Experimenting with a Multi-Iteration Systematic Review in Software Engineering

Fabiano Cutigi Ferrari[1]*, **José Carlos Maldonado**[1]

[1]Computing Systems Department – University of São Paulo
Caixa Postal: 668 – 13560-970 – São Carlos – SP – Brazil

{ferrari,jcmaldon}@icmc.usp.br

Abstract. ***Context:*** *Systematic reviews promise to become a common practice within the Software Engineering (SE) domain, having the replicability as an important requirement that allows for result contrasting and updates.* ***Objectives:*** *To provide background in systematic review updating process by reporting our experience in performing a multi-iteration systematic review within the SE domain.* ***Method:*** *We updated the original review twice based on an adapted three-phase process for systematic reviews. We discuss the main issues and present the lessons learned.* ***Results:*** *Following the adapted process led us to update the original review successfully, even considering the lack of background in review updates and the limited supporting mechanisms.* ***Conclusions:*** *Replicability of systematic reviews is feasible specially thanks to the rigorous planning process.*

1. Introduction

Systematic review is a rigorous, well-defined approach for identifying, evaluating and interpreting all available research with respect to (w.r.t.) a particular topic of interest [Kitchenham 2004]. It represents an underlying, largely used practice in the evidence-based medicine domain [Kitchenham et al. 2004]. Recently, Kitchenham [2004] provided guidelines for performing systematic reviews in the Software Engineering (SE) field, based on a set of existing guidelines for medical research. There is a belief that SE researchers and practitioners may benefit from the evidence-based Software Engineering (EBSE) approach [Kitchenham et al. 2004] to produce more reliable, better-founded results based on empirical evidences collected in a systematic fashion.

Since then, despite of still not being widely adopted by the SE research community [Brereton et al. 2007], we can observe a growth in the number of reported experiences and results of systematic reviews in SE-related subjects. A quick look at online repositories such as IEEE Xplore and ACM Digital Library reveals a number of related publications. Moreover, we can find several discussions and evaluations concerning systematic reviews in SE [Bailey et al. 2007, Brereton et al. 2007, Dybå et al. 2007, Dieste et al. 2008b].

Some important requirements for a systematic review are transparency, replicability and auditability [Kitchenham 2004, Biolchini et al. 2005]. They are strongly dependent on the review documentation, which must clearly include an overview of the procedures and enough details of the selected studies. Omitting these elements may risk all the process. In spite of that, reporting experiences of updating systematic reviews has not been given the required attention to date. Feedback on whether and how researchers and practitioners have

*Financially supported by FAPESP (Process 05/55403-6) and CAPES (Process 0653/07-1), Brazil.

been replicating and updating systematic reviews would be of interest in several ways. For example, it could be used to:

- provide researchers and practitioners with guidance on how to plan and undertake systematic review updates and replications;
- help researchers identify weaknesses of current guidelines and processes; and
- help service providers improve current search engine capabilities and publication databases (e.g. the type of information that is made available).

This paper reports our experience in conducting and updating a systematic review in the software testing context. We call it a *multi-iteration systematic review*, which so far has consisted of the original systematic review (hereafter called simply "original review") and two updates. We call "*iteration*" every full cycle of a systematic review, from the planning to the review documentation. Notice that we do not focus on the review results themselves (e.g. number and quality of primary studies). Instead, we use the review as an example throughout this paper to support our discussions specially regarding replicability issues.

We planned our review according to available guidelines and templates [Kitchenham 2004, Biolchini et al. 2005]. After finishing the first iteration, we figured out that running regular updates could provide us with some interesting insights about the replicability of planned reviews. Our experience shows that replicability is a feasible requirement even in the presence of limitations, which includes the lack of knowledge about the updating process. We also highlight the poor support most of the search engines currently provide for composing search strings and refining search scopes. However, some search engines have evolved lately towards providing better support for searching activities.

We expect this experience report can somehow contribute to the establishment of well-defined processes for performing systematic reviews, as well as to the improvement of current resources such as search engines and publication databases. In addition, we expect to help researchers identify a set of variables that might impact the final results of a review replication, which may include adequate planning, quality of selected databases (and related search engines) and researcher expertise.

The remaining of this paper is organised as follows: Section 2 presents some related work. Section 3 introduces the example we use throughout the paper. Section 4 presents the main steps of the iterations we have performed. Our discussions are presented in the Section 5. Finally, Section 6 brings our conclusions and future work.

2. Related Work

Brereton et al. [2007] report their experience of conducting three systematic reviews: one complete and two partially finished. Among the lessons learned, we highlight the problems the study participants faced while composing search strings to submit to different engines. The problems are similar to the ones we have found for an overlapping subset of search engines (namely, IEEE Xplore, ACM Digital Library and ScienceDirect). Similarly, Dybå et al. [2007] report their experience gained from a systematic review of agile software development though focusing the discussion on the quality criteria for study selection.

Bailey et al. [2007] analyse the overlap of results that may occur while performing searches for SE-related publications using different search engines. After running three systematic searches, they found out that there is little overlap due to two main reasons: either different databases are targeted or the search engines are incomplete. Moreover, they suggest that non-standardised publication meta-data and different keyword taxonomies also help narrowing the search results, reducing the overlap. Although the authors suggested these limitations may harden study replications, they did not run any replication to realise which problems may in fact arise while performing a new review iteration.

In spite of being a very important requirement of systematic reviews, replication-related issues have not been tackled by most of the authors who have reported systematic review experiences lately. Aware of that, Dieste et al. [2008] introduced a systematic review process which comprises updating tasks. The process resulted from the author's experience of updating an original review, whose results will be soon available [Dieste et al. 2008b]. The authors underwent some difficulties while conducting the new iteration, mainly regarding data extraction and aggregation. Although the authors claim that the process formalises a review updating process, it only gives little emphasis on updating activities. They only briefly discuss how to manage individual primary studies[1] and their contents (e.g. how to treat the new studies independently but still considering the whole set of results) and how to aggregate the extracted data following the original procedures.

3. Running Example

This section introduces the example we use throughout this paper and upon which we perform our discussion and draw our conclusions. The example consists of a systematic review of Aspect-Oriented (AO) software testing which has been originally performed and then updated twice over the past 30 months. The first iteration (hereafter also called "original review") occurred between May and July, 2006, and results were already reported [Ferrari and Maldonado 2006, Ferrari and Maldonado 2007]. The two updates were performed in the subsequent years, at intervals of 13 months (i.e. August 2007 and September 2008). Next we briefly present the review planning (mainly objectives, research questions and search repositories).

3.1. Overview of the Original Protocol

We planned our systematic review in conformance to the template defined by Biolchini et al. [2005] and the guidelines provided by Kitchenham [2004] . Our goals were two-fold. First, identifying which software testing approaches researchers and practitioners have been applying to AO software. Second, identifying fault types that are specific to AO software and have been characterised to date. We defined one primary and three secondary research questions as follows:

- Primary question: *Which testing techniques/criteria have been applied to AO software to date?*.
- Secondary question #1: *Which of these techniques/criteria are specific to AO software?*
- Secondary question #2: *Which AO fault types have been described to date?*
- Secondary question #3: *Which kinds of experimental studies have been performed in order to validate AO software testing approaches?*

Our search strategy can be summarised as follows: studies of interest would be searched at online publication repositories and general purpose search engines. We have

[1] The term "primary study" has so far been used in the EBSE domain to describe a variety of research results, from well-founded experimental procedures to incipient research approaches. On the other hand, in the evidence-based paradigm, systematic reviews are considered "secondary studies".

targeted more restricted search engines as, for example, the IEEE Xplore and the ACM Digital Library, semi-restricted engines such as Scirus, and Google as a general purpose search engine, but not limited to them. The main search terms we used to compose our search string are "*aspect-oriented*" and "*testing*". The general review procedures (e.g. preliminary selection, final selection, data extraction and review documentation) follow the Biolchini et al.'s template. More details about how we performed these tasks appear in this paper as long as they are required. The reader can also refer to the original protocol template [Biolchini et al. 2005] and also to our previous work [Ferrari and Maldonado 2007].

4. Performing Multiple Iterations of Systematic Reviews

This section reports the process of performing a multi-iteration systematic review. It is based on the experience we acquired while performing a three-iteration systematic review on AO software testing, briefly introduced in the previous section. We first describe the first iteration (or original review) in terms of the process we followed and some specific issues that we observed while planning and executing the activities. In the sequence, we report our experience of performing two updates on the original review, focusing on the required adaptations to the original process.

4.1. The Original Review

The Process: According to Section 3, our review protocol was designed based on the template proposed by Biolchini et al. [2005] . In addition to the template, they proposed a three-phase process for systematic reviews which is shown in Figure 1. The main activities each phase comprises are: (1) *Planning:* definition of objectives and protocol. (2) *Execution:* identification of primary studies (searches), preliminary selection, evaluation of primary studies and final selection. (3) *Result analysis:* extraction and synthesis of data. Notice that *Packaging* is not considered a phase but an activity that is performed through the whole process in order to support the storage of intermediary and final results.

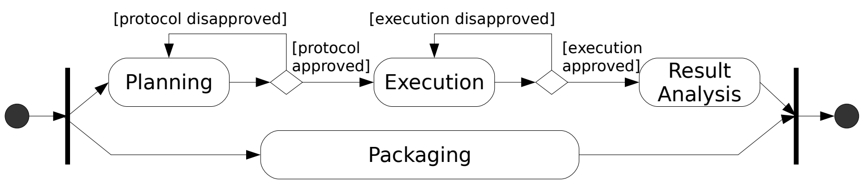

Figure 1. The process for systematic reviews defined by Biolchini et al. [2005].

Planning and Executing the Review: During the planning phase, one important decision comprised the selection of sources of primary studies. Such decision should take into account the objectives and the cost constraints. As far as possible, we selected a reasonable set of sources that included the IEEE, the ACM and the Springer's repositories. In addition, we included the Scirus and Google search engines as a mean of broadening the search results, and also consulted leading researchers in the AO software testing domain whose knowledge could contribute for the identification of extra relevant research.

The identification of primary studies basically relied on the search engines returning the expected results. We defined the following search string[2]:

[2] In fact, this is an updated string used in our last two iterations, which does not include redundant terms.

```
(aspect-oriented software OR aspect-oriented application OR aspect-
oriented app OR aspect-oriented program OR AOP) AND (testing OR fault
OR defect OR error OR incorrect OR failure)
```

We were aware that limitations of some search engines could risk the search process and consequently the final results. This led us to adopt alternative strategies in face of problems while running the searches. For example, defining shorter strings that included the most relevant terms (e.g. “aspect-oriented” and “testing”) and running some pilot searches aiming at figuring out the *modus operandi* of each search engine. The two checkpoints shown in Figure 1 allowed us for going backwards and forwards through the process phases when any adjustment or activity re-execution were required.

Another issue regarding the preliminary study selection was that for some studies the decision could not be made only based on the title and abstract reading. Therefore, an alternative approach was followed: scanning the whole document in order to find individual search terms. In cases where the occurrence of such terms provided us with any hint about the relevance of a study, we fully read the section which contained the terms and then decided for its inclusion or exclusion.

The remaining activities of the original review (evaluation of primary studies, final selection, and extraction and synthesis of data) were performed according to the protocol. The data extraction form showed to be suitable for our purposes and the review was already documented [Ferrari and Maldonado 2006, Ferrari and Maldonado 2007]. More details about how we dealt with search engine limitations are discussed in Section 5.

4.2. Performing New Iterations

While planning our first update on the original review, we realised that we would need to adapt the original process [Biolchini et al. 2005] in order to succeed. The adaptation resulted in the process shown in Figure 2. In this figure, the Execution and the Result Analysis phases (the two large rounded boxes with grey background) are split into the respective activities. The rounded boxes with white background represent the activities we have either modified or introduced into the process, namely *Planning Update*, *Filtering Results* and *Merging Results*. The remaining activities are the same as defined in the original process (Figure 1) and appear with grey background.

We next present the changes we made in order to make the process suitable for our purposes. We also report some issues we dealt with while performing the updates.

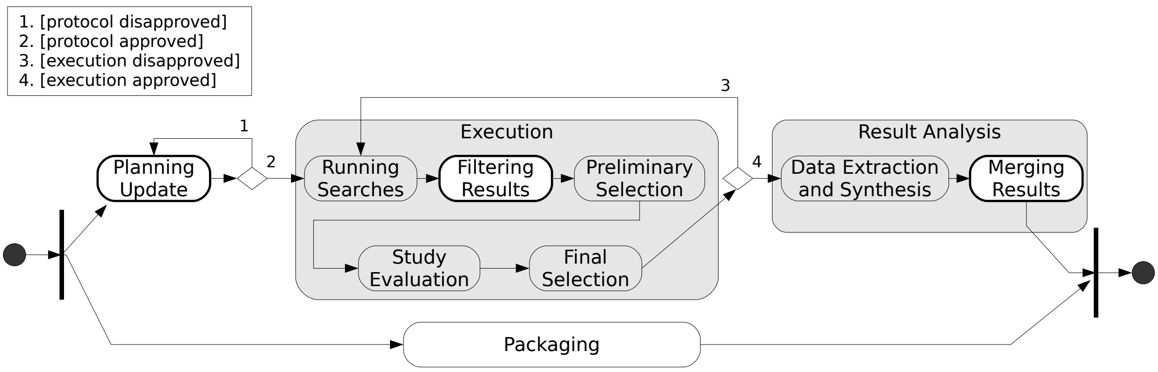

Figure 2. The process for systematic review updates.

Planning the Update: The protocol was revisited in order to identify required changes to make it suitable for a new, updating iteration of our review. After analysing the original research questions and procedures, we realised that the inclusion and exclusion criteria required changes. This was due to the need for restricting the search only for studies published since 2006 (for the first update) and since 2007 (for the second update). Doing so, we could reduce the overlap among results considering the set of studies retrieved in the previous iterations. Therefore, we added the statement "*published since <YEAR>*" to each inclusion criteria and "*published before <YEAR>*" to each exclusion criteria, where "*<YEAR>*" meant the current year of the review update.

Still regarding the planning activity, we accepted the suggestion of an independent expert to also consider the ScienceDirect search engine. This engine accesses the Elsevier's repositories and has been included since the first review update. The reader should notice that any incorrect change in the inclusion or exclusion criteria may risk the goals of the review update, leading to incomplete or inconsistent results.

Filtering the Results: After running the searches and getting the lists of publications, we manually "filtered" the results in order to disconsider overlapping subsets of studies to be analysed afterwards. Due to the nature of the studies of interest for our systematic review (i.e. studies related to AO software testing), applying such filter was not very time-consuming. However, tooling support is desirable when considering larger result sets.

This activity can be seen as a complement of the *Planning Update* if we consider that a filter were also applied to the inclusion and exclusion criteria. However, search engines generally only allow the user to filter the searches by year of publication. Considering that our original review and consecutive updates were not performed in the beginning of each year, overlapping results could not be avoided. In addition, not all studies are made available right after they are published. Therefore, we believe this manual selection of studies is likely to be required in most of cases.

Merging the Results: This activity basically consists of including into the data set the data extracted from studies selected in the current iteration. Once the previously extracted data has been stored in a standardised fashion, this same standard should be applied to the new extracted data. Then, both data sets can be merged into a single one.

The *Merging Results* activity was partially performed for our two review updates, since we only extracted and merged part of the data of interest, mostly regarding aspect-oriented fault types and testing criteria. Such data has already been used to underlie other research initiatives [Ferrari et al. 2008], and we plan to finish this activity in a near future.

5. Discussion

This section brings our discussions about two central issues we observed while performing our review. First, we analyse some characteristics of the selected search engines. Our analysis is drawn from a specific point of view, i.e. how suitable these engines are in order to support SE-related systematic reviews. Moreover, we also discuss how they have evolved through the last few years and how they have behaved for replicated searches. Notice that we do not intend to present a comprehensive evaluation regarding, for example, usability or performance criteria, since this is not a goal of our work. In the sequence, we report some insights and lessons learned regarding the replicability of systematic reviews. Doing so, we expect to contribute to other updating-related initiatives and also for process improvements.

5.1. Suitability of Search Engines

Tables 1 and 2 summarise some characteristics for each search engine according to the observations we made during the three iterations. We believe these characteristics are relevant for the searching process. They are following described and, when necessary, discussed in terms of some noticed issues for specific search engines, such as limitations and evolutions. Note that evolutions are highlighted in bold in the tables.

Table 1. Search engines – characteristics and evolution (1 of 2)

	IEEE Xplore			ACM Digital Library			SpringerLink		
Characteristic	2006	2007	2008	2006	2007	2008	2006	2007	2008
1. Centralised results	yes	yes	yes	no	no	no	yes	yes	yes
2. Full string composition	yes	yes	yes	no	no	**yes**	no	no	**part**
3. Adequate scope narrowing	yes	yes	yes	yes	yes	yes	no	**yes**	yes
4. Consistent results	yes	yes	yes	yes	yes	yes	no	**yes**	yes

Table 2. Search engines – characteristics and evolution (2 of 2)

	ScienceDirect			Scirus			Google		
Characteristic	2006	2007	2008	2006	2007	2008	2006	2007	2008
1. Centralised results	—	yes	yes	no	no	no	no	no	no
2. Full string composition	—	yes	yes	part	part	**yes**	no	no	no
3. Adequate scope narrowing	—	yes	yes	no	no	no	no	no	no
4. Consistent results	—	yes	yes	no	no	no	no	no	no

Centralised results: Results are considered *centralised* when all the retrieved publications come from a central repository. According to tables 1 and 2, results gathered from the IEEE Xplore, SpringerLink and ScienceDirect engines are centralised since they only come from IEEE, Springer and Elsevier's repositories, respectively. On the other hand, engines such as ACM and Scirus are not restricted to a single source. For example, while performing a search with the ACM Digital Library engine, the user can also find papers published at the IEEE and Springer's repositories, as well as with Scirus and Google.

Full string composition: A search string is considered *full* when it includes all the terms of interest, combined by means of logical operators and nested expressions. Moreover, it is considered *partially full* ("part" in tables 1 and 2) when there are restrictions for: (i) the number of search terms; or (ii) the length of the string, regardless the number of terms.

The second line of tables 1 and 2 shows that, on the one hand, we have only the IEEE Xplore and ScienceDirect apparently not imposing limits for search string composition. On the other hand, we were required to use alternative approaches to overcome string composition limitations. For example, in the first two iterations, we had to compose smaller strings (e.g. $<$ `+abstract:"aspect-oriented" +abstract:testing` $>$ and $<$ `+abstract:"aspect-oriented" +abstract:fault` $>$) to be submitted to the ACM Digital Library engine and then run multiple searches in order to obtain reasonable results. As a consequence, we could observe overlaps among the retrieved list of publications, which in turn required extra effort during the preliminary selection. However, we noticed that the ACM search engine currently allows the user to compose full search strings, reducing the search to a single run and the results to a single set. The Scirus engine has also shown some evolution, since it allowed for a long search string instead of a truncated 255-character string as in the first two iterations.

Adequate scope narrowing: a search engine supports *adequate scope narrowing* when it allows the user to specify which parts of a document is going to be scanned. For example, users should be able to narrow their search to titles and abstracts of publications. Moreover, once such parts are defined, the engine should behave as expected, i.e. limiting the scanning only to these parts.

Line 3 of tables 1 and 2 shows that only Scirus and Google do not support scope narrowing. This is due to the nature of these search engines, hence it should not be seen as a limitation. However, researchers and practitioners should have in mind that the poor support for scope narrowing will certainly require more time for the preliminary study selection, since they will have to manually analyse each study part according to the protocol.

We can also see that the SpringerLink engine has evolved since the second iteration. During the first search, although scope narrowing was allowed according to engine specification, full documents were analysed. This problem seems to be solved as we noticed during the next iterations.

Consistent results: Results are considered *consistent* when, once the search string is defined, they conform to the following requirements: (i) every run with this string results in the same set of publications w.r.t. a specific search engine; and (ii) composed search terms are treated as atomic terms (i.e. *"term1 term2"* is different from *"term1" and "term2"*). Notice that "same set of publications" may not necessarily be true in case new publications have been inserted into a repository. Nevertheless, a late search should include all the results from a former one.

Tables 1 and 2 show that more specific search engines such as IEEE Xplore and ScienceDirect tend to bring more consistent results. On the other hand, less restricted and general purpose engines such as Scirus and Google present varied result sets when a search is replicated. This is likely to happen because these engines target not only publication repositories but also Web pages generally.

Still regarding consistency of results, during the first iteration of our review, the SpringerLink engine did no behave as expected. Compound search terms were not considered atomic. For example, searches for "aspect-oriented" also retrieved publications containing "aspect" and "oriented" not necessarily together. Therefore, extra effort was required for preliminary study selection. During the second and third iteration, however, this problem seemed to be solved and the engine behaved as expected.

5.2. Replicability of Planned Reviews

Replicability is a fundamental requirement for a systematic review [Kitchenham 2004]. It ensures that a secondary study can be re-executed following the original (possibly updated) protocol and, later on, be contrasted to the original results. Moreover, replicability capabilities also support the updating process in order to keep the results up-to-date. We following report some insights and lessons learned while running our three-iteration systematic review, where the last two iterations consisted of replications that aimed at keeping the results updated. We focus our reporting on some specific topics comprising the planning process, search engines modifications and data extraction and synthesis.

Rigorousness of the planning process allows for replicability: The general guidelines for systematic reviews in SE [Kitchenham 2004] has shown to be adequate according to their purposes, resulting in well-defined protocols for which only minor changes are required when planning a review replication. This is a consequence of the rigour present during the definition of the research questions and all other involved elements. The protocol should be elaborated by all study participants and should be reviewed by independent experts. This process ensures that only minor changes might be required when a replication is considered to be undertaken. For example, in our updating plan, we were required only to add a new constraint (regarding the year of publication) to each inclusion and exclusion criteria, and also to include a new search engine (ScienceDirect), as shown in Section 4. This need for few adaptations was a consequence of the original plan being rigorously defined, and certainly contributed to simplify our review updating iterations.

Search engine modifications do not risk the replication process and results: It is known that application software is modified through the time, so are the search engines. This may require extra effort for the user to get adapted to the new operational model. For example, we noticed that some search engines have been modified during the last few years such that additional pilot searches were required to adapt our search strings to these modifications. However, all the noticed modifications comprised improvements, as we can see in tables 1 and 2. This led us to believe that current and future changes in search engines are likely to improve instead of risking the review updating process and results.

Standardised data extraction and synthesis help replications succeed: Data extraction and synthesis represent central issues in a systematic review and rely on the quality and accuracy of the data extraction forms. Such forms are designed in advance, and should provide means for the researchers to collect of all the relevant data according to the proposed objectives. Despite of their importance, data extraction and synthesis are still prone to subjectivity and to particular writing styles. Therefore, while replicating or updating a systematic review, researchers should follow as much as possible the original fashion. Doing so, they can achieve more standardised results, in turn reducing the likelihood of misunderstandings.

During our updating iterations, data extraction and synthesis, added to the merging activity, were straightforward. This might be due to the small number of researchers involved in the study, which may ease the whole process. However, we believe that as the number of involved researchers grows, the protocol should provide more detailed information about how to fill in the data extraction forms, which should be strictly followed.

6. Conclusions and Future Work

Recently, we could observe a growing interest in the evidence-based Software Engineering, particularly in systematic reviews. It has motivated the definition of guidelines, templates and processes for systematic reviews [Kitchenham 2004, Biolchini et al. 2005, Dieste et al. 2008a]. However, little emphasis has been given on the replicability property, which is a very important requirement that allows the researcher to reproduce or update a review and then to contrast or merge the new results with previously reported results.

This paper reported our experience of performing a multi-iteration systematic review within the SE domain. We focused on reporting the process of conducting the review itself instead of reporting the achieved results. The review process was based on a three-phase process that we have adapted in order to include updating-related activities.

Performing and updating our systematic review has provided us with some insights and also has helped us identify the limitations of currently available search engines. However, such limitations have not risked our review results specially because of the well-defined, rigorous planning which has been adequately adjusted and strictly followed. When it comes to former systematic reviews, we do not believe that the search engine limitations have risked the results since they tend to broaden the resulting lists of studies, which in turn requires extra effort during the preliminary study selection. However, building either comprehensive or at least reasonable search strings may be tricky and relies on the researcher's ability of doing so, supported by pilot searches.

Our future research includes the full documentation of our review in the form of a paper, comprising the original and the updated results. We also plan to keep the results up-to-date by performing regular updates according to the adapted three-phase process. This will provide important feedback and hints for process improvement and refinements.

References

Bailey, J., Zhang, C., Budgen, D., Turner, M., and Charters, S. (2007). Search engine overlaps: Do they agree or disagree? In *REBSE'2007 - in conjunction with ICSE'2007*, Minneapolis - USA. IEEE Comp. Society.

Biolchini, J., Mian, P. G., Natali, A. C. C., and Travassos, G. H. (2005). Systematic review in software engineering. Tech. Report RT-ES 679/05, Systems Engineering and Computer Science Dept., COPPE/UFRJ, Rio de Janeiro - Brazil.

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., and Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4):571–583.

Dieste, O., López, M., and Ramos, F. (2008a). Formalizing a systematic review updating process. In *SERA'2008*, pages 143–150, Prague - Czech Republic. IEEE Comp. Society.

Dieste, O., López, M., and Ramos, F. (2008b). Obtaining well-founded practices about elicitation techniques by means of an update of a previous systematic review. In *SEKE'2008*, pages 769–772, Redwood City - USA. Knowledge Systems Institute Graduate School.

Dybå, T., Dingsøyr, T., and Hanssen, G. K. (2007). Applying systematic reviews to diverse study types: An experience report. In *ESEM'2007*, pages 225–234, Madrid - Spain. IEEE Comp. Society.

Ferrari, F. C. and Maldonado, J. C. (2006). A systematic review on aspect-oriented software testing. In *WASP'2006*, pages 101–110, Florianópolis/SC - Brasil. (in Portuguese).

Ferrari, F. C. and Maldonado, J. C. (2007). Aspect-oriented software testing: A systematic review. Tech. Report 291, ICMC/USP, São Carlos - Brasil. (in Portuguese).

Ferrari, F. C., Maldonado, J. C., and Rashid, A. (2008). Mutation testing for aspect-oriented programs. In *ICST'2008*, pages 52–61, Lillehammer - Norway. IEEE Comp. Society.

Kitchenham, B. (2004). Procedures for performing systematic reviews. Joint Tech. Report TR/SE-0401, Department of Computer Science - Keele University and National ICT Australia Ltd, Keele/Staffs-UK and Eversleigh-Australia.

Kitchenham, B. A., Dybå, T., and Jørgensen, M. (2004). Evidence-based software engineering. In *ICSE'2004*, pages 273–281, Edinburgh - Scotland. IEEE Computer Society.