



A systematic literature review of literature reviews in software testing



Vahid Garousi^{a,b,*}, Mika V. Mäntylä^c

^aSoftware Engineering Research Group, Department of Computer Engineering, Hacettepe University, Ankara, Turkey

^bMaral Software Engineering Consulting Corporation, Calgary, Canada

^cM3S, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

ARTICLE INFO

Article history:

Received 17 December 2015

Revised 8 September 2016

Accepted 9 September 2016

Available online 12 September 2016

Keywords:

Secondary studies

Tertiary study

Software testing

Systematic mapping

Systematic literature reviews

Surveys

ABSTRACT

Context: Any newcomer or industrial practitioner is likely to experience difficulties in digesting large volumes of knowledge in software testing. In an ideal world, all knowledge used in industry, education and research should be based on high-quality evidence. Since no decision should be made based on a single study, secondary studies become essential in presenting the evidence. According to our search, over 101 secondary studies have been published in the area of software testing since 1994. With this high number of secondary studies, it is important to conduct a review in this area to provide an overview of the research landscape in this area.

Objective: The goal of this study is to systematically map (classify) the secondary studies in software testing. We propose that tertiary studies can serve as summarizing indexes which facilitate finding the most relevant information from secondary studies and thus supporting evidence-based decision making in any given area of software engineering. Our research questions (RQs) investigate: (1) Software-testing-specific areas, (2) Types of RQs investigated, (3) Numbers and Trends, and (4) Citations of the secondary studies.

Method: To conduct the tertiary study, we use the systematic-mapping approach. Additionally, we contrast the testing topics to the number of Google hits to address a general popularity of a testing topic and study the most popular papers in terms of citations. We furthermore demonstrate the practicality and usefulness of our results by mapping them to ISTQB foundation syllabus and to SWEBOK to provide implications for practitioners, testing educators, and researchers.

Results: After a systematic search and voting process, our study pool included 101 secondary studies in the area of software testing between 1994 and 2015. Among our results are the following: (1) In terms of number of secondary studies, model-based approach is the most popular testing method, web services are the most popular system under test (SUT), while regression testing is the most popular testing phase; (2) The quality of secondary studies, as measured by a criteria set established in the community, is slowly increasing as the years go by; and (3) Analysis of research questions, raised and studied in the pool of secondary studies, showed that there is a lack of 'causality' and 'relationship' type of research questions, a situation which needs to be improved if we, as a community, want to advance as a scientific field. (4) Among secondary studies, we found that regular surveys receive significantly more citations than SMs ($p = 0.009$) and SLRs ($p = 0.014$).

Conclusion: Despite the large number of secondary studies, we found that many important areas of software testing currently lack secondary studies, e.g., test management, role of product risk in testing, human factors in software testing, beta-testing (A/B-testing), exploratory testing, testability, test stopping criteria, and test-environment development. Having secondary studies in those areas is important for satisfying industrial and educational needs in software testing. On the other hand, education material of ISTQB foundation syllabus and SWEBOK could benefit from the inclusion of the latest research topics, namely search-based testing, use of cloud-computing for testing and symbolic execution.

© 2016 Elsevier B.V. All rights reserved.

Contents

1. Introduction	196
2. Background and related work	197
2.1. Other tertiary studies in SE	197

* Corresponding author at: Software Engineering Research Group, Department of Computer Engineering, Hacettepe University, Ankara, Turkey.

E-mail addresses: vgarousi@gmail.com, vahid.garousi@hacettepe.edu.tr (V. Garousi), mika.mantyla@oulu.fi (M.V. Mäntylä).

2.2.	A review of research questions investigated in other SE tertiary studies	198
3.	Research method	198
3.1.	Goal and research questions	198
3.2.	Research process	199
4.	Article selection	199
4.1.	Source selection and search keywords	199
4.2.	Quality assessment	199
4.3.	Application of inclusion/exclusion criteria	200
4.4.	Final pool of secondary studies and the online repository	201
5.	Development of the systematic map	201
5.1.	Iterative development of the systematic map	201
5.2.	Final systematic map	201
5.3.	Metrics and data extraction	201
6.	Results	201
6.1.	RQ1: software testing areas investigated in the secondary studies	201
6.1.1.	Testing methods	201
6.1.2.	Systems Under Test (SUT)	204
6.1.3.	Testing types/phases	204
6.1.4.	Other factors	204
6.2.	RQ2: research questions being investigated	204
6.3.	RQ3: annual trends of types, quality, and number of primary studies	205
6.3.1.	Annual trends and types	205
6.3.2.	Quality, number of primary studies and number of RQs	207
6.4.	RQ4: highest-cited secondary studies, and citation comparison of secondary versus primary studies	208
6.4.1.	Highly-cited secondary studies	208
6.4.2.	Comparing citations to regular surveys, SMs and SLRs	209
6.4.3.	Comparing citations between secondary and primary studies	209
7.	Discussions	210
7.1.	Interpretations and implications	210
7.1.1.	Implications for the industry (ISTQB syllabus)	210
7.1.2.	Implications for the software engineering education (based on SWEBOK guide)	211
7.1.3.	Implications for the research community	211
7.2.	Threats to validity	213
7.2.1.	Internal validity	213
7.2.2.	Construct validity	214
7.2.3.	Conclusion validity	214
7.2.4.	External validity	214
8.	Conclusions and future work	214
	Acknowledgements	215
	References	215

1. Introduction

Secondary studies are common in software engineering (SE). A secondary study is defined as a study of studies [1], i.e., a review of individual (or, primary) studies. Example types of secondary studies include: regular surveys, Systematic Literature Reviews (SLR), and Systematic Mapping (SM) studies.

Software testing is an active area of SE. According to our search, over 101 secondary studies have been published in the area of software testing since 1994. With this high number of secondary studies in this area, it is important to conduct a tertiary review in this area to provide an overview of the research landscape in this area. A tertiary review is a study of secondary

studies (or, a systematic review of systematic reviews) [2]. Tertiary studies 'review the reviews' in a given area in order to provide an overview of the state of evidence in that area. The SE community as a whole believes that secondary and tertiary studies are useful, e.g., [2–5]. There are relatively high number of citations to secondary and tertiary studies in SE, and also there are studies such as [4] which report the usefulness and value of these studies.

There have been tertiary studies in various areas of SE (e.g., [6–15]), but none focusing on testing yet. As discussed above, any newcomer researcher or industrial practitioner is likely to experience difficulties in digesting large volumes of knowledge in software testing. Also, in an ideal world, all knowledge used in industry, education and research should be based on high-quality evidence. Since no decision should be made based on a single study, secondary studies become essential in presenting the evidence. We propose that tertiary studies can serve as summarizing indexes which would facilitate it to find the most relevant information from secondary studies and thus supporting evidence-based decision making in any given area of software engineering.

The authors believe that a tertiary studies should be like the "index" of a book. Such a tertiary study will be useful in that it is read first by the people (e.g., new PhD students and practitioners) who want to know what is out there in a given area (software testing, in our case). A complain often heard from practitioners is that academic literature is impenetrable due to the sheer volume of the literature. A tertiary study such as the current paper should make it more penetrable. Also, such a tertiary study could be used as an aid when constructing contents of research-intensive courses on software testing. If a sub-field of software testing has a large body of research literature, then a review of this literature (via secondary studies) has most likely been performed or should be performed.

Based on the above needs and motivations, in this work, we systematically classify the body of knowledge in secondary studies in software testing via a tertiary study [16]. Our study aims at answering the following four research questions (RQs):

- RQ1: What software-testing-specific areas have been investigated in the secondary studies? Answering this RQ will enable us to determine the software-testing-specific areas covered and not covered by secondary studies. Knowing the areas not covered will pinpoint the need for conducting secondary studies in those areas.
- RQ2: What types of RQs are being investigated? This allows us to characterize the studies in software testing from the viewpoint of philosophy of science. This can help us find gaps and trends in type of secondary studies being conducted.
- RQ3: What are the annual trends of types, quality, and number of primary studies reviewed by the secondary studies? Answering this RQ will allow us to get a big picture of the landscape in this area.
- RQ4: What are the highest cited secondary studies and are the secondary studies cited more often than primary studies? Given the importance of citations to determine scientific merit, we decided to investigate what secondary studies are the most cited. For the same reason, we investigate whether secondary studies receive more citations than primary studies.

As a part of this study, we define inclusion (selection) and exclusion criteria of relevant secondary studies, and systematically develop and refine a systematic map (classification schema) of all the selected studies.

The remainder of study is organized as follows. Section 2 reviews the related work. Section 3 describes our research method, including the overall SM process, the goal and research questions tackled in this study. Section 4 discusses the article selec-

Download English Version:

<https://daneshyari.com/en/article/4972310>

Download Persian Version:

<https://daneshyari.com/article/4972310>

[Daneshyari.com](https://daneshyari.com)