Information and Software Technology 52 (2010) 14–30

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

A systematic review on regression test selection techniques

Emelie Engström, Per Runeson *, Mats Skoglund

Department of Computer Science, Lund University, SE-221 00 Lund, Sweden

ARTICLE INFO

Article history:
Received 4 February 2009
Received in revised form 24 June 2009
Accepted 10 July 2009
Available online 18 July 2009

Keywords:
Regression testing
Test selection
Systematic review
Empirical studies

ABSTRACT

Regression testing is verifying that previously functioning software remains after a change. With the goal of finding a basis for further research in a joint industry-academia research project, we conducted a systematic review of empirical evaluations of regression test selection techniques. We identified 27 papers reporting 36 empirical studies, 21 experiments and 15 case studies. In total 28 techniques for regression test selection are evaluated. We present a qualitative analysis of the findings, an overview of techniques for regression test selection and related empirical evidence. No technique was found clearly superior since the results depend on many varying factors. We identified a need for empirical studies where concepts are evaluated rather than small variations in technical implementations.

Contents

* Corresponding author. Tel.: +46 46 222 93 25.
E-mail addresses: emelie.engstrom@cs.lth.se (E. Engström), per.runeson@cs.lth.se (P. Runeson), mats.skoglund@cs.lth.se (M. Skoglund).

0950-5849/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.infsof.2009.07.001

1. Introduction

Efficient regression testing is important, even crucial, for organizations with a large share of their cost in software development. It includes, among other tasks, determining which test cases need to be re-executed, i.e. regression test selection, in order to verify the behavior of modified software. Regression test selection involves a trade-off between the cost for re-executing test cases, and the risk for missing faults introduced through side effects of changes to the software. Iterative development strategies and reuse are common means of saving time and effort for the development. However both require frequent retesting of previously tested functions due to changes in related code. The need for efficient regression testing strategies is thus becoming more and more important.

A great deal of research effort has been spent on finding cost-efficient methods for different aspects of regression testing. Examples include test case selection based on code changes [1,6,13,17,20,22,43,49,62,64,67] and specification changes [38,40,54,68], evaluation of selection techniques [48], change impact analysis [44], regression tests for different applications, e.g. database applications [18], regression testing of GUIs and test automation [39], and test process enhancement [31]. To bring structure to the topics, researchers have typically divided the field of regression testing into (i) test selection, (ii) modification identification, (iii) test execution, and (iv) test suite maintenance. This review is focused on test selection techniques for regression testing.

Although techniques for regression test selection have been evaluated in previous work [3,15,36,65], no general solution has been put forward since no technique could possibly respond adequately to the complexity of the problem and the great diversity in requirements and preconditions in software systems and development organizations. Neither does any single study evaluate every aspect of the problem; e.g. Kim et al. [27] evaluate the effects of regression test application frequency, Elbaum et al. [11] investigate the impact that different modifications have on regression test selection techniques, several studies examine the ability to reduce regression testing effort [3,11,15,27,36,65,66] and to reveal faults [11,15,27,49].

In order to map the existing knowledge in the field, we launched a systematic review to collect and compare the existing empirical evidence on regression test selection. The use of systematic reviews in the software engineering domain has been subject to a growing interest in the last years. In 2004 Kitchenham proposed a guideline adapted to the specific characteristics of software engineering research. This guideline has been followed and evaluated [5,30,57] and updated accordingly in 2007 [29]. Kitchenham et al. recently published a review of 20 systematic reviews in software engineering during 2004–2007 [28].

Ideally, several empirical studies identified in a systematic review evaluate the same set of techniques under similar conditions on different subject programs. Then there would be a possibility to perform an aggregation of findings or even meta-analysis and thus enable drawing general conclusions. However, as the field of empirical software engineering is quite immature, systematic reviews have not given very clear pictures of the results. In this review we found that the existing studies were diverse, thus hindering proper quantitative aggregation. Instead we present a qualitative analysis of the findings, an overview of the existing techniques for regression test selection and of the amount and quality of empirical evidence.

There are surveys and reviews of software testing research published before, but none of these has the broad scope and the extensive approach of a systematic review. In 2004 Do et al. presented a survey of empirical studies in software testing in general [8] including regression testing. Their study covered two journals and four conferences over 10 years (1994–2003). Other reviews of regression test selection are not exhaustive but compare a limited number of chosen regression test selection techniques. Rothermel and Harrold presented a framework for evaluating regression test techniques already in 1996 [48] and evaluated the, by that time, existing techniques. Juristo et al. aggregated results from unit testing experiments [25] of which some evaluated regression testing techniques, although with a more narrow scope. Binkley et al. reviewed research on the application of program slicing to the problem of regression testing [4]. Hartman et al. reported a survey and critical assessment of regression testing tools [21]. However, as far as we know, no systematic review on regression test selection research has been carried through since the one in 1996 [48]. An early report of this study was published in 2008 [12], which here is further advanced especially with respect to the detailed description of the techniques (Section 3.4), their development history and the analysis of the primary studies (Section 3.5).[1]

This paper is organized as follows. In Section 2 the research method used for our study is described. In Section 3 the empirical studies and our analyses are reported. In Section 4 the results are discussed, and in Section 5 the work is concluded.

2. Research method

2.1. Research questions

This systematic review aims at summarizing the current state of the art in regression test selection research by proposing answers to a set of questions below. The research questions stem from a joint industry-academia research project, which aims at finding efficient procedures for regression testing in practice. We searched for candidate regression test selection techniques that were empirically evaluated, and in case of lack of such techniques, to identify needs for future research. Further, as the focus is on industrial use, issues of scale-up to real-size projects and products are important in our review. The questions are:

(RQ1) Which techniques for regression test selection in the literature have been evaluated empirically?
(RQ2) Can these techniques be classified, and if so, how?
(RQ3) Are there significant differences between these techniques that can be established using empirical evidence?
(RQ4) Can technique *A* be shown to be superior to technique *B*, based on empirical evidence?

Answers to these research questions are searched in the published literature using the procedures of systematic literature reviews as proposed by Kitchenham [29].

2.2. Sources of information

In order to gain a broad perspective, as recommended in Kitchenham's guideline [29], we searched widely in electronic sources. The advantage of searching databases rather than a limited set of journals and conference proceedings is also empirically motivated by Dieste et al. [7]. The following seven databases were covered:

- Inspec (<www.theiet.org/publishing/inspec>).
- Compendex (<www.engineeringvillage2.org>).

[1] In this extended analysis, some techniques that originally were considered different ones, were considered the same technique. Hence, the number of techniques differ from [10]. Further, the quality of two empirical studies was found insufficient in the advanced analysis, why two studies were removed.