.	a) Sending of the questionnaire to selected groups, in co-operation with Knowit Oy Survey open Feb 22nd - Mar 20th 2016.	**Completed 03/2016**
	b) Analysis of the data from 3a & previous study from 2013. Web-scraping utilized as and additional, supportive method for acquiring data about the tools. Paper with Kari Kakkonen, Knowit Oy & Professor Mika Mäntylä	**Completed 05/2016**
4. Publication 1 (RQ2)	**Full paper,** PROFES2016, https://profes2016.idi.ntnu.no/ Accepted 9th Aug 2016.	**Completed 05/2016**
5. Set-up of the DSS – test phase	Design & implement the preliminary DSS, in co-operation with UBICOMP, with intention to collect experiences of software professionals in the industry.	**Completed 06/2016**
6. Survey 2 Collect data for the DSS	a) Collection of the data from software professionals	**Ongoing 08-10/2016**
	b) Analysis of the data from 6a	**10-11/2016**
7. Publication 2 (RQ1 & RQ2)	Mapping study	**10 /2016**
8. Publish the DSS	Configure the system (based on Survey 2) and publish & collect data	**12/2016**
9. Focus Group Study / Interviews	Focus Group Study/Interviews of selected software professionals to compare and analyze surveys and Web-scraping methods for acquiring data.	**05-07/2017**
10. Publication 3 (RQ1 & RQ2)	Based on empirical findings (and possibly Focus Group Study / Interviews)	**09/2017**
11. Analysis of the results	Analysis of the results from different phases, conclusions	**11/2017**
12. Publication 4	Summary of the findings – How to support the selection of software test automation tools?	**01/2018**
13. Complete the PhD Thesis	Writing the thesis	**11/2018**
14. Finalize the research	Publish the results	**01/2019**

Please note, the details will be clarified as the work gets started & progresses, the plan will be refined over time.

The empirical data for the analysis is collected via utilizing the DSS, Table 4 (8). The idea is to provide feedback for the user, based on his/her answers: e.g. describe how the given answers relate to answers submitted by other respondents or point out problematic areas as a result. The data will be analyzed to conclude the findings for the dissertation and for publishing the results, Table 4 (9-14). The further studies and schedule for later publications will be defined in more detail as the work progresses. Figure 1 defines the process for the development of the DSS.

The intention is to survey existing literature (both academic, peer-reviewed literature and industrial surveys) and acquire empirical data for study by utilizing questionnaires & surveys, also utilizing focus group method (and possibly interviews, in Finland & in Turkey) and the to-be-generated DSS. Focus group method is claimed to be useful for collecting data about experiences and knowledge from a group of people to find out – not just *what* people think, but also *how* they think and *why* people think the way they do [43]. The idea is to utilize those methods as vehicles to harness the concept of "the wisdom of the crowds" for the processes. The concept is also defined as "the collective opinion of a group of individuals rather than that of a single expert" [44]. Another idea is to utilize web-scraping (within legal and ethical limits), where required and applicable.

5. RESEARCH WORK & SCHEDULE

The schedule is divided into 14 tasks, see Table 4. Planned publications are defined as tasks of their own to clearly point out their importance and for allowing easy changes to the schedule, if needed. The already accomplished tasks are marked as "Completed" in Table 4 (currently tasks 2-5).

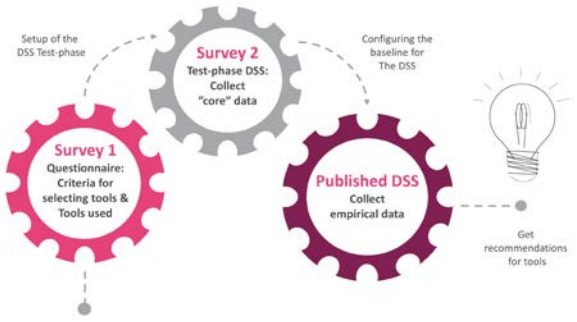

Figure 1. The process of developing the DSS.

The future work and related details will be clarified as the work progresses – obviously, the plan will be refined over time, several times. The research is at its early stage, thus all details and decisions have not been nailed, yet.

6. RESULTS TO DATE

A paper has been submitted to the industry track of PROFES2016 with the title "Using Surveys and Web-Scraping to Select Tools for Software Testing Consultancy". The paper has been written in co-operation with Kari Kakkonen, Knowit Oy & Professor Mika Mäntylä, University of Oulu.

The objective of the study was to analyze how findings from data collected utilizing surveys and web-scraping could support a software testing consultation company in the process of selecting the best tools. A similar type of survey had been conducted in 2013 (also amongst mostly Finnish software professionals), however the survey questionnaires of the two surveys were rather different. The results from that survey were compared with the recent survey of 2016. About 82% of the respondents of the recent survey reported to be working in Finland.

The survey of 2016 was utilized to acquire a list of tools used for software testing in industry and criteria used for tool selection. A quantitative analysis was done for the top 15 tools of the latter survey utilizing data collected by web-scraping. The observation period for the data collected by web-scraping was the first quarter of 2016 (January 1st – March 31st 2016), to provide some variation and recent body to the content. Data was acquired for tool specific Wikipedia page views, Google hits, number of StackOverflow questions and view counts for those and for number of Twitter tweets.

The tools included clear local preferences, e.g. Robot Framework being particularly popular in Finland due to its origin). However, four and five out of nine tools reported for small and large size organizations [45], respectively, were included in the list of Top 15 tools of the survey. Interestingly, Selenium and Jenkins, appeared in 2nd and 3rd places (respectively) in both our survey of 2016 and the results from web-scraping. Also, the appearance of Appium, a mobile-suitable tool, seemed as an expected development, both in the results of survey of 2016 and web-scraping.

In the study, we contrasted methods and results from web-scraping and surveys. As with methods, web-scraping was identified as a rather cost-effective, quick way to acquire large amounts of data compared with surveys. However, finding suitable sources of data for web-scraping and methods to process the data were identified as possible obstacles for the method. Surveys give voice to the people in specific market sectors and contexts and web-scraping is considered as an additional source to surveys. Regarding the results from surveys and web-scraping, it was concluded that survey results are always biased by the size and origin of the sample but help to position the results to more concrete contexts. Web-scraping was, again, considered to provide additional quantitative perspective for the problem.

There seems to be a rather common consensus on criteria for selecting the tools, e.g. ease of use and compatibility and applicability of the tools were among the most important ones. The respondents also emphasized the importance of the tool support and costs related to the tool and its usage. We concluded that multiple sources and methods should be used to get a comprehensive view of the popularity of the tools in the markets. Web-scraping was seen as a very cost-effective support for more traditional survey instruments for collecting data.

7. THREATS TO VALIDITY

One of the main issues considering threats to validity is related to the participants of the surveys. There is always the possibility of having just a small number of participants. Also, the possible national bias of the participants needs to be acknowledged (e.g. if most participants for the surveys are from Finland). However, there has been initial plans to conduct another survey and Focus Group study (or interviews) in Turkey, too. I also hope to be able to co-operate with partners from the industry but again, there is always the chance those people are too busy to contribute to my research.

8. CONCLUSIONS

The topic and the research problem emerges from the industry. The intention is to exploit the knowledge and experience of software professionals in the industry. The research is at its early stage. Only the very first steps towards the development of the DSS have been taken, and the research will continue from now on with baby steps.

9. REFERENCES

[1] Tassey, G. 2002. The economic impacts of inadequate infrastructure for software testing. *National Institute of Standards and Technology, RTI Project.* 7007.

[2] Strigel, W., Juristo, N., and Moreno, A. M. 2006. Guest Editors' Introduction: Software Testing Practices in Industry. *IEEE Software.* 23, 0019-21.

[3] Blackburn, M., Busser, R., and Nauman, A. 2004. Why model-based test automation is different and what you should know to get started. In *International Conference on Practical Software Quality and Testing.* 212-232.

[4] Engel, A. and Last, M. 2007. Modeling software testing costs and risks using fuzzy logic paradigm. *J. Syst. Software.* 80, 817-835.

[5] Harrold, M. J. 2000. Testing: A roadmap. In *Proceedings of the Conference on the Future of Software Engineering.* 61-72.

[6] Chittimalli, P. K. and Harrold, M. J. 2009. Recomputing coverage information to assist regression testing. *Software Engineering, IEEE Transactions.* 35, 452-469.

[7] Andersson, C. and Runeson, P. 2002. Verification and validation in industry-a qualitative survey on the state of practice. In *Empirical Software Engineering, Proceedings. 2002 International Symposium N.* 37-47.

[8] Budnik, C. J., Chan, W. K., and Kapfhammer, G. M. 2010. Bridging the gap between the theory and practice of software test automation. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering. 2,* 445-446.

[9] Graham, D. and Fewster, M. 2012. *Experiences of Test Automation: Case Studies of Software Test Automation.* Addison-Wesley Professional.

[10] Rafi, D. M., Moses, K. R. K., Petersen, K., and Mäntylä, M. V. 2012. Benefits and limitations of automated software testing: Systematic literature review and practitioner survey. In *Proceedings of the 7th International Workshop on Automation of Software Test.* 36-42.

[11] Taipale, O., Kasurinen, J., Karhu, K., and Smolander, K. 2011. Trade-off between automated and manual software testing.

International Journal of System Assurance Engineering and Management. 2, 114-125.

[12] Fewster, M. 2001. Common mistakes in test automation. In *Proceedings of Fall Test Automation Conference*.

[13] Stobie, K. 2009. Too much automation or not enough? when to automate testing. In *Pacific Northwest Software Quality Conference*.

[14] Sahaf, Z., Garousi, V., Pfahl, D., Irving, R., and Amannejad, Y. 2014. When to automate software testing? decision support based on system dynamics: An industrial case study. In *Proceedings of the 2014 International Conference on Software and System Process.* 149-158.

[15] Ramler, R. and Wolfmaier, K. 2006. Economic perspectives in test automation: Balancing automated and manual testing with opportunity cost. In *Proceedings of the 2006 International Workshop on Automation of Software Test.* 85-91.

[16] Whittaker, J. 2000. What is software testing? And why is it so hard? *Software, IEEE.* 17, 70-79.

[17] Kasurinen, J., Taipale, O., and Smolander, K. 2009. Analysis of problems in testing practices. In *Software Engineering Conference, 2009. APSEC'09. Asia-Pacific.*309-315.

[18] Capgemini Consulting. 2015. *World Quality Report 2015-2016*. Capgemini, Sogeti and HP. https://www.capgemini.com/thought-leadership/world-qualityreport-2015-16.

[19] ISTQB (International Software Testing Qualifications Board). 2016. ISTQB® worldwide software testing practices report 2015 - 2016. ISTQB. http://www.istqb.org/references/surveys/istqb-worldwidesoftware-testing-practices-report.html.

[20] Capgemini Consulting. 2014. World quality report 2014-2015, sixth edition. Capgemini, Sogeti and HP. https://www.capgemini.com/resources/world-quality-report-2014-15.

[21] Poston, R. M. and Sexton, M. P. 1992. Evaluating and selecting testing tools. *Software, IEEE*. 9, 33-42.

[22] Garousi, V. and Zhi, J. 2013. A survey of software testing practices in Canada. *J. Syst. Software*. 86, 1354-1376.

[23] Michael, M. and Tragoudas, S. 2002. ATPG tools for delay faults at the functional level. *ACM Trans. Design Autom. Electron. Syst.* 7, 33-57.

[24] Ng, S., Murnane, T., Reed, K., Grant, D., and Chen, T. 2004. A preliminary survey on software testing practices in australia. In *Software Engineering Conference, 2004. Proceedings. 2004 Australian.* 116-125.

[25] Kaur, M. and Kumari, R. 2011. Comparative study of automated testing tools: Testcomplete and quicktest pro. *International Journal of Computer Applications*, vol. 24, pp. 1-7, 2011.

[26] Portillo-Rodríguez, J., Vizcaíno, A., Piattini, M., and Beecham, S. 2012. Tools used in Global Software Engineering: A systematic mapping review. *Information and Software Technology.* 54, 663-685.

[27] Kaur, H. and Gupta, G. 2013. Comparative Study of Automated Testing Tools: Selenium, Quick Test Professional and Testcomplete. *Int.Journal of Engineering Research and Applications ISSN*. 1739-1743.

[28] Hussain, S., Wang, Z., Toure, I. K., and Diop, A. 2013. Web service testing tools: A comparative study. *arXiv Preprint arXiv:1306.4063*.

[29] Petersen, K. and Wohlin, C. 2009. Context in industrial software engineering research. In *Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement*. 401-404.

[30] Dybå, T., Kitchenham, B., and Jorgensen, M. 2005. Evidence-based software engineering for practitioners. *Software, IEEE.* 22, 58-65.

[31] Dybå, T., Sjøberg, D. I., and Cruzes, D. S. 2012. What works for whom, where, when, and why?: On the role of context in empirical software engineering. In *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. 19-28.

[32] Johns, G. 2006. The essential impact of context on organizational behavior. *Academy of Management Review.* 31, 386-408.

[33] Sjoberg, D. I., Dybå, T., and Jorgensen, M. 2007. The future of empirical methods in software engineering research. In *2007 Future of Software Engineering*. 358-378.

[34] Garousi, V. and Mäntylä, M. V. 2016. When and what to automate in software testing? A multi-vocal literature review. *Information and Software Technology.* 76, 92-117.

[35] Garousi, V. and Varma, T. 2010. A bibliometric assessment of canadian software engineering scholars and institutions (1996-2006). *Computer and Information Science.* 3, 19.

[36] Lehtinen, T. O., Mäntylä, M. V., Vanhanen, J., Itkonen, J., and Lassenius, C. 2014. Perceived causes of software project failures– An analysis of their relationships. *Information and Software Technology.* 56, 623-643.

[37] Mosley, D. J. and Posey, B. A. 2002. *Just enough Software Test Automation.* Prentice Hall Professional.

[38] Hendrickson, E. 1998. The differences between test automation success and failure. *Proceedings of STAR West*.

[39] Hoffman, D. 1999. Cost benefits analysis of test automation. *STAR West.* 99.

[40] Meszaros, G., Smith, S. M., and Andrea, J. 2003. The test automation manifesto. In *Extreme Programming and Agile Methods-XP/Agile Universe 2003*. Springer. 73-81.

[41] Surowiecki, J. 2005. *The Wisdom of Crowds.* Anchor.

[42] Hosio, S., Goncalves, J., Anagnostopoulos, T., and Kostakos, V. 2016. Leveraging wisdom of the crowd for decision support. In *Proceedings of British Human Computer Interaction Conference (BCS-HCI)*.

[43] Kitzinger, J. 1995. Qualitative research. Introducing focus groups. *BMJ*. 311 (July 29), 299-302.

[44] Wikipedia. *Wisdom of the crowd*. https://en.wikipedia.org/wiki/Wisdom_of_the_crowd

[45] Yehezkel, S. 2016. Test Automation Survey 2016. March 16, 2016. http://blog.testproject.io/2016/03/16/test-automation-survey-2016/.