Master Thesis
Software Engineering
Thesis no: MSE-2012:97
September 2012

On the quality of grey literature and its use in information synthesis during systematic literature reviews

- Master Thesis Report

Affan Yasin & Muhammad Ijlal Hasnain

School of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona
Sweden

This thesis is submitted to the School of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Software Engineering. The thesis is equivalent to 2 x 20 weeks of full time studies.

Contact Information:
Authors:
Affan Yasin
afya10@student.bth.se

Muhammad Ijlal Hasnain
Muhf10@student.bth.se

University advisor:
Dr. Richard Torkar
Department of Software Engineering

School of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona
Sweden

Internet : www.bth.se/com
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

ABSTRACT

Context: The internet has become a vital channel for disseminating and accessing scientific literature for both the academic and industrial research needs. Nowadays, everyone has wide access to scientific literature repositories, which comprise of both "white" and "grey" literature. The "grey" literature, as opposed to "white" literature, is non-peer reviewed scientific information that is not available using commercial information sources such as IEEE or ACM. A large number of software engineering researchers are undertaking systematic literature reviews (SLRs) to investigate empirical evidence in software engineering. The key reason to include grey literature during information synthesis is to minimize the risk of any bias in the publication. Using the state of the art non-commercial databases that index information, the researchers can make the rigorous process of searching empirical studies in SLRs easier. This study explains the evidence of grey literature while performing synthesis in Systematic Literature Reviews.

Objectives: The goals of this thesis work are,

1. To identify the extent of usage of Grey Literature in synthesis during systematic literature reviews.
2. To investigate if non-commercial information sources primarily Google Scholar are sufficient for retrieving primary studies for SLRs.

Methods: The work consists of a systematic literature review of SLRs and is a tertiary study and meta-analysis. The systematic literature review was conducted on 138 SLRs' published through 2003 until 2012 (June). The article sources used are IEEEXplore, ACM Digital Library, SpringerLink and Science Direct.

Results: For each of the selected article sources such as ACM, IEEEXplore, SpringerLink and Science Direct, we have presented results, which describe the extent of the usage of grey literature. The qualitative results discuss various strategies for systematic evaluation of the grey literature during systematic literature review. The quantitative results comprise of charts and tables, showing the extent of grey literature usage. The results from analysis of Google Scholar database describe the total number of primary studies that we are able to find using only Google Scholar database.

Conclusion: From the analysis of 138 Systematic Literature Reviews (SLRs'), we conclude that the evidence of grey literature in SLRs is around 9%. The percentage of grey literature sources used in information synthesis sections of SLRs is around 93.2%. We were able to retrieve around 96 % of primary studies using Google Scholar database. We conclude that Google Scholar can be a choice for retrieving published studies however; it lacks detailed search options to target wider pool of articles. We also conclude that grey literature is widely available in this age of information. We have provided guidelines in the form of strategies for systematic evaluation of grey literature.

Keywords: Systematic Literature Review, SLR, Grey literature, Google Scholar.

Acknowledgements

We are thankful to Almighty Allah who gave us this opportunity and strength to complete this study to the best of our efforts.

We would like to thank our research advisor **Dr. Richard Torkar** for his continuous guidance and encouragement throughout our thesis work. His brisk responses to our questions were constructive, detailed and his positive attitude was a driving force behind this study.

We also thank our parents for their moral and financial support during our stay in Sweden. We are thankful to Software Engineering researchers **Dr.Tore Dybå** (Chief Scientist and research manager at SINTEF), **Professor Martin Shepperd** (Brunel University), **Dr.Marco Torchiano** (Associate Professor at Politecnico di Torino), **Dr.Siffat Ullah Khan** (Head of Department of Software Engineering University of Malakand), **Dr.Sira Vegas** (Professor at the Computing School – UPM), **Saad Bin Saleem** (PhD Student at The Open University,UK)and **Dr.Wasif Afzal** (Assistant Professor at Bahria University) for contributing in our study with additional literature, along with this special thanks to **Dr.Emilia Mendes** (Associate Professor at Zayed University), **Dr Robert Glass** and **Dr.Cigdem Gencel** (Assistant Professor / Senior Researcher at Free University of Bozen-Bolzano) for their guidance through email.

Lastly, we thank our friends and colleagues, who were helpful and encouraging throughout our study.

Table of Contents

List of Tables

List of Figures

Glossary

We/ authors refer to Affan Yasin and Muhammad Ijlal Hasnain, writer of this thesis work. Researchers refer to general researchers all over the world. The terms “systematic review” and “systematic literature reviews” are interchangeable in this study context.

1 INTRODUCTION

1.1 Background

Over the past decade, the internet has emerged as an essential source of information for everyone [1]. It has become the first source to which people normally turn if they need any information. In scientific community, academic researchers are now equipped with state of the art sources of scientific articles and meta-data research tools for their research. The online presence of scientific communities, discussion boards and blogs owned by notable authors is an important source of up-to-date scientific information [2]. However, most of the information published in online communities, blogs and discussion boards is considered as "Grey" by the definition of Grey Literature.

The Grey Literature, by Luxembourg definition, is, "Iinformation produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body." [3]. In general, grey literature publications are volatile in nature and lack bibliographic controls such as place and date of publication, details of author and publisher. These tendencies of grey literature make it difficult to index and categorize it. The grey literature is often referred as "fugitive literature" as it is semi-published and difficult to locate [4] [16]. Grey literature includes reports, theses, conference proceedings, bibliographies, technical specification and government policy documents. Grey literature, though not peer-reviewed thoroughly, is still an important source of information [6]. It complements in a research in many different ways. Typically, at the start of a research endeavor, the first-hand information about a particular problem is collected through grey literature. This includes a quick search for the problem topic over the internet and further discussion with other research personnel [7] [17].

It is worthwhile to note that grey literature, although not peer-reviewed, is often produced by scholars and scientists of their respective fields and is of high quality and detail [6]. According to Soule and Ryan [7], grey literature is becoming a common means for information exchange because it is available on a timelier basis than literature published by commercial information sources. For instance, the conference papers are in access to public long before the published articles. Beside these traits, grey literature is focused, in-depth and up-to-date information about any topic [26]. The growth of internet has immensely broadened the access to grey literature [1] [14].

1.2 Grey Literature in Current Era

The internet has changed the overall publishing nature of scientific literature. It offers tools and channels for researchers as well as individuals to produce, publish and assess scientific information [8]. The internet has not only widened the access to grey literature but also produced new challenges for researchers, like how to include grey literature into research. The main challenge concerns the ability to provide reliable assistance in using grey literature and maintain authentication of grey literature [14]. A recent example of such challenge was faced by the journal, Science. A journal paper in which researchers have claimed to identify a set of genes that can predict human longevity with 77% accuracy, received rapid feedback and enough criticism just after an hour of online publication [5]. The online researchers were much skeptic about the environment and controls in which the study was conducted. Therefore, a large number of scientist authors discussed this specific paper in their blogs, twitter and other social media platforms. The same situation was experienced by the authors of a

paper in which they claimed that heralding bacteria use arsenic rather than phosphorus in their DNA backbone [4]. The online publication of this paper was followed by immense criticism from the scientific community.

The academic research endeavor often starts with a literature review. With abundant sources of information both producing grey and white literature, it is important to carefully examine the information sources while exploring for literature. The grey literature plays a vital role by giving the up-to-date and summarized first-hand information about the research problem [6]. This helps researcher to quickly familiarize about current developments in the research topic.

Franco-Dutch study analyzed 64 scientific articles for grey literature citations. The analysis of thousands of citation showed the proportion of grey literature cited in scientific publications [17]. According to the study, the relative importance of grey literature depends on the research fields. For instance, in medical and health sciences, there is a preference for journals and conventional information sources among researchers, while in other engineering sciences, grey literature sources are used around half of total sources cited [18].

In software engineering research, there is a continuous growth in the conducted empirical studies [9]. The researchers are adopting systematic approaches to produce summary of evidence for a particular research problem. The inclusion of grey literature as a source for primary studies is essential to control and minify the publication bias. The threat of publication bias arises while searching the primary studies for inclusion in SLR [7]. The term publication bias refers to the problem that the studies with positive results are most likely to be included as primary studies in systematic review than the studies with negative results. The concept of "positive" and "negative" is subject to the viewpoint of researcher and can lead to inflated findings [11]. Some of the strategies to tackle this issue are to scan for grey literature, conference proceedings and unpublished results by contacting colleagues and researchers [9].

The importance of including grey literature in systematic literature review demands a study to investigate the evidence of grey literature used in synthesis during systematic literature reviews.

As stated above, the inclusion of grey literature is necessary to remove publication bias in systematic literature reviews [13] [15]. Acquiring grey literature is considered a tiresome task. But, fortunately, with the growth of internet, we have efficient search engines for searching scholarly articles. We can now control and refine our search results more effectively. Google Scholar (GS) provides a simple interface for searching scholarly literature over internet. The search results are from many different sources such as journals, theses, books, abstracts, reports, proceedings, literature published on university websites and from lots of online literature repositories. GS ranks the documents on the basis of different attributes like where the article was published, who are the authors and how recently the article has been cited by other scholars [28]. GS indexes both "white" and "grey" literature published around the web from peer reviewed journal articles to lecture notes and documents published through universities.

The process of selecting primary studies for systematic literature review can be very laborious, time-consuming and rigorous [27]. Manual searches are conducted on different information sources to pile up primary studies. On the other hand, we have Google Scholar that retrieves results from all major databases and orders them on the basis of different attributes described above [28].

It will be interesting to know if researchers can rely only on Google Scholar for searching research articles instead of searching separately in each of the databases. The results may help in saving a lot of effort that researchers exert in collecting scholarly content for their research.

1.3 Research Problem

The internet is transforming the whole value chain of publishing by offering tools and channels for disseminating and assessing grey literature in forms of research blogs, discussion boards and social media [3]. The inclusion of grey literature is inevitable for minimizing publication bias during systematic literature reviews. The importance of including grey literature in systematic literature review demands a study to investigate the evidence of grey literature being used in synthesis during systematic literature reviews.

1.4 Aims and Objectives

The overall aim of conducting this study is to find how much grey literature has been used (in percentage) in systematic literature reviews in software engineering and explore if it is feasible to only use Google Scholar for finding scholarly articles for academic research.

The aim was achieved by accomplishing the following objectives.

- Identify the strengths and weaknesses of grey literature.
- Explore the usage of grey links/ literature in systematic literature reviews.
- Investigate and examine the Google Scholar database by performing a rigorous search for literature.
- Identify method(s) to evaluate the quality of grey literature.

1.5 Research Methodology

We have chosen systematic literature review methodology to investigate the evidence of grey literature in SLRs. Developing a review protocol is a critical element of systematic review. The review protocol helps in controlling bias and to provide scope for study replication [12]. We have developed the review protocol and then evaluated it iteratively. The understanding of problem increased with each step of the iteration and helped us to refine our search strategy and research questions.

"Most authors claimed that the rationale behind SLRs was the formalized, repeatable process for systematically searching a body of literature to document the state of knowledge on a particular subject." [30]

1.6 Suitability of Selected Methodology

The study is conducted as a systematic literature review based on guidelines proposed by Kitchenham. Systematic literature reviews are recommended methods for aggregating empirical evidence in software engineering [11] [12]. The SLR methodology is driven by using a predefined review protocol that aims to be unbiased by being auditable and repeatable [30]. There are two different types of SLRs; i) Conventional SLRs ii) Mapping Studies [31].

We have selected the conventional SLR type for this study. The conventional SLR approach aims to aggregate results related to a specific research question by ignoring the topic area. The research questions are like "Is TDD (Test Driven Development) methodology more effective at defect detection than BDD (Behavioural Driven Development)?" In systematic mapping studies, the focus is to find and classify evidence on a specific topic area. The research questions are like "What do we know about Integration Testing?" This type of questions may help to identify available literature about a topic area prior to conducting conventional systematic literature reviews. In our study, we are not concerned with the topic area of primary studies. We

are concerned with collecting SLRs of any topic area within Software Engineering to find the evidence of grey literature within the primary studies of our collected SLRs.

The other research methodologies i.e. surveys, experiments and case studies seem to be irrelevant to our study context. Surveys are conducted when the use of a technique has already taken place. In our study, we are not assessing any specific technique rather collecting overall evidence of grey literature in SLRs. Similarly, case studies are most suitable for industrial evaluations. In our study, we are not evaluating any industry adopted practice. Our study is focused on SLRs published in academia.

For the data synthesis, we have chosen quantitative synthesis i.e. meta-analysis. An alternative would have been to use thematic analysis. Thematic analysis is used for qualitative analysis. In thematic analysis, themes are identified that reflect their textual data. In our opinion, the thematic analysis would lead to undesirable results in forms of data themes, which would be hard to synthesize for finding grey literature evidence. In our study, we are investigating the extent (in percentage) to which grey literature has been used and a conventional SLR with meta-analysis was considered as a most suitable methodology for this study.

1.7 Related Work

We have searched for all types of studies i.e. SLRs, case studies and experiments about grey literature evidence in Software Engineering. We have found that the usage of grey literature in Software Engineering has not been adequately investigated before and that is why we are unable to find appropriate related work studies. Although, we have retrieved around 120 results while searching for exact term "grey literature" in computer science/ software engineering sections of major databases (IEEE, Science Direct, ACM Digital and SpringerLink), we have found that most of the articles are irrelevant and discussed about grey literature traits rather than investigating its evidence in particular area.

To broaden our understanding about grey literature and its evidence, we have looked up for work performed in other research fields such as health sciences. We have found that grey literature has been given significant importance in health science field. In health sciences, research has been undertaken to investigate grey literature usage [37], assess the quality of grey trials [34] and if this literature inclusion in meta-analysis and systematic literature reviews is elementary [35] [36] [38]. The study by Hopewell, McDonald et al concluded that the published trials show an overall greater treatment effect than grey trial, therefore, researchers undertaking SLRs should search for trials in both the published and grey literature in order to help minimize the effects of publication bias [35]. The same type of results is reported by McAuley, Pham et al. [38] and later by Conn, Valentine et al [36] in their study about meta-analyses. Their studies conclude that meta-analyses that exclude grey literature likely over-represent studies with statistically significant findings, inflate effect size estimates, lead to exaggerated estimates of intervention effectiveness, and provide less precise effect size estimates than meta-analyses including grey literature.

Perhaps, the most notable work in grey literature analysis has been done by Farace and Schopel. They conducted a citation analysis study [33] and a follow-up survey study [18] while analyzing grey literature in publications from various science disciplines. They performed analysis on nearly thousand citations and concluded that the relative importance of grey literature is largely dependent on research disciplines and subjects, methodological approaches, and sources used. They also observed that in some fields, especially the life sciences and medical sciences, there has been a traditional preference for conventional distribution media (journals), while in others, such as agriculture and the engineering sciences in general, grey literature resources

tend to predominate. These couple of studies contain comprehensive analysis of grey literature and added valuable initial input for our research.

1.8 Research Questions

The research questions are the driving force behind entire systematic literature review methodology [12]. The research questions are derived from the need of performing systematic literature reviews. The following research questions were derived from aims and objectives.

RQ1. What is the extent of usage of grey literature in systematic reviews?
RQ2. What are the strategies that can be used to categorize grey literature (non-peer reviewed)?
RQ2.1 What are the strengths and weaknesses of those categorization strategies?
RQ3. What is the extent of usage of grey literature in information synthesis (in order to reject or strengthen any arguments) ?
RQ4. To what extent does Google Scholar bring the same results as we get from commercial research databases such as ACM, IEEE, SpringerLink, and ScienceDirect ?
RQ4.1 To what extent does GS has indexed research papers?

1.9 Thesis Outline

This thesis report is structured based on following sections.

Chapter 1, Introduction, introduces the research problem and objective of this study.

Chapter 2, Systematic Literature Review, reports the systematic literature review undertaken in this study.

Chapter 3, Results and Analysis, presents the results and analysis of the systematic literature review.

Chapter 4, Strategies for Grey Litearture, presents the strategies for taking grey literature into account more systematically.

Chapter 5, Discussion, has a discussion about grey literature strengths weaknesses and the discussion about Google Scholar results.

Chapter 6, Conclusion, briefly describes the overall thesis and results.

Chapter 7, References, contains references used in this study.

Chapter 8, Appendix, contains figures, tables and other complimentary material of this thesis.

1.10 Terminology

The terminologies used in the report are tabulated below.

Table 1: Terminologies

Terms/Abbreviations	Definition
SLR(s)	Systematic Literature Review(s)
GL	Grey Literature
RQ	Research Question
GS	Google Scholar
EBSE	Evidence Based Software Engineering

2 SYSTEMATIC LITERATURE REVIEW

Systematic literature reviews (SLRs) are the recommended evidence based software engineering (EBSE) method for aggregating evidence [11] [12]. SLRs help to summarize the existing empirical evidence particular to a research question or topic area. A systematic review synthesizes existing work in a fair manner and reports the findings in a summarized format. The number of undertaken empirical studies in software engineering is on rise therefore; systematic literature reviews in software engineering are becoming more frequent [12].

2.1 Method

The SLR reported in this thesis is undertaken based on the original guidelines of performing systematic literature reviews proposed by Kitchenham [9]. The objective of conducting this systematic review is to find and summarize all existing evidence of grey literature found in systematic literature reviews in the field of Software Engineering. The secondary objective of this study is to investigate the Google Scholar, search engine, and see if we come up with as many primary studies as our SLRs do. Figure 1 illustrates the development of review protocol used in this study.

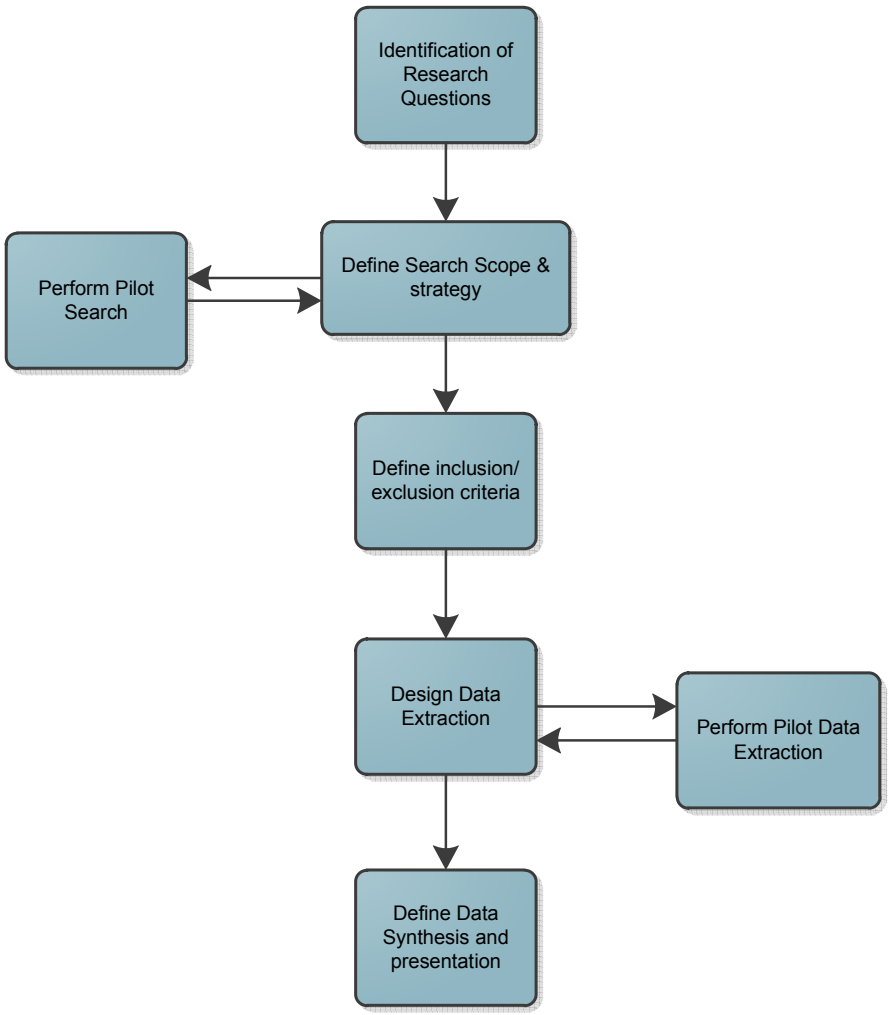


Figure 1: Development of Review Protocol

The review is classified as tertiary literature review since the goal of this study is to assess systematic literature reviews (that are referred to as secondary studies). The research questions addressed by this study are already presented in the introduction section of this document.

2.2 Search Strategy

Our search strategy was aimed to extract maximum number of systematic reviews from the selected databases. We examined our research questions to derive the initial search terms and keywords. We limited our search of articles published from year 2004 till 2012 (June) because the first guidelines for performing systematic literature reviews were proposed in year 2004 by Kitchenham [9]. Table 2 presents the research questions along with their motivation.

Table 2: Research Questions

ID	Research Question	Rationale
1	What is the extent of usage of grey literature in systematic reviews?	The aim is to investigate the evidence of grey literature in synthesis during systematic literature reviews.
2	What are the strategies that can be used to categorize grey literature (non-peer reviewed)?	The aim of this question is to come up with strategies that can help to consider grey literature systematically.
3	What are the strengths and weaknesses of those categorization strategies?	The question aims to explain the strengths and weaknesses of formed strategies in RQ2.
4	What is the extent of usage of grey literature in information synthesis (in order to reject or strengthen any arguments) ?	The aim is to investigate if selected grey primary studies are indeed used in synthesis sections of SLRs.
5	To what extent does Google Scholar bring the same results as we get from commercial research databases such as ACM, IEEE, SpringerLink, and ScienceDirect? To what extent does GS has indexed research papers?	The primary motivation is to investigate if we can use only Google Scholar to retrieve primary studies for research endeavours.

The main search strings were formed from the research questions by;

1. Altering the spellings, identifying alternative terms and synonyms of major search terms.
2. Consulting the keywords in already published SLRs.
3. Using Boolean OR for incorporating search terms of alternative spellings and synonyms.
4. Using Boolean AND to link the major terms with other terms and for combining different terms.

Thus, the finalized search keywords for this study are given below,

1. Systematic Review.
2. Systematic Literature Review.

3. Meta-Analysis.
4. Empirical Evidence.
5. Empirical Studies.
6. Empirical Study.
7. Empirical Studies OR Empirical Study
8. Systematic Review AND Kitchenham
9. Systematic Literature Review AND Kitchenham.
10. Meta-Analysis AND Kitchenham.
11. (Empirical Studies OR Empirical Study) AND Kitchenham.
12. “Systematic review” AND (Software Engineering) etc.

Figure 2: Search Strategy

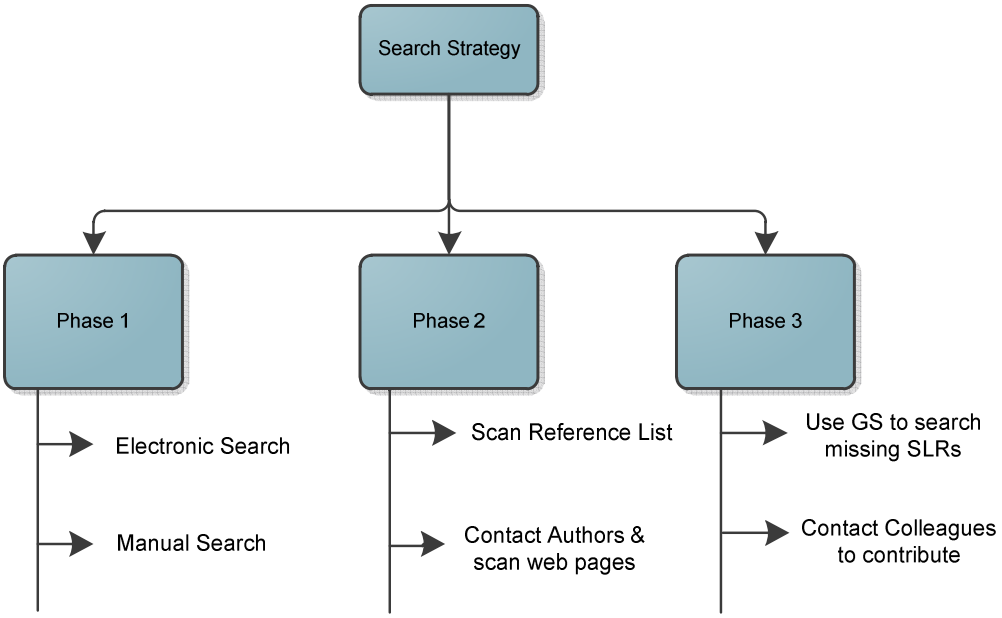


Figure 2 illustrates the different phases and steps of our search strategy. We made sure that each keyword should not retrieve more than 300 - 400 search results in one go. This was done by using Boolean operators in between the search terms. For instance, by using only the keyword “systematic”, we were able to retrieve nearly thousands of search results from each database, which made the result sample impractical for SLRs lookup. The goal was to find most efficient combination of keywords for retrieving maximum relevant SLRs from databases.

Before diving into the search process, we ensured theselection of proper and relevant keywords for our search strategy. We conducted a pilot search before the actual search process to verify the strength of search terms. This was an attempt to avoid time being wasted because of the inadequately designed keywords and to improve search keywords [13] [27]. After finalizing the pilot studies, we performed a search for the pilot studies. The rationale was that if we got more than 90% percent pilot studies using our keywords, then we could select our keywords; otherwise there was a need to improve them.

2.3 Selecting Pilot Studies

The pilot search was carried out on a total of 37 SLRs. These candidate studies were selected as follow;

- 22 studies were selected from the Kitchenham tertiary study "Systematic Literature Review in SE- SLR".
- 15 studies were selected from the publications by the faculty of Blekinge Institute of Technology (Dr. Tony Gorshek, Dr. Wasif Afzal, Dr. Richard Torkar, Dr. Darja Smite, Dr. Claes Wohlin, Dr. Robert Feldt etc).
- We also assured that in the pilot population, we have at least one SLR of each year from year 2004 to year 2012.

The selected pilot studies are heterogeneous and cover nearly all major software engineering disciplines i.e. Requirement Engineering, Global Software Engineering, Software Quality and EBSE (Evidence Based Software Engineering) etc. The pilot search has been conducted by the both authors individually. The inter rating agreement has been found through Cohen Kappa application to results.

2.4 Cohen Kappa Application

Cohen Kappa is a robust statistical measure for examining the agreement level among two authors (raters) [19]. Kappa is useful when all disagreements may be considered equally serious [20]. For Kappa usage, the number of raters should be two and the same two raters rate each subject. The Kappa coefficient value ranges from 0.0 to 1.0 with better reliability on the higher end.

In the pilot search, each author assessed 37 SLRs. The Cohen Kappa was performed to find inter-rater agreement among raters. The SPSS software was used to find the coefficient value. The value was found to be 0.63 which is considered as substantial agreement [20]. The findings are listed in Appendix A.

2.5 Selected Information Sources

The following databases were selected for this study.

1. ACM Digital Library
2. IEEEXplore
3. ScienceDirect
4. SpringerLink

In the beginning, Microsoft Academic Research Database was also included in the list. The Microsoft Academic Research Database is fairly new database for finding scholarly literature. It did not show promising results when the authors searched for basic keywords of this study, leading to exclude it from final list of selected databases.

2.5.1 Question: Why to target databases and not journals or electronic Sources?

The selected databases cover the most relevant journals, conference and workshop proceedings within software engineering, as confirmed by Dybå et al. [11].

In order to gain a broader perspective, as recommended in Kitchenham guideline [9], 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. [10].

Table 3 shows the databases with respect to the content and type of each database.

Table 3: Database Selection

Database	Type	Content	Selection
ACM	Association / Publisher Database	Journals, Articles, Proceedings	Include
IEEE	Association / Publisher Database	Journal, Articles, Proceedings, Standards	Include
Science Direct	Association / Publisher Database	ScienceDirect database contains more than 10 million journal articles and book chapters.	Include
Springer Link	Association / Publisher Database	Books, Journals, Reference Works, Protocols	Include
Google Scholar	Search Engine	Web pages, Link to articles, Conferences, Journals etc.	Exclude
Libris	Library Catalogue	Books, Journals	Exclude
ebrary	EBook Collection	eBooks	Exclude
Safari	EBook Collection	eBooks	Exclude
Inspec	Expert Database	Journal, Articles, Proceedings	Exclude
ISI	Citation Database	References	Exclude
Summon @ BTH	Meta Search Engine	Everything BTH (Blekinge Tekniska Högskola) Subscribes.	Exclude

2.5.2 Motivation for Databases Selection

For selecting the databases for study, the authors first discussed the possible choices with thesis advisor. The authors also approached couple of researchers in field of Software Engineering e.g. Dr. Wasif and Nauman Ghazi (PhD student), since they have also worked in the similar research area.

Along with this, the authors had a positive discussion with colleagues regarding criteria for the selection of databases for retrieving maximum number of SLRs in Software Engineering. Getting all the information, analyzing the previous systematic reviews and searching on the internet, we were able to finalize our selection.

The searching of the digital libraries for systematic reviews was conducted by two primary reviewers. The digital libraries were randomly allocated to the two reviewers. One reviewer searched studies in IEEE Xplore and ACM while the other reviewer searched in SpringerLink and Science Direct databases.

2.6 Search documentation

The authors have documented all steps of pilot search process to make it systematic and repeatable. The pilot search process is the same that we have used in our full-scale search. The authors have documented the keywords used, the number of

search results, the number of found SLRs and the attributes of search such as year range, publication type, full-text or meta-data search, content types etc. Since all databases have different search attributes to filter the results, a link to Google Drive excel file in Appendix B has been included, which shows the layout used to document search process.

2.7 Search result management

The study required a high collaboration level between authors for managing the search results. This was achieved using Google Drive and DropBox software services. Google Drive was mainly used for sharing and documenting valuable resources for thesis as well as to document the pilot search process. DropBox was used mainly for managing full-scale search results.

For each SLR obtained, after the inclusion and exclusion filters, an excel document was created. The document contained all the primary studies included in that particular SLR. For each primary study, we had columns of "Source" and "Google Scholar Hit/ Miss". The "database" column showed the information source to which the study belonged i.e. IEEE, Springer, Journals or Conferences etc. The Google Scholar column showed whether we were able to find that primary study in Google Scholar. At the end of excel file, a comprehensive summary was given. The summary showed the total number of primary studies belonging to each category IEEE, ACM and Springer etc.

2.8 Inclusion criteria

This review included every article returned by the protocol that met at least one of the following criteria for inclusion (IC) and did not meet any of the criteria for exclusion (EC).

- The article should be a systematic literature review that follows guidelines from Kitchenham.
- The SLR should contain included primary studies list.
- The SLR is peer-reviewed.
- The SLR language should be English.
- The SLR is published between year 2003 and June 2012.

While searching for articles, the authors included both individual systematic literature reviews and combined studies (that have SLR as well as other research methodology) to possible candidates for inclusion and exclusion. Only most recent versions were selected, when the authors were confronted with multi-version SLRs.

2.9 Exclusion criteria

Publications that satisfy at least one of the following exclusion criterions were excluded.

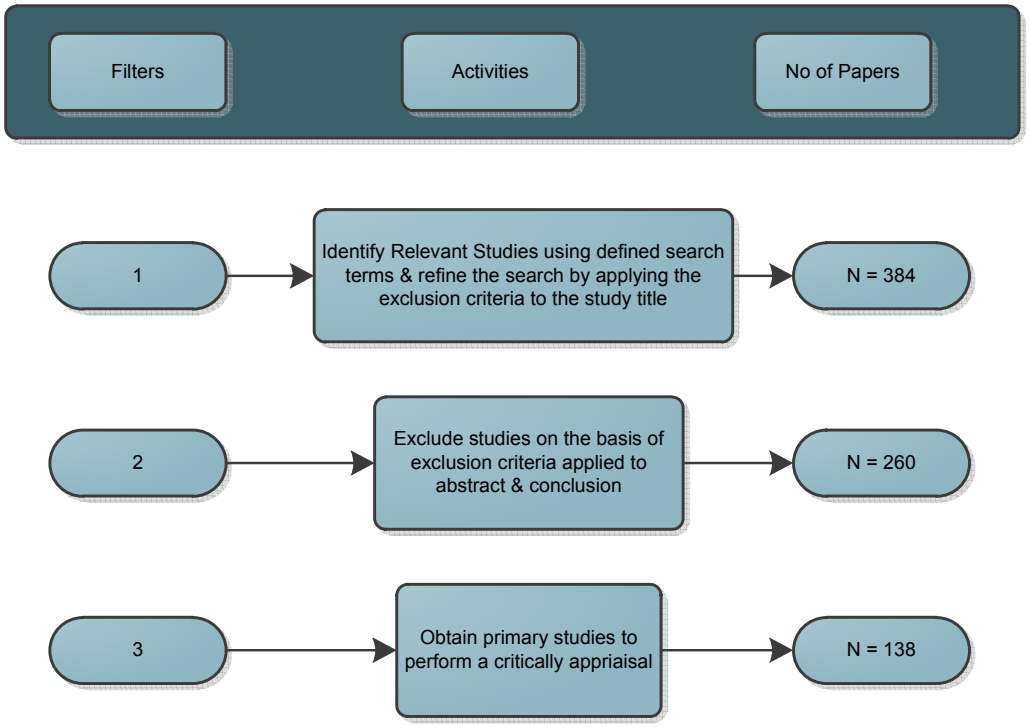
- Articles not available in full-text will be excluded.
- Articles (SLRs) that are not available through our university subscription will be excluded.
- Articles (SLR) that are not specifically from Software Engineering will be excluded.
- If the same study has been published more than once, the most relevant version (i.e. the one explaining the study in greater details) will be used and the others will be excluded.
- If the same study has been published more than once, the most recent version will be included and old version will be excluded.
- Exclude editorials, position papers, keynotes, reviews, tutorial summaries and panel discussions.

- Exclude papers reporting lessons learned, expert judgments or anecdotal reports, and observations.

2.10 Study Selection Procedure

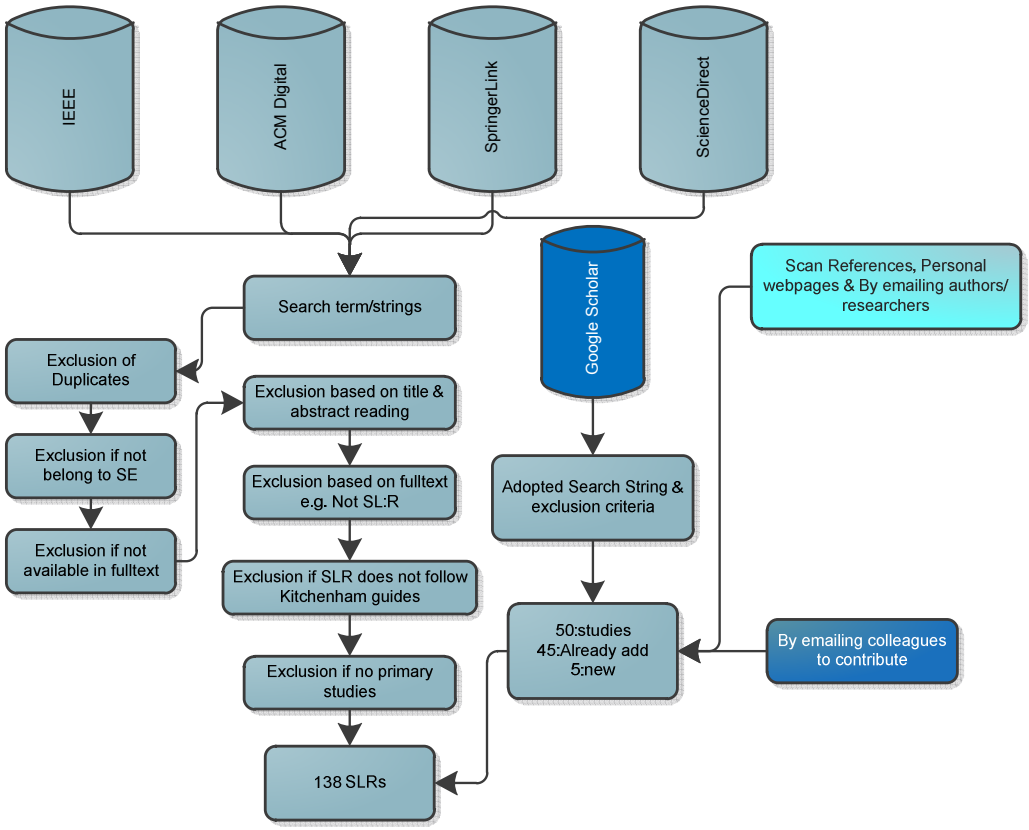
The basic inclusion/ exclusion criterion was to identify only systematic literature reviews and select the most relevant studies. The basic criterion was complemented by detailed inclusion/ exclusion criteria. The detailed inclusion/ exclusion criteria are discussed in above sections. The initial search brought over 6000 articles. The basic inclusion/ exclusion criteria were applied to the search results. The authors read the titles, abstract, and keywords of the articles to filter them out based on the basic criteria i.e. the article is an SLR. Figure 3 illustrates the stages for study selection process.

Figure 3: Stages of Study Selection Process



Both authors applied the inclusion/ exclusion criteria and selected individually. At the end of the selection procedure, the included studies were cross-checked by both authors. Figure 4 illustrates the complete study selection procedure.

Figure 4: Study Selection Procedure



2.11 Quality Assessment Criteria

The authors did not perform study quality assessment as a separate step. For each SLR, inclusion and exclusion criteria worked as quality assessment criteria. Besides that, the authors also tried using the quality assessment criteria of Database of Abstracts of Reviews of Effects (DARE) [29]. The authors excluded it later because during application of the DARE criteria on each SLR, it was time consuming and the results did not seem to add any value to main study objective.

2.12 Data Extraction Strategy

The authors extracted the data that was most relevant for study focus. The data extracted from each study were:

1. The source (journal or conference) and full reference.
2. The author(s),the institution and the country where the institution is situated.
3. How many primary studies used in the SLR
4. List of all the primary studies used in the SLR.

The data extraction was followed by a more detailed data extraction of primary studies used in the SLRs. The extraction process from primary studies is detailed in following section.

2.12.1 Data Extraction from Primary Studies

Similar to pilot search, the authors performed pilot data extraction from each SLR. The primary studies were extracted by reading the study selection section of SLRs. Most of the SLRs have mentioned the included primary studies explicitly. In some

cases, the primary studies were not mentioned in the study selection section thus the authors extracted the primary studies by going through the full-text in those cases.

The authors created an excel file for each SLR, containing primary studies included in the SLR. For each primary study, the authors searched for the information source (i.e. IEEE, ACM, Journals, or Grey Literature) of that primary study. This was important to classify the primary study as grey literature or white literature (RQ1). Some articles had already categorized their included primary studies by information source. In such article cases, the authors used the list provided by the article . According to the definition of grey literature, conference proceedings, technical reports or lecture notes are considered grey. The grey literature classification was done based on information source for each primary study.Each included primary study was searched in Google Scholar to find a hit or miss. A Google Scholar hit means that we are able to find the study in Google Scholar and Google Scholar miss represents the opposite i.e. study is not found in Google Scholar. There were three columns in each excel file named as Study Name, Article Source and Google Hit/ Miss respectively. At the end of each excel file, a summary was written about the grey literature and Google Scholar finding. The purpose of writing summary was to help in aggregating results later.

The databases were randomly allocated to the authors with one researcher focusing on IEEE and ACM while the other focused on SpringerLink and Science Direct. Each author extracted and verified data extraction of his each allocated database. After one author finished complete process of a database, the results were given to the other author to cross-check. This cross-checking was done on partial level because of time constraint. This crosscheck approach was used because it allowed to crosscheck findings more thoroughly. When there was a disagreement, we the issue was discussed with the advisor until there was an agreement.

2.13 Evaluating Review Protocol

The review protocol is vital for systematic literature reviews. The evaluation of protocol should be done by independent experts. The authors reviewed and evaluated the protocol through discussions with supervisor. This helped to properly review and formulate research questions and other elements of protocol including data extraction strategies and inclusion/ exclusion criteria for primary studies. The keywords and search strings were optimized with the help of librarians to make sure that search strings are appropriately derived from the research questions.

The authors piloted research protocol on a limited scale. Piloting the research protocol helps in identifying potential mistakes in data collection procedures [12]. The pilot review protocol process greatly helped to improving data collection procedure by identifying more relevant attributes to extract.

2.14 Data Synthesis

The data synthesis involves summarizing the results of included primary studies. The collected data can be synthesized for valuable information. Data synthesis can be descriptive (narrative) and quantitative [9]. Often, the descriptive synthesis requires a complementary quantitative synthesis. The quantitative synthesis uses statistical techniques (mean, median, odds ratios, relative risks, etc.). Quantitative synthesis is generally referred to as meta-analysis.

The first stage of synthesis was to identify the extent to which the grey literature has been used in systematic literature reviews. For achieving this first stage, the extracted data was analyzed by both authors. Each of the included primary study in

SLR was reviewed thoroughly by author assigned to that particular database and was marked as grey or white. Similar steps were taken for the Google Scholar findings.

For each selected database, the authors created a excel file that contained all the aggregated data of individual SLRs excel files. . This aggregation was necessary to investigating grey literature evidence of SLRs published in each database. The columns were named as [SLR Name], [Total Primary Studies], [IEEE Count], [ACM Count], [Springer Count], [ScienceDirect Count] and [Grey Literature]. Corresponding to each database, there were four main files, containing the aggregation of grey literature evidence and Google Scholar findings.

Chapter 3 enlightens the readers to the results of data synthesis.

3 RESULTS AND ANALYSIS OF SYSTEMATIC REVIEW

In this chapter, we present results of systematic literature review. We have used descriptive analysis and complemented it with a meta-analysis.

3.1 Classification of Studies

A total of 138 SLRs were selected for the review. These SLRs covered four major databases as well as data suggested by the different authors. We present our results separately for each database and later, in the end, we will draw the overall picture of grey evidence and Google Scholar findings.

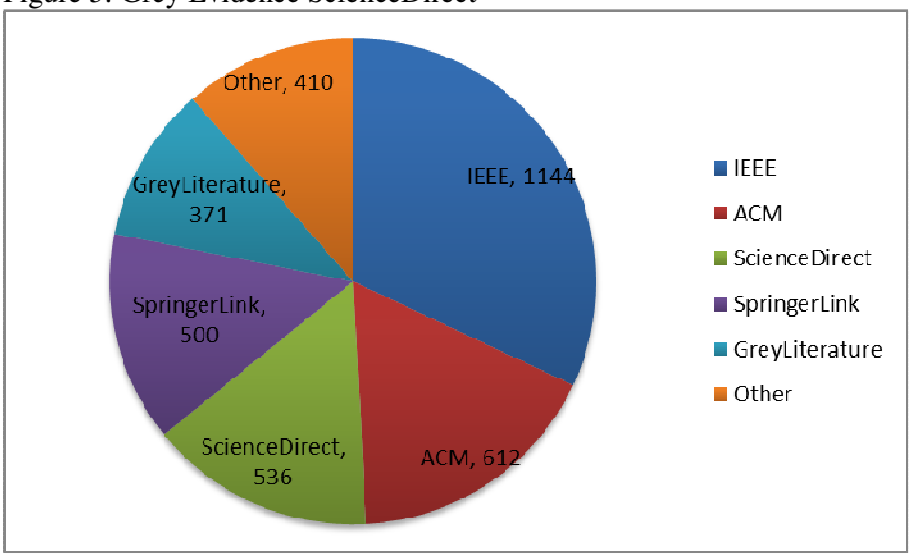
3.2 Science Direct SLRs

A total of 67 SLRs were selected from ScienceDirect database. There are a total of 3573 primary studies in all the SLRs.

3.2.1 Grey Literature Extent

From 3573 primary studies, 1144 belonged to IEEE, 536 to ScienceDirect, 612 to ACM, 500 to SpringerLink while 371 were classified as grey literature. A total of 2868 primary studies cover four databases i.e. IEEE, ACM, ScienceDirect, SpringerLink and grey literature sources. The remaining 410 primary studies cover other journals and books. In this study, we present results of the four selected databases i.e. IEEE Xplore, ACM, ScienceDirect and SpringerLink. Figure 6 illustrates the evidence of grey literature in SLRs, which are selected from ScienceDirect database. We observe that 371 primary studies are from grey sources.

Figure 5: Grey Evidence ScienceDirect



The percentage of primary studies from ScienceDirect database is tabulated below. The numbers are rounded to nearest 10. From the table, we observe that most of the primary studies in our selected SLRs, nearly 32%, belong to IEEE database. Similarly ACM, Science Direct and SpringerLink articles cover 17%, 15% and 15% of total primary studies respectively. The grey evidence is 10% of the total primary studies.

Table 4: Grey Evidence in Science Direct

Information Source	Total Primary Studies	Percentage (%)
IEEE	1144	32
ACM	612	17
Science Direct	536	15
Springer Link	500	15
Other Journals/Books	410	12
Grey literature	**371**	**10**
Total	**3069**	

3.2.2 Google Scholar Findings for ScienceDirect

We searched **3573** primary studies in Google Scholar database and found **3383** studies as hit and **190** studies as miss. A total of **94.6** percent of primary studies were found in the Google Scholar. The more granular breakdown of each information source primary studies is tabulated below in Table 5.

Table 5: Google Scholar Science Direct

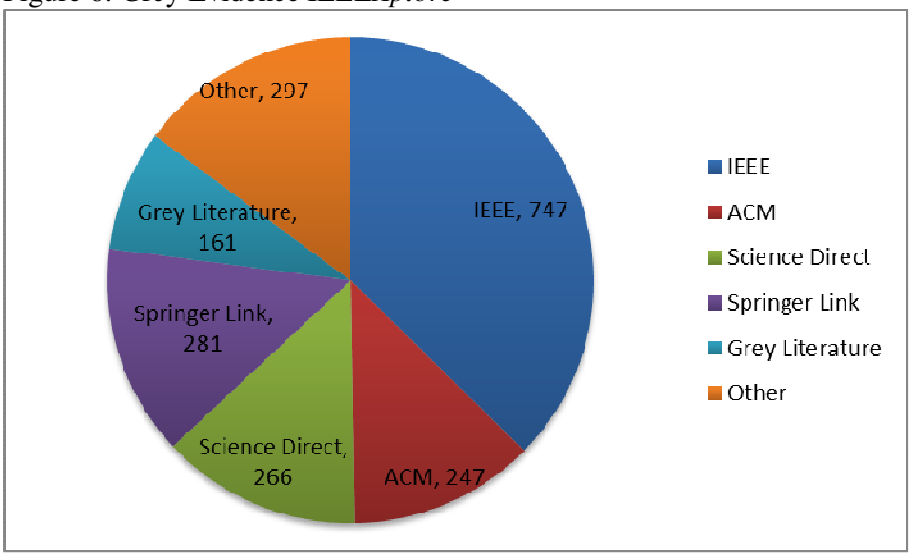
Information Source	Found in GS	Not Found in GS
IEEE	1129	15
ACM	602	10
ScienceDirect	530	6
SpringerLink	491	9
Grey Literature	302	69
Total	**3383**	**190**

Table 5 shows that out of 1144 primary studies from IEEE, we were able to find 1129 primary studies in Google Scholar while 15 primary studies were not found in Google Scholar. Similarly, out of 612 primary studies of ACM, we found 602 primary studies in Google Scholar while 10 primary studies were not found using Google Scholar. Overall, the Google Scholar hit ratio for SLRs published in ScienceDirect is around **94.6 percent** (3383 / (3383+190)). Please note that, while searching for primary studies, we were able to find some studies on Google.com website that were not found in Google Scholar database.

3.3 IEEE SLRs

There were 48 SLRs retrieved from IEEE, which consisted of 1999 primary studies. Figure 7 shows the evidence of grey literature in SLRs selected from IEEE *Xplore* database.

Figure 6: Grey Evidence IEEE*Xplore*



From 1999 primary studies, 747 were from IEEE *Xplore*, 243 from ACM, 266 from ScienceDirect, 281 from SpringerLink and 161 primary studies were categorized as grey literature. The grey literature evidence in IEEE database is found to be as **8% (161 primary studies)**. Around 297 primary studies were taken from other information sources like journals or books. The full coverage of primary studies with respect to the databases is tabulated below in Table 6.

Table 6: Grey Evidence in IEEE

Information Source	Total Primary Studies	Percentage %
IEEE	747	38
ACM	247	12
Science Direct	266	13
Springer Link	281	14
Other Journals/Books	297	15
Grey literature	**161**	**8**
Total	**1999**	

3.3.1 Google Scholar Findings for IEEE

We searched 1999 primary studies in Google Scholar (GS). A total of 1928 primary studies were found using GS while 71 primary studies were not found. Overall **96 percent** of primary studies are found using Google Scholar. The results of Google Scholar findings are tabulated below in table 7.

Table 7: Google Scholar IEEE

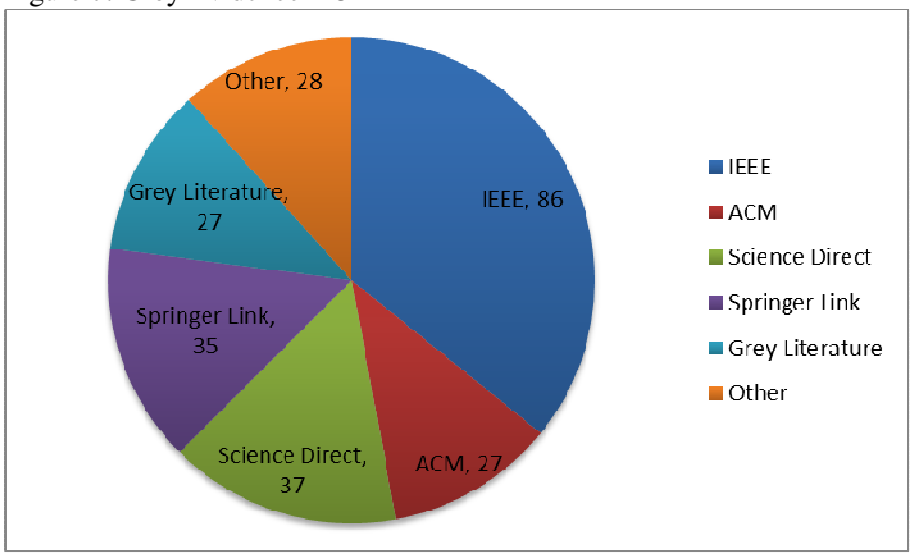
Information Source	Found in GS	Not Found in GS
IEEE	745	2
ACM	247	0
ScienceDirect	266	0

SpringerLink	275	6
Grey Literature	133	28
Total	**1928**	**71**

3.4 ACM SLRs

We retrieved 9 SLRs consisting of total 240 primary studies. There were 27 grey sources used as primary studies in SLRs selected from ACM database.

Figure 7: Grey Evidence ACM



Out of 240 primary studies, 86 were from IEEE, 27 from ACM, 37 from Science Direct, 35 from SpringerLink and 27 (11%) primary studies were classified as grey literature. There were 28 primary studies that are taken from other information sources such as Journals or Books. The primary studies coverage with respect to databases is tabulated below in table 8.

Table 8: Grey Evidence in ACM

Information Source	Total Primary Studies	Percentage %
IEEE	86	36
ACM	27	11
Science Direct	37	15
Springer Link	35	15
Other Journals/Books	28	12
Grey literature	**27**	**11**
Total	**240**	

3.4.1 Google Scholar Findings for ACM

We searched 240 primary studies on Google Scholar. Out of these 240 primary studies, we were able to find 229 primary studies using Google Scholar. So, overall we were able to find about **95 percent** of total primary studies of **ACM** SLRs using Google Scholar. The results of Google Scholar finding are tabulated below in table 9.

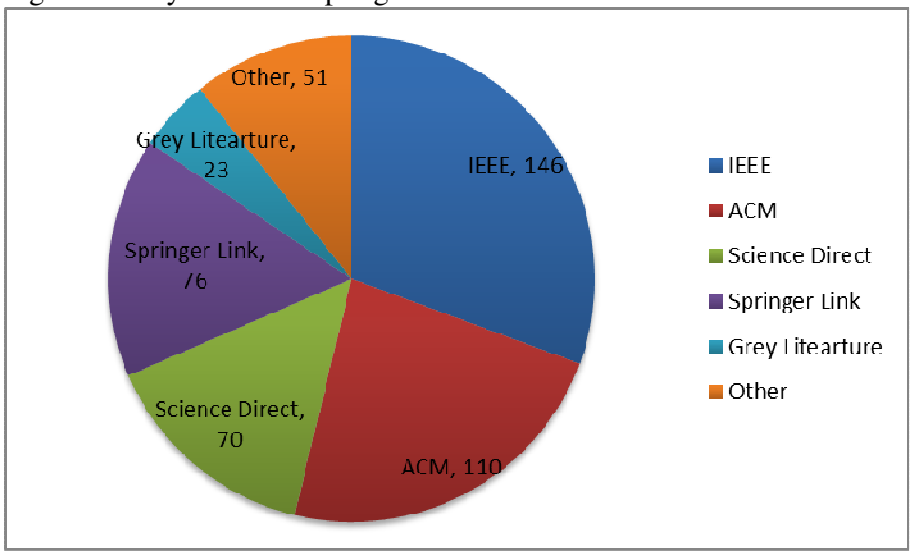
Table 9: Google Scholar ACM

Information Source	Found in GS	Not Found in GS
IEEE	85	1
ACM	27	0
ScienceDirect	37	0
SpringerLink	38	0
Grey Literature	20	7
Total	**229**	**11**

3.5 SpringerLink SLRs

There are 476 primary studies extracted from 14 SLRs of Springer Link database. Around 23 primary studies were found to be from grey sources.

Figure 8: Grey Evidence SpringerLink



The grey literature evidence in SpringerLink SLRs is found to be as 5% (23 primary studies). The primary studies coverage with respect to database is tabulated below in table 10.

Table 10: Grey Evidence in SpringerLink

Information Source	Total Primary Studies	Percentage %
IEEE	146	30
ACM	110	23
Science Direct	70	15
Springer Link	76	16
Other Journals/Books	51	11
Grey literature	**23**	**5**
Total	**476**	

3.5.1 Google Scholar Findings in SpringerLink

We searched 476 primary studies in Google Scholar for SpringerLink SLRs. We found 468 primary studies while 8 studies were found using Google Scholar. Overall 98 percent of the total 476 primary studies are found using Google Scholar.

Table 11: Google Scholar SpringerLink

Information Source	Found in GS	Not Found in GS
IEEE	146	0
ACM	110	0
ScienceDirect	70	0
SpringerLink	75	1
Grey Literature	19	4
Total	**468**	**8**

3.6 Data Synthesis

In this section, we have presented data synthesis of results of all four databases as detailed in the above sections. We collected total of 138 SLRs for our study. There were 6307 primary studies extracted from the SLRs. The total SLRs and the primary studies are given below for each database in table 12.

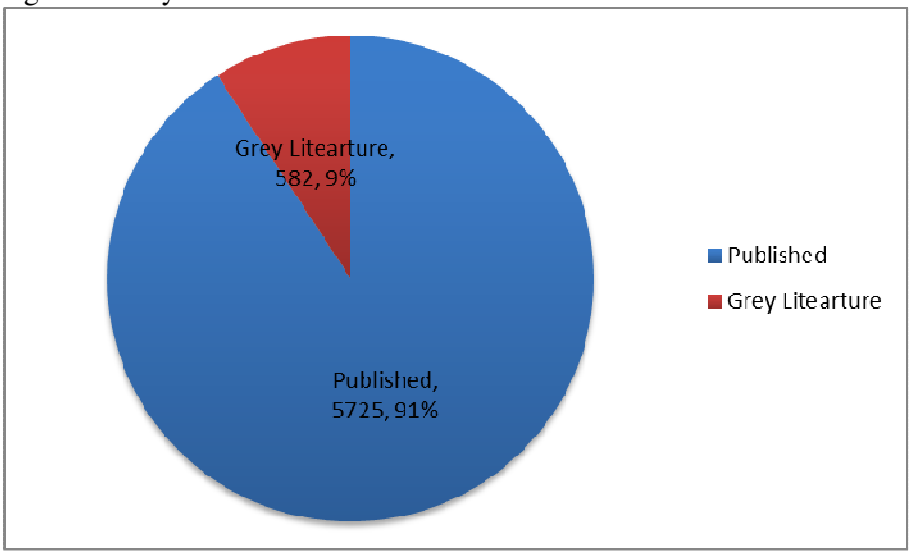
Table 12: Total SLRs/Primary Studies

Database	Total SLRs	Primary Studies
IEEE	48	2018
ScienceDirect	67	3573
ACM	9	240
SpringerLink	14	476
Total	**138**	**6307**

3.6.1 Total Grey Evidence

A total of 6307 primary studies included in 138 SLRs are investigated. We have found out that 582 primary studies are from grey sources. The percentage of grey evidence is around 9.22 percent in the selected 138 SLRs of Software Engineering. Figure 10 shows the extent to which grey literature has been used in SLRs in Software Engineering (SE).

Figure 9: Grey Evidence in SE SLRs



The figure above shows the overall evidence of grey literature usage in systematic literature reviews in Software Engineering. We observe that published and peer reviewed primary studies cover a major part of 91% while the grey evidence only covers a minor portion of 9%. Also, in published literature, most of the primary

studies (more than 50%) belong to major information sources like IEEE *Xplore* and ACM. We have noticed that most of the grey literature that has been included as primary studies in SLRs are conference proceedings and technical reports.

3.6.2 Grey Literature Usage in Information Synthesis

After finding the grey evidence in primary studies of 138 SLRs, we proceeded to find out whether the included grey evidence was used in synthesis section of SLR. The motivation is to find the of extent grey evidence to strengthen or reject arguments in synthesis section.

For this analysis, we selected 582 already found grey references and checked them in their corresponding SLRs synthesis sections. We scanned the synthesis section of each SLR and found that **93.2 percent** of grey evidence was actually used in the synthesis section. The table 13 shows the grey literature usage in each database.

Table 13: Extent of Grey Evidence in Information Synthesis

Database	SLRs	Grey Literature Links	Found in Synthesis
ScienceDirect	67	371	340
IEEE	48	161	153
ACM	9	27	27
SpringerLink	14	23	23
Total	138	582	543
Percentage			**93.2 %**

The table above shows the results of finding grey evidence in synthesis section. From the table, we can see that out of total 582 grey references from 138 SLRs, we have found 543 grey references in the synthesis section.

3.6.3 Google Scholar Synthesis

We searched for 6307 primary studies in Google Scholar and found 6027 primary studies as hit. Only 280 primary studies were not found using Google scholar. The GS hit percentage is around 96 percent.

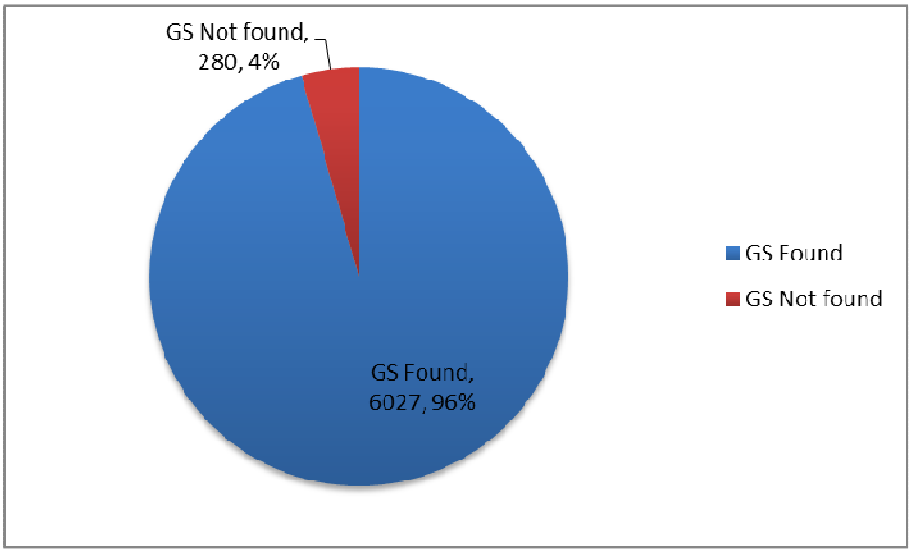


Figure 10: Google Scholar Results for Primary Studies

Looking into more details, we have noticed that out of 6027 primary studies found in Google Scholar, a large portion belonged to major databases i.e. IEEE has 25% while ACM has 30% .During the analysis of 280 primary studies, which are not found

in Google Scholar, we have noticed that most of the primary studies are grey sources. Around 40% of the primary studies that are not found in Google Scholar, are grey literature. We believe that this is because of the fact that grey literature is volatile in nature. This can also be the fact that sometimes the grey literature is not published in electronic formats or is not published over the web at all.

3.6.4 Definitions and Indicators

The following indicators are calculated based on our gathered data. The definition of each indicator is given below.

- ***Frequency of GL use***;
  The proportion of SLRs with grey primary studies, out of all the SLRs examined.

- ***Frequency of GL citing;***
  The proportion of grey primary studies out of all the primary studies examined.

- ***Intensity of GL uses***;
  The frequency of grey primary studies divided by the frequency of GL use. This is indicator of the average frequency of GL primary studies per SLR with GL.

3.6.4.1 Frequency of SLRs with Grey Primary Studies

This indicator shows the total number of SLRs which have grey primary studies. We have seen from the table 14 that 76% (105 SLRs) of total SLRs (138 SLRs) have used grey literature as source for their primary studies. The breakdown of each database SLRs is given in below table.

Table 14: Frequency of Grey Primary Studies

Database	Number of SLRs	Number of SLRs with grey primary studies	Frequency of GL use (%)
ScienceDirect	67	55	82.1
IEEEXplore	48	36	75.0
ACM Digital	9	6	66.67
SpringerLink	14	8	57.14
Total	138	105	76.7

3.6.4.2 Frequency of Grey Primary Studies

This indicator shows the total number of grey primary studies percentage for each database. We have seen from table 15 that total 582 primary studies are identified as grey literature out of 6307 primary studies. This indicator is the total grey literature evidence in SLRs of software engineering. From the table below, we see that total evidence of grey primary studies in SLRs is 9.22 percent.

Table 15: Total Primary Studies, Total Grey Primary Studies

Database	Total Primary Studies	Grey Primary Studies	Frequency of Grey Primary Studies (%)
ScienceDirect	3573	371	10.38
IEEEXplore	1999	161	8.05
ACM Digital	240	27	11.25
SpringerLink	476	23	4.8
Total	6307	582	9.22

3.6.4.3 Intensity of GL use

The intensity of GL use is the average number of grey primary studies in SLRs. It is calculated as below;

Total Grey Primary Studies / Total SLRs with Grey primary studies

We observe that the intensity of grey literature usage in 105 SLRs is 5.54. Table 16 shows the intensity of GL usage indicator for each selected database.

Table 16: Average of grey primary studies per SLR

Database	Frequency of GL		Intensity of GL use
	Usage (%)	Grey Primary Studies (%)	Total Grey Studies / Total SLRs with Grey studies
ScienceDirect	82.1	10.38	6.75
IEEEXplore	75.0	8.05	4.47
ACM Digital	66.67	11.25	4.5
SpringerLink	57.14	4.8	2.88
Total	76.7	10.83	5.54

3.6.5 Characteristics of GL Documents

While analyzing the characteristics of grey primary studies, we have noticed that some of the grey primary studies lack proper bibliographical controls required for analysis. Some SLRs have missing published dates. Although we have not encountered any typical problems of grey literature like typographical errors, omissions and inconsistencies, we have been unable to find information about some grey primary studies. The Grey literature analysis requires not only the name(s) of the individual author(s), but also the corporate author or organization responsible for the publication.

3.6.5.1 Forms of GL cited

The distribution analysis of grey literature primary studies is shown in table 17. The analysis is performed on the document type of grey primary studies cited. We have observed that conference papers and proceedings are the most cited (43%) document type in SLRs followed by technical reports (25.2%) and theses/ dissertation (12.4%) as grey literature. The technical reports category includes several subtypes of reports like research reports, internal reports and review reports.

Table 17: Ranking of type of Grey Primary Studies Documents

Type of Grey Literature	Number of Primary Studies	%
Conference Papers/Proceedings	252	43.3
Technical Reports	147	25.2
Thesis/Dissertation	71	12.4
Workshop/Seminars	44	7.62
Guidelines/Lecture Notes	36	6.08
Public Reports	19	3.3
Preprints	12	2.1

We have observed that conference proceedings papers and technical reports cover major proportion (68%) of the grey primary studies. The conference papers and technical reports are designed for external diffusion and usually the scientific community is officially informed about these reports. However, dissertations are not intended for external diffusion and most of the times remain unnoticed by scientific communities.

3.6.5.2 Origin of documents

Table 18 shows the split of grey primary studies by the type of producers. We observe that universities and international organizations are the biggest producers of grey literature documents, covering around 60% of total grey primary studies. We have noticed that the grey studies produced by international organizations and research labs like Carnegie Mellon Software Engineering Institute (SEI), IBM Research Center and Simula Research Lab contained well-formed bibliographical details. All of the publications that have been issued by international organizations and research centers are highly accessible and have decent bibliographical controls.

Table 18: Grey Studies split in terms of individual producer

GL Primary Study Origin	Number of Primary Studies	%
Universities	223	38.39
International Organizations	120	20.53
Research Institutes/Labs	114	19.64
Scientific Societies	57	9.82
Others	47	8.03
Govt. Organizations	21	3.57

These organizations produce state of the art technical reports and articles about software engineering.

3.6.5.3 Date of Publication

We found only 12 grey primary studies (2%) that did not have date of publication. The split of primary studies with year of publication is given is table 19. We have found out that majority of grey primary studies included in the SLRs are concentrated in the recent past. Almost 48% (280) of included grey primary studies have been published during the last 5 years.

Table 19: Dated Grey Primary Studies

Year	Number of Grey Primary Studies	%
1970-1990	15	2.56
1991	5	0.85
1992	10	1.70
1993	5	0.86
1994	10	1.70
1995	11	1.81
1996	5	0.85
1997	6	0.91
1998	34	5.98
1999	20	3.41
2000	15	2.56
2001	24	4.27
2002	34	5.98

2003	53	9.40
2004	43	7.69
2005	86	15.38
2006	29	5.12
2007	58	10.58
2008	48	8.54
2009	29	5.12
2010	20	3.41
2011	10	1.70

3.7 Validity Threats Related to SLR

Kitchenham guidelines suggest to study the validity threats at the end of SLR [12]. We have observed the following threats related to our study.

3.7.1 Conclusion Validity

Conclusion validity is related to the ability to draw correct conclusion from the research results. It is the reliability factor of study results.

Generally, the conclusion of systematic literature review greatly depends upon the correct selection of primary studies publications and data extraction procedures. During this research process, we remained more concerned with the latter because it involved the manipulation of data. We collected large amount of data from each SLR and this data involved arithmetic computations for analysis. Our study is a tertiary study and primary studies comprise only of SLRs. We tried to collect as many SLRs as possible. This has naturally reduced our threat of finding incorrect studies since any SLR of Software Engineering is a correct study.

For data extraction, we used Google Drive and MS Excel for managing data and collaboration. The data extraction process was trivial but time consuming. We have extracted Title, Year of Publication, Authors and most importantly, the primary studies from each SLR. We have noticed that most SLRs follow the Kitchenham guidelines and mentioned the primary studies clearly. We came across some SLRs in which the primary studies were not clear and hard to extract. In such cases, we both authors individually analysed the article for possible primary studies and reported to each other. We found no disagreement among ourselves in these cases.

3.7.2 Construct Validity

Construct validity threat deals with the generalization of the research results. Sometimes, it is possible that the result of research may be generalized to extent that is not intended. We have minimized this threat by clearly mentioning the aims and objective of this study in the review protocol. The aims and objectives help to define the scope and boundary of the study. Also, our topic of research i.e. Grey Literature is already very limited in scope of information science. This has helped in minimizing the threat of misinterpretations of results by the other authors.

3.7.3 Internal Validity

3.7.3.1 Publication Bias

Qualitative and observational study results are often vitiated by systematic error. The systematic error is related to publication biasness. Publication bias is the tendency that the studies with positive results are more widely disseminated than the studies with negative results [16]. For minimizing this threat, we conducted our review based on a pre-defined review protocol. The review protocol defines in advance, research questions, inclusion/exclusion criteria for selected studies and methods to be used.

This helps to minimize the internal bias during the systemic literature review and ensures internal validity.

3.7.4 External Validity

Threats to the identification of primary studies

To avoid the external validity threat of missing a primary study i.e. SLR, we contacted different research authors, our research colleagues for any contribution regarding our research topic. Most of the authors replied with their recent SLRs publications. We collected around 25 SLRs with the help of this process and included them in our study.

4 STRATEGIES FOR GREY LITERATURE

The use of scientific publications that have not been formally published, for example in journals, is widespread in some areas. The internet has made it possible to spread these unpublished scientific publications to a much broader audience [21]. Open Access (OA) based data sources are of interest as an alternative to the traditional citation databases, mainly because of the coverage. OA based data sources are not limited to journal articles as the traditional citation databases, meaning that the working papers and monographs can be better analyzed by a data source based on the OA resources. However, coverage of OA is not easy to determine [22]. With the advent of the Internet and electronic publishing, new models of research communication has emerged that simultaneously complement and challenge established systems. The term 'Open Access'(OA) generally means to access however; downloading and reading the material is free for the entire population of Internet users [27].

4.1 Data Never Sleeps: The Need to Categorize Grey Literature

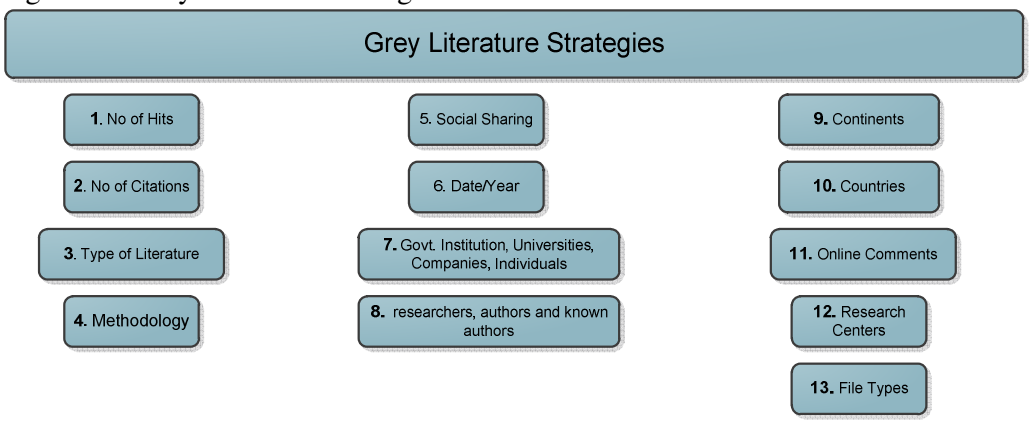
Marketers, business men, scientists and engineers share vast amount of data every day on the internet. Although it is hard to derive a meaningful conclusion from this vast amount of data, the internet is loaded with research, blogs, photos and information data every day. Recently, the two internet research organizations such as WOMMA and Domo (http://www.domo.com/) have analyzed exactly how much data we consume every day every minute. The results show that huge amount of data is produced over internet every day. According to the research, more than 680000 Facebook users share content every minute, 48 hours of video shared by You Tube users every 60 seconds and 47000 Apps downloads request recieved by Apple every minute.

Similarly, while looking at the academic and scientific communities, we observe that students, researchers and scholars producing personal opinions, reports and articles over internet. Now a question arises that how much more data do we need? Or are we utilizing this huge amount of data properly? "Just because the information is before us, does not mean it is significant. Eliminating useless data and determining relevant data will allow us to make the meaningful conclusions needed to get ahead. Only then, we will be able to get some rest while the mountains of data continue to pile in" [24]. Therefore in order to properly and systematically utilize this massive amount of information in research, we need to suggest a categorization strategy for the unpublished literature. In the following section, we have presented strategies to categorize grey literature based on various attributes. These strategies are the result of our experience and knowledge gained about grey literature while investigating the grey evidence in SLRs.

4.2 Strategies for Grey Literature

In our opinion, the solution to filter available grey literature using various categorization techniques can be adopted not only by researchers of software engineering but also by researchers from all other fields such as Health, Engineering and Social Sciences etc. Figure 11 illustrates the strategies that we have concluded.

Figure 11: Grey Literature Strategies



4.2.1 No. of hits

4.2.1.1 Explanation

If we get the number of hits on a particular article, then we can get an idea that this article is accessed by "x" number of people, giving a good idea if this article is worth to read and include in research.

4.2.1.2 Strengths

- The number of hits of a webpage is a good measure to know about the webpage popularity. By using this strategy and defining a popularity scale, we can categorize grey literature in popular/ unpopular category e.g. 200 hits to be popular.

4.2.1.3 Weaknesses

- The process of finding the total number of hits is not easy and requires technical knowledge and familiarity with online web based services.
- Although number of hits is a good measure of webpage popularity, it does not account for the quality of content available on the webpage. An article which has higher number of hits could be because of the negative or controversial data available in the article.
- Some good articles with low number of hits may be missed in this strategy.

4.2.2 Number of Citations

4.2.2.1 Explanation

The number of citations shows the number of times any specific article is quoted or cited in other articles. We can get the citations from online services like Google Scholar or Microsoft Academic Research.

4.2.2.2 Strengths

- We can get an insight on the quality of article if it has been citied or quoted by large number of authors.
- We can categorize grey literature in two main categories as high cited literature and low cited literature. The high and low citation of literature may be defined according to the study perspective.
- As compared to the number of hits, the number of citations is more related to scholarly and academic research articles.

4.2.2.3 Weaknesses

- The previous studies have shown a correlation between the free online availability or Open Access status of articles and citation counts [21]. (Presumably due to higher visibility and/ or access, OA data has higher citation [21] [22]). That means OA articles have usually high number of citations as compared to other articles available in other databases.
- This strategy does not contain information about the quality of literature, as it was the case with first strategy i.e. number of hits.

4.2.3 Type of Literature (Annual reports, Theses/dissertations, Conference abstracts/proceedings)

4.2.3.1 Explanation

In this strategy, we categorize grey literature according to the different types of scientific literature. Some of the widely used types of grey literature found over internet are tabulated below.

Table 20: Types of Grey Literature

1. Annual reports	2. Theses/ dissertations
3. Conference abstracts/proceedings	4. Workshops
5. Working papers	6. Evaluation reports
7. Facts sheets	8. Study or research reports
9. Scientific and technical reports	10. Government documents
11. Patent documents	12. Newsletters
13. Standards and specifications	14. Codes of practice
15. Article Pre-prints	16. Web Objects
17. Legal Tests	18. Lecture Notes
19. Press Releases	20. Leaflets
21. Statistical Surveys	22. Brochures Booklets

By arranging the literature according to the types as mentioned in the table above, we can categorize grey literature.

4.2.3.2 Strengths

- Using this strategy, we can filter data by type and can use that particular type of grey literature in our research. For instance, a researcher may find it convenient to use 'Conferences and Proceedings' data and thus will apply Conference Proceedings grey literature category in research.

4.2.3.3 Weaknesses

- It is difficult to rank grey literature with respect to quality within particular type. We are unable to identify the quality of literature in a particular type.

4.2.4 Methodology/ Source data (e.g. survey, experiment, sase-study, etc.) used in the document.

4.2.4.1 Explanation

In this strategy, we categorize the grey literature with respect to the methodology used in that literature e.g. survey, experiment, case studies, systematic literature reviews etc.

4.2.4.2 Strengths

- Researchers can target a particular methodology in grey literature and utilize it as a useful resource.

4.2.4.3 Weaknesses

- The strategy cannot take into account all the grey literature and it is not guaranteed that the authors will follow any guidelines. For example, In Systematic Literature Review, it is appreciated to use Kitchenham guidelines and now-a-days researchers follow these guidelines while performing SLR. However, in grey literature, authors may or may not use any guidelines. After reviewing the material, it is up to the author to add or reject the grey data.
- The researchers from underdeveloped countries do not have easy access to commercially published materials. They rely on grey literature as a major source of knowledge. However, they have to investigate grey data thoroughly before adding it into their studies.

4.2.5 Social Sharing

4.2.5.1 Explanation

According to this category, researchers help others in getting relevant content by adopting social media sharing systems. We have today Google +1, Facebook share and Twitter functionality [23] [24]. We can always Google +1, share article on Facebook and tweet it on Twitter. The count of shares from colleagues can be helpful in accessing content over the internet.

4.2.5.2 Strengths

Using this strategy, we can get the referred content related to our problem topic in a less amount of time . For instance, in a research centre 5 to10 members work on some topic (e.g. Search based testing) and they all decide to use Google 1+ technique to filter out the grey material. This will be helpful for all the team members to get content referred by other fellow members and check how many people from the team referred particular literature.

4.2.5.3 Weaknesses

One of the natural human instincts is to resist change even if it is designed to be beneficial.

The way any change is introduced has its own power to motivate or demotivate, and can often be the key to success or failure [25]. The bigger challenge will be how to motivate researchers to adopt this method. The effectiveness of this technique increases as the maximum number of researchers help each other on the internet.

4.2.6 Date/Year (New or Old)

4.2.6.1 Explanation

In this strategy, we categorize the grey literature with respect to published year.

4.2.6.2 Strengths

This strategy helps in filtering out the most recent published content that addresses the latest research problems. The latest published content can be useful in identifying research gaps.

4.2.6.3 Weaknesses

It will be difficult to implement this strategy since it will lead to deal with a massive amount of data.

4.2.7 Govt. Institution, University, Company or Individual

4.2.7.1 Explanation

In this strategy, we categorize the grey literature with respect to source of grey literature. The grey literature published by government Institution, university, company or individual author etc.

4.2.7.2 Strengths

The literature produced by these institutions is well researched and processed. The type of grey literature normally results from R&D (Research and Development) activities and projects within different institutions.

4.2.7.3 Weaknesses

- As researchers/ authors of literature belong to the institution e.g. government institution, company, university etc. there will be chances of biasness in the reports.
- The literature produced on Govt. level or individual companies may not be understandable to everyone.

4.2.8 Famous researchers or the familiar authors in any specific field.

4.2.8.1 Explanation

In this strategy, we categorize the grey literature based on the famous authors. For instance, in Systematic Reviews "Barbara Kitchenham" is one of the famous researchers, so any data related to that author(s) may be taken into account.

4.2.8.2 Strengths

- By using this categorization, there are high chances to get the quality work as the famous authors from top of their fields are selected .

4.2.8.3 Weaknesses

- One may not get a lot of grey literature on a specific topic as filtering is done by authors/ researchers criterion.
- By selecting particular authors/ researchers for one's research, it may raise the question of biasness.

4.2.9 Continents (Asia, Africa, Europe, America, Australia etc.)

4.2.9.1 Explanation

In this strategy, we categorize the grey literature with respect to different regions i.e. continents of the world.

4.2.9.2 Strengths

- This strategy is helpful in filtering material into two categories i.e. material from one trusted continent and material from rest of the world. We generally know that European, American and Australian continents are good in research, so may categorize and take the research from that particular continent.

4.2.9.3 Weaknesses

- This categorization technique does not guarantee to filter the quality material.

4.2.10 Countries (America, Sweden, Norway, UK etc.)

4.2.10.1 Explanation

In this type of categorization, we categorize the grey literature based on grey literature producing countries.

4.2.10.2 Strengths

The research and scholarly articles produced in developed countries like Sweden, Germany, England etc. are generally of high quality as compared to the developing countries. Thus, it is possible to focus grey literature from one certain famous country.

4.2.10.3 Weaknesses

This strategy may limit the scope of the grey literature because a certain country may be producing more literature in a specific field than the targeted research field.

4.2.11 Online Comments (on forums, blogs, Discussion boards)

4.2.11.1 Explanation

In this strategy, we categorize the grey literature published on the internet i.e. blogs, discussion boards and forums by the number of comments posted on the certain blog post.

4.2.11.2 Strength

- A highly commented blog post is likely to have most of its flaws/ obvious issues identified.
- The quality of the literature will be increased within a less amount of time.

4.2.11.3 Weaknesses

The spam comments in blogs may increase the total number of comments without adding any value to the published literature .

4.2.12 Research Centres

4.2.12.1 Explanation

There are 62 research groups working only in Norway in the field of Information and Technology. Similarly, different research centres produce grey literature in different forms such as white papers, working papers and technical reports. In this strategy, we categorize grey literature based on the research centres.

4.2.12.2 Strengths

The quality of grey literature produced in research centres is eminent because this literature is produced by scholars working specifically in the field of research.

4.2.12.3 Weaknesses

Although most of the research centres produce literature in English language, there are chances that the language of grey literature may be other than English e.g. Spanish, German etc. This will reduce the usefulness of grey literature for researchers from other countries.

4.2.13 File Types (pdf, ppt, doc)

4.2.13.1 Explanation

In this strategy, we categorize the grey literature, which is available over the internet, with respect to the file formats such as pdf, doc and ppt.

4.2.13.2 Strength

Researchers can target the particular presentation type in grey literature and can use that in strengthening or rejecting arguments.

4.2.13.3 Weaknesses

Some file types depend on the operating systems (OS). In this case, third party tools may be used to convert files into Operating System allowed file formats. In addition, files published in large websites are maintained online for a longer time while the files published over internet have relatively shorter life span.

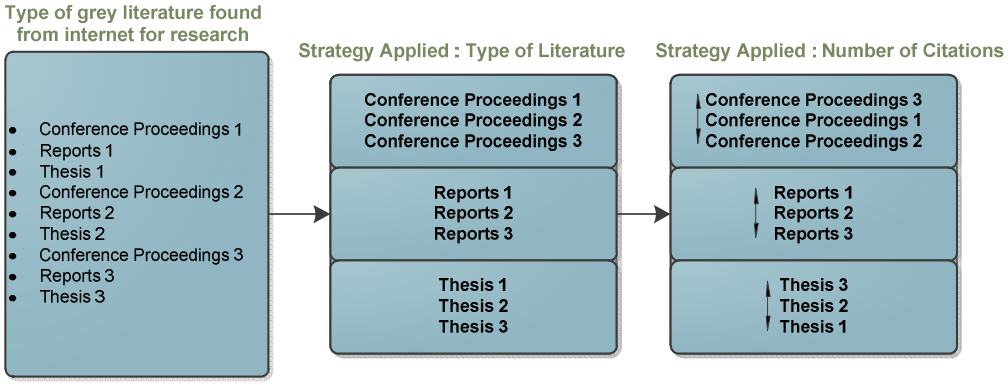
4.3 Conclusion

All the strategies presented above have their own pros and cons. The recommendation is to use hybrid approach while using these strategies. An example combination of these strategies can be as follows;

1. Search the String/ Keyword.
2. Categorize by grey literature type (Conference Proceedings, Thesis, Reports etc.) .
3. Categorize by no. of hits or no. of citations.

After the categorization in step 2, proceed with the application of strategy in the sub categories of list.

Figure 12: Application of Combination of Strategies



Another sample combined strategy may be as follows;

1. Search the String/ Keyword.
2. Categorize by Methodology (SLR, Case-Study, Experiment etc.).
3. Categorize the sub-category by Country.

The following strategy can also be used;

1. Search the String/ Keyword.
2. Categorize by Source (Govt. Institution or University or Company or Individual).
3. Categorize by Country.

There are many different combinations which can be adopted in order to fetch quality data from internet. It totally depends on the researcher to select a certain combination of strategies which suits his research requirements.

4.4 Quality of Grey Literature

During analysis of grey literature types, we have found that the quality of grey literature is subjective. The quality of grey document is hard to assess. This is due to the below characteristics of grey literature;

1. Doesn't follow any research methodology
2. Doesn't report results of study in a proper format
3. Validation procedures for study results are not defined.

Grey Literature does not usually pass through a rigorous peer-review so the quality becomes questionable. Based on our experience, we came up with the following checklist to assess the quality of grey literature. The checklist contains questions formulated specifically to assess the quality of grey literature.

The checklist is inspired and motivated from DARE (Database of Abstracts of Reviews of Effects) quality assessment checklist widely used for assessing the quality of systematic literature reviews. Also, the same checklist is used by Kitchenham et al [9] in order to assess the quality selected SLRs. We formed the questions and scores based on our experience gained throughout this study.

Table 21: Checklist for Grey Literature Assessment

No.	Question	Answer
1	What is the research methodology?	Yes/ No/ Partially
2	Does the article produce repeatable results?	Yes/ No/ Partially
3	Does the article have sufficient bibliographical information? (Author, Stakeholders, Date)	Yes/ No/ Partially
4	Does other reviewed paper cite the document?	Yes/ No
5	Is the document produced by international organization or research lab?	Yes/ No
6	Is the article commented online in forums/ discussion boards?	Yes/No
7	If discussed online, do the comments contribute to the original article findings?	Yes/No
8	Does the author belong to academic/ research institute?	Yes/No

A Score of 1 can be assigned for a 'YES', 0.5 for 'PARTIALLY' and 0 for a 'NO' for the checklist questions. The questions are based on our analysis of grey literature documents during this study. During our study, we have examined a wide range of grey literature and found that the documents produced by international organizations like IBM, Carlen Melon Software Engineering institute and research labs like Simula are of high quality. These documents follow defined research methodology and the validity of results is evaluated frequently. Some of these grey documents are results of

funded research projects. Similarly, we have observed that grey documents having important findings are likely to be discussed online or cited in other papers [5]. Although, this particular criterion does not say much about the quality of content however; it can give an indicative value of importance of document. We have also observed that filtering out irreverent content/ literature is now becoming a major challenge in evaluating the quality of grey literature and the categorization strategies presented in this chapter can help to filter out irreverent information.

We have not evaluated the validity of the formulated checklist because of time constraints and it is included in the future extension of this work.

5 DISCUSSION

The section will discuss the results and limitations to our thesis study. Despite the importance of grey literature during systematic literature reviews, we have found out that the minimal level of grey literature evidence is 9% . Most of the literature, which is included as primary studies in SLRs, is published and peer-reviewed. Grey literature has gained more importance in 'Health and Medical Science' research because of the sensitivity of research topics about human health and life. The inclusion of grey trials is necessary to limit any publication bias in Health Science [15]. The publication bias is a research that focuses mainly on positive published results. We have found out that in the field of Software Engineering, researchers undertake systematic literature reviews to strengthen their synthesis with the use of peer-reviewed articles. In the following section, we state our answers to research questions, which were figured out in the beginning of this study.

RQ1. What is the extent of usage of grey literature in systematic reviews?

After investigation of 6307 primary studies, we have found out that the percentage of grey evidence is 9% in our selected SLRs. Among the total 6307 primary studies, 582 studies were classified as grey literature. While analyzing the 582 grey links, we noticed that most of the grey literature consisted of conference proceedings and technical reports (68%). The research results in reports and proceedings are more detailed and specific than in journals and these results are available months before the official publication in traditional databases. The main concern with grey literature is that it is not properly peer-reviewed and if any peer-reviewed material exists, it is not thorough and exhaustive. On the contrary, the Software Engineering workshops are closer to conferences in terms of the rigor of their review process. Section 3.6.1 presents the results of grey evidence finding.

RQ2. What are the strategies that can be used to categorize grey literature (non-peer reviewed)?
RQ2.1 what are the strengths and weaknesses of those categorization strategies?

We didn't find any evidence of literature that report the categorization strategies for grey literature. The section no. 4 presents the strategies that can be utilized while performing grey literature review. The strategies are a result of the knowledge learned about the grey literature during this course of study. We have formulated a total of 13 strategies with their respective strengths and weaknesses. We believe that if these strategies are used in a hybrid way, one can retrieve specific grey literature according to one's research interest.

RQ3. What is the extent of usage of grey literature in information synthesis (in order to reject or strengthen any arguments) ?

Section 3.6.2 presents the result of analysis of grey literature in the synthesis section of SLRs. While finding for the evidence of grey literature in the synthesis section of SLRs, we have observed the extent to be 93.2 percent. A total of 543 grey links out of 582 were found in the synthesis section. This implies that 93.2 percent of the grey articles selected as primary studies are used to strengthen or reject arguments in SLRs. However, we also noticed that the SLRs don't follow any predefined or standard format for reporting the results and synthesis section. Some SLRs have only discussed the selection and extraction process. Also, the terminologies used for the synthesis section vary among researchers. The synthesis section is mentioned as "Analyses", "Interpretation", "Aggregation" and "Inferences". We also came across

SLRs that did not explicitly mention the studies in the synthesis section; rather a summarized aggregation is presented in the synthesis sections [30]. In such cases, it was not possible to check for the grey evidence in synthesis section. Thus, we assumed that all the primary studies were used in synthesis.

RQ4. To what extent does Google Scholar bring the same results as we get from commercial research databases such as ACM, IEEE, SpringerLink, and ScienceDirect?
RQ4.1 To what extent does GS has indexed research papers?

While finding for the evidence of grey literature in the synthesis section of SLRs, we have observed the evidence extent to be 93.2 %. This means that 93.2 percent of the grey articles selected as primary studies are used to strengthen or reject arguments in SLRs. However, we have also noticed that the SLRs don't follow any predefined or standard format for reporting the results and synthesis section. Some of the SLRs have only discussed the selection and extraction process. The terminologies used for the synthesis section also vary among researchers. The synthesis section is mentioned as "Analyses", "Interpretation", "Aggregation" and "Inferences". We also came across SLRs that do not explicitly mention the studies in the synthesis section; rather a summarized aggregation is presented in the synthesis sections [30]. In such cases, it was not possible to check for the grey evidence in the synthesis section thus we assumed that all the primary studies are used in synthesis.

We have presented strategies about how to take grey literature into account for the research needs. The problem with grey literature is the quality which is highly subjective. Grey literature often stems from unreliable and unreferenced sources [7]. The open source information is at ease of access but suffers from the low 'signal-to-noise ratio' i.e. after filtering we only get few articles that are of reasonable quality and can be used to strengthen or reject arguments in research. This situation worsens in the online grey domain as there are hundreds of scientific blogs, forums and websites, producing grey literature daily.

Based on our experience during this study, we believe that the volatility and limited access problem of grey literature is now scaled down, due to the wide spread of internet. The reason is that majority of all content published in print form is currently available on internet for worldwide access. The government, corporations and individuals publish most of their data online for public access. Not only websites but also blogs, discussion boards and Facebook are disseminating large amounts of information. According to our experience, the social media, if used responsibly, can be a good source to categorize related content. The tools like Google+1, Facebook shares and number of tweets can be used to assess content popularity and quality subjectively.

Based on the results of Google Scholar, we have seen that it was able to retrieve a slew such as 96% of primary studies of SLRs. Most of the primary studies that were not found using Google Scholar were of grey sources, a total of 108 studies (almost 40 percent) were not found. We tried to find the characteristics of studies, which were not found in Google Scholar, however did not succeed. We also deduced that the primary studies, which were not found in Google Scholar, had heterogeneous characteristics. Thus, we could not infer much information about the type of studies that Google Scholar was generally not able to retrieve. Based on our observations, Google Scholar does not provide many options for searching the content on it. The traditional databases like IEEE and ScienceDirect provide more search fields to create search strings. However, Google Scholar provides much simpler and clean interface to search articles as compared to the traditional databases.

During our Google Scholar analysis, we noticed that the primary studies, which were not retrievable in GS, were retrievable through Google. There were only few primary studies that were not found in both Google Scholar and Google. All of these few primary studies, which were not found in Google and Google Scholar, were presented either in conferences or in workshops. One more interesting finding was that these few primary studies were either published before year 2000 or belonged to specific conference proceedings. So collectively, we were able to find nearly all the primary studies using combination of Google Scholar and Google. Based on our study in this thesis work, we believe that Google Scholar can become a foremost choice for searching literature in the future because of its efficacy in finding scholarly literature .

6 CONCLUSION

The aim of this study is to explore the evidence of grey literature and its usage in information synthesis during systematic literature reviews. The secondary aim of this study is also to investigate the efficacy of using Google Scholar for the retrieval of primary studies. We conducted a comprehensive systematic literature review of 138 SLRs to find the evidence of grey literature and its usage in information synthesis during systematic literature reviews. There were total 6307 primary studies that were investigated for grey literature evidence. The total evidence of grey literature in SLRs is found to be 9%. The study also presents strategies to categorize grey literature for systematically considering grey literature. We have also found out that conference proceedings and technical reports produced by international organizations and research labs form major proportion (68%) of grey primary studies included in SLRs. Our findings support that the GL-conference proceedings are of significant importance to the field of grey literature. The study also shows that software engineering researchers have a trend to include the recent research publications and documents as primary studies in their SLRs.

We also searched 6307 primary studies in Google Scholar and found 6027 of the total (i.e. 96%) were in GS. Looking at our study results, we conclude that Google Scholar has a great potential to become a major source of scientific articles for research purpose. We were able to find large number of published and peer reviewed primary studies from GS. The studies not found in Google Scholar, were mostly grey literature. However, we also saw that the literature not found in Google Scholar, could be found at Google. In our point of view, if we use combination of Google Scholar and Google for our research, search percentage is significantly increased.

In the beginning of our research, we expected the evidence of grey literature to be significant in SLRs however; our findings are quite different now. Our study shows that there isn't significant evidence of grey literature in SLRs' of Software Engineering. Most of the literature included in SLRs is published and peer reviewed. So, the question still remains that with the increase in grey sources and rapid feedback over internet, will the researchers in software engineering be able to adapt grey literature as an important source of information?

We believe that our categorization strategies of grey literature will help the researchers to consider grey literature more systematically. Although the quality of grey literature is subjective, the categorization strategies and quality assessment checklist can help to filter out irreverent information from the pile of data.

6.1 Limitations

We have considered the following as main limitations to our study.

Access to literature: While analyzing primary studies, we confronted some studies that were not accessible using our university subscription. We had no other choice but to exclude them from our analysis.

Primary Studies List: In our tertiary study, we came up with SLRs that had not mentioned included primary studies. In some cases, even if the primary studies were mentioned in SLR, it was hard to extract them since they were mixed up with the references included in the SLR. Also in some cases, the authors had used images for the included primary studies and it was not possible to extract data from those images.

Time Constraint: This study was conducted on a small-scale because of time frame limitation. The study was conducted over 6 months of time frame.

Human Error: We tried to remain honest and credible throughout the study. Despite, there are still chances of human error because of dealing with large amount of data.

6.2 Future Work

The study presents the results of only four major databases i.e. IEEE, ACM, ScienceDirect and SpringerLink. A future consideration will be to include more databases for analysis and conduct the study on a larger scale. We have extracted data from SLRs in a systematic and easily understandable format in Microsoft Excel files and there is a great opportunity to extract other useful analysis from the data. Some of the analysis opportunities include analyzing the types of grey literature in the grey evidence and forming scale to measure quality of grey literature. We have found a trend of grey literature usage for the past few years and can predict usage of grey literature in the coming years.

As an initial step, we have proposed categorization strategies for grey literature. As a next step to utilizing the grey literature to maximum extent, categorization strategies can be verified through experiment(s), finalized methods, and techniques.

7 REFERENCES

[1] Gelfand, J.; King, J.L.; "Grey market science: research libraries, grey literature, and the legitimization of scientific discourse in the Internet age," Socioeconomic Dimensions of Electronic Publishing Workshop, 1998. Proceedings, vol., no., pp.115-120, 23-25 Apr 1998.

[2] "Response required," Nature, vol. 468, no. 7326, p. 867, Dec. 2010. [Online]. Available: http://dx.doi.org/10.1038/468867a [July 23, 2012]

[3] Wikipedia, "Grey Literature", [Online], Available: http://en.wikipedia.org/wiki/Gray_literature [January 25, 2012]

[4] Denda, Kayo. "Fugitive Literature in the Cross Hairs: An Examination of Bibliographic Control and Access". Collection Management 27:2 (2002): 75-86

[5] "Trial by Twitter", Nature, vol. 469, p.286-287, [Online], Available: http://www.nature.com/news/2011/110119/pdf/469286a.pdf [July 23, 2012]

[6] O. Osayande, O. C. Ukpebor. Grey Literature Acquisition and Management: Challenges in Academic Libraries in Africa. Library Philosophy and Practice 2012.

[7] H. Soule, R. Paul Ryan. Gray literature. Technical Briefing, Information Technology summit 1999.

[8] J. Schöpfel. Observations on the Future of Grey Literature. The Grey Journal 2, 2 (2006) 67-76.

[9] B.A. Kitchenham, Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3. Technical Report S.o.C.S.a.M. Software Engineering Group, Keele University and Department of Computer Science University of Durham, 2007.

[10] O. Dieste, A. Grimán, N. Juristo, Developing search strategies for detecting relevant experiments, Empirical Software Engineering (2008).

[11] T. Dybå, T. Dingsöyr, G.K. Hanssen, Applying systematic reviews to diverse study types: an experience report, in: First International Symposium on Empirical Software Engineering and Measurement, 2007 (ESEM 2007), 2007, pp. 225–234.

[12] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," INFORMATION AND SOFTWARE TECHNOLOGY, vol. 51, no. 1, Jan. 2009.

[13] Kitchenham, B.; Brereton, P.; Turner, M.; Niazi, M.; Linkman, S.; Pretorius, R.; Budgen, D.; , "The impact of limited search procedures for systematic literature reviews — A participant-observer case study," Empirical Software Engineering and Measurement, 2009. ESEM 2009. 3rd International Symposium on , vol., no., pp.336-345, 15-16 Oct. 2009

[14] De Castro, P. and Salinetti, S. (2004). Quality of grey literature in the open access era: Privilege and Responsibility. In Farace, Dj, editor, GL5, CONFERENCE PROCEEDINGS. INIST CNRS; City Amsterdam, Dept Commun, External Affairs.

[15] I. Ahmed, A. J. Sutton, and R. D. Riley, "Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey," BRITISH MEDICAL JOURNAL, vol. 344, Jan. 2012. [Online]. Available: http://dx.doi.org/%7B10.1136/bmj.d7762%7D

[16] A. F. J. van Raan, "Bibliometrics and Internet: Some observations and expectations," SCIENTOMETRICS, vol. 50, no. 1, Jan. 2001. [Online]. Available: http://dx.doi.org/%7B10.1023/A:1005690102069%7D

[17] FARACE et J. FRANTZEN (ed.). GL7 Conference Proceedings. Seventh International Conference on Grey Literature: Open Access to Grey Resources. Nancy, 5-6 December 2005.
[18] FARACE D.J., FRANTZEN J., SCHÖPFEL J., STOCK C. et BOEKHORST A.K. (2006). "Access to Grey Content: An Analysis of Grey Literature Based on Citation and Survey Data". In: Quality Assessment of Grey Litearture. vol. 1, pp. 31-40, 2005.
[19] F.L. Joseph, Measuring nominal scale agreement among many raters,‖ *Psychological Bulletin*, vol. 76, 1971, pp. 378-382.
[20] J.R. Landis and G.G. Koch, ―The measurement of observer agreement for categorical data.,‖ *Biometrics*, vol. 33, 1977, p. 159.
[21] Frandsen, T.F. (2009), "The effects of open access on unpublished documents: a case study of economics working papers", Journal of Informetrics, Vol. 3 No. 2, pp. 124-33.
[22] Antelman, K. (2004). "Do open-access articles have a greater research impact?" College and Research Libraries, 65, 372–382.
[23] Google +1 Button." [Online]. Available: http://www.google.com/+1/button/. [Accessed: 19-Jul-2012].
[24] "How to use Google +1 social recommendation feature." [Online]. Available: http://www.techiezone.net/internet/how-to-use-google-1-social-recommendation-feature/. [Accessed: 19-Jul-2012].
[25] Watson, "Resistance to change," American behavioral scientist, vol. 14, no. 5, p. 745, 1971.
[26] M. Vaska, J. Schöpfel, I. Fürstová. Grey Literature Repositories. Radim Bačuvčík - VeRBuM, Zlín 2010.
[27] Felizardo, K.R.; Salleh, N.; Martins, R.M.; Mendes, E.; MacDonell, S.G.; Maldonado, J.C.; , "Using Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews," Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on , vol., no., pp.77-86, 22-23 Sept. 2011
[28] "Google Scholar". Wikipedia. [Online]. Available http://en.wikipedia.org/wiki/Google_Scholar. (Accessed 19-July-2012)
[29] DARE. Database of Abstracts of Reviews of Effects. [Online]. Available at http://www.dimdi.de/static/en/db/dbinfo/cdar94.htm (Last accessed 19-July-2012)
[30] D. S. Cruzes and T. Dybå, "Research synthesis in software engineering: A tertiary study," Information and Software Technology, vol. 53, no. 5, pp. 440–455, May 2011.
[31] B. Kitchenham, R. Pretorius, D. Budgen, O. Pearl Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering – A tertiary study," Information and Software Technology, vol. 52, no. 8, pp. 792–805, Aug. 2010
[32] THE EVALUATION OF GREY LITERATURE USING BIBLIOMETRIC INDICATORS A METHODOLOGICAL PROPOSAL Rosa Di Cesare, Roberta Ruggieri CNR – Istituto di Ricerche sulla Popolazione e le Politiche Sociali
[33] FARACE D.J., FRANTZEN J., SCHÖPFEL J., STOCK J. Citation Analysis and Grey Literature: Stakeholders in the Grey Circuit. In Quality Assessment of Grey Litearture. vol. 1, pp. 31-40
[34] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J., "How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study." Health Technology Assessment 2003, vol. 7: No.1
[35] Hopewell S, McDonald S, Clarke M, Egger M (2007) Grey literature in meta-analyses of randomized trials of health care interventions. Cochrane Database Syst Rev. Issue 2: MR000010. doi:10.1002/14651858.MR000010.pub3.

[36] Conn VS, Valentine JC, Cooper HM, Rantz MJ. Grey literature in meta-analyses. Nurs Res. 2003;52(4):256–261

[37] The use of grey literature in health sciences: a preliminary survey. V Alberani, P De Castro Pietrangeli, A M Mazza Bull Med Libr Assoc. 1990 October; 78(4): 358–363.

[38] L. McAuley, B. Pham, P. Tugwell, D. Moher. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? The Lancet, Volume 356, Issue 9237, Pages 1228 - 1231, 7 October 2000.

8 APPENDIX

8.1 Appendix A: Cohen Kappa Application

FREQUENCIES VARIABLES=Difference
/ORDER=ANALYSIS.

Frequencies

Notes

Output Created		29-Mar-2012 16:47:15
Comments		
Input	Active Dataset	DataSet5
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	71
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data.
Syntax		FREQUENCIES VARIABLES=Difference /ORDER=ANALYSIS.
Resources	Processor Time	00 00:00:00.000
	Elapsed Time	00 00:00:00.005

[DataSet5]

Statistics

Difference

N	Valid	37
	Missing	34

Difference

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1.00	4	5.6	10.8	10.8
	.00	31	43.7	83.8	94.6
	1.00	2	2.8	5.4	100.0
	Total	37	52.1	100.0	

Missing	System	34	47.9		
Total		71	100.0		

```
CROSSTABS
  /TABLES=RaterA BY RaterB
  /FORMAT=AVALUE TABLES
  /STATISTICS=KAPPA
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.
```

Crosstabs

Notes

Output Created		29-Mar-2012 16:50:04
Comments		
Input	Active Dataset	DataSet5
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	71
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
Syntax		CROSSTABS /TABLES=RaterA BY RaterB /FORMAT=AVALUE TABLES /STATISTICS=KAPPA /CELLS=COUNT EXPECTED /COUNT ROUND CELL.
Resources	Processor Time	00 00:00:00.000
	Elapsed Time	00 00:00:00.008
	Dimensions Requested	2
	Cells Available	174762

[DataSet5]

Case Processing Summary

	Cases		
	Valid	Missing	Total

	N	Percent	N	Percent	N	Percent
RaterA * RaterB	37	52.1%	34	47.9%	71	100.0%

RaterA * RaterB Crosstabulation

			RaterB			
			0	1	3	Total
RaterA	0	Count	6	4	0	10
		Expected Count	2.2	7.3	.5	10.0
	1	Count	2	23	0	25
		Expected Count	5.4	18.2	1.4	25.0
	3	Count	0	0	2	2
		Expected Count	.4	1.5	.1	2.0
Total		Count	8	27	2	37
		Expected Count	8.0	27.0	2.0	37.0

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.636	.137	4.606	.000
N of Valid Cases		37			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

8.2 Appendix B: Pilot Search File Format

Given is the link to Google Drive excel file used to document pilot search.

https://docs.google.com/spreadsheet/ccc?key=0AgpdubbHEGGjdHU1SXJuSHdqNDc1aHpPcTJ2TmM0a1E

8.3 Included Primary Studies (SLRs)

8.4 Primary studies from ScienceDirect

- S.-J. Huang, N.-H. Chiu, and Y.-J. Liu, "A comparative evaluation on the accuracies of software effort estimates from clustered data," *Information and Software Technology*, vol. 50, no. 9–10, pp. 879–888, Aug. 2008.
- K.-J. Stol, M. A. Babar, P. Avgeriou, and B. Fitzgerald, "A comparative study of challenges in integrating Open Source Software and Inner Source Software," *Information and Software Technology*, vol. 53, no. 12, pp. 1319–1336, Dec. 2011
- Ø. Hauge, C. Ayala, and R. Conradi, "Adoption of open source software in software-intensive organizations – A systematic literature review," *Information and Software Technology*, vol. 52, no. 11, pp. 1133–1154, Nov. 2010.
- A. Talaei-Khoei, P. Ray, N. Parameshwaran, and L. Lewis, "A framework for awareness maintenance," *Journal of Network and Computer Applications*, vol. 35, no. 1, pp. 199–210, Jan. 2012.
- M. N. Wicks and R. G. Dewar, "A new research agenda for tool integration," *Journal of Systems and Software*, vol. 80, no. 9, pp. 1569–1585, Sep. 2007.
- M. J. Escalona, J. J. Gutierrez, M. Mejías, G. Aragón, I. Ramos, J. Torres, and F. J. Domínguez, "An overview on test generation from functional requirements," *Journal of Systems and Software*, vol. 84, no. 8, pp. 1379–1393, Aug. 2011.
- M. Jørgensen, "A review of studies on expert estimation of software development effort," *Journal of Systems and Software*, vol. 70, no. 1–2, pp. 37–60, Feb. 2004.
- S. Heckman and L. Williams, "A systematic literature review of actionable alert identification techniques for automated static code analysis," *Information and Software Technology*, vol. 53, no. 4, pp. 363–387, Apr. 2011.
- L. M. Karg, M. Grottke, and A. Beckhaus, "A systematic literature review of software quality cost research," *Journal of Systems and Software*, vol. 84, no. 3, pp. 415–427, Mar. 2011.
- G. S. Walia and J. C. Carver, "A systematic literature review to identify and classify software requirement errors," *Information and Software Technology*, vol. 51, no. 7, pp. 1087–1109, Jul. 2009.
- P. A. da Mota Silveira Neto, I. do Carmo Machado, J. D. McGregor, E. S. de Almeida, and S. R. de Lemos Meira, "A systematic mapping study of software product lines testing," *Information and Software Technology*, vol. 53, no. 5, pp. 407–423, May 2011.
- F. Elberzhager, J. Münch, and V. T. N. Nha, "A systematic mapping study on the combination of static and dynamic quality assurance techniques," *Information and Software Technology*, vol. 54, no. 1, pp. 1–15, Jan. 2012.
- G. Holl, P. Grünbacher, and R. Rabiser, "A systematic review and an expert survey on capabilities supporting multi product lines," *Information and Software Technology*, vol. 54, no. 8, pp. 828–852, Aug. 2012.
- M. S. Ali, M. Ali Babar, L. Chen, and K.-J. Stol, "A systematic review of comparative evidence of aspect-oriented programming," *Information and Software Technology*, vol. 52, no. 9, pp. 871–887, Sep. 2010.
- M. Khurum and T. Gorschek, "A systematic review of domain analysis solutions for product lines," *Journal of Systems and Software*, vol. 82, no. 12, pp. 1982–2003, Dec. 2009.
- L. Chen and M. Ali Babar, "A systematic review of evaluation of variability management approaches in software product lines," *Information and Software Technology*, vol. 53, no. 4, pp. 344–362, Apr. 2011.
- M. Höst and A. Oručević-Alagić, "A systematic review of research on open source software in commercial software product development," *Information and Software Technology*, vol. 53, no. 6, pp. 616–624, Jun. 2011.

- W. Afzal, R. Torkar, and R. Feldt, "A systematic review of search-based testing for non-functional system properties," *Information and Software Technology*, vol. 51, no. 6, pp. 957–976, Jun. 2009.
- D. Mellado, C. Blanco, L. E. Sánchez, and E. Fernández-Medina, "A systematic review of security requirements engineering," *Computer Standards & Interfaces*, vol. 32, no. 4, pp. 153–165, Jun. 2010.
- H. P. Breivold, I. Crnkovic, and M. Larsson, "A systematic review of software architecture evolution research," *Information and Software Technology*, vol. 54, no. 1, pp. 16–40, Jan. 2012.
- F. J. Lucas, F. Molina, and A. Toval, "A systematic review of UML model consistency management," *Information and Software Technology*, vol. 51, no. 12, pp. 1631–1645, Dec. 2009.
- E. Engström, P. Runeson, and M. Skoglund, "A systematic review on regression test selection techniques," *Information and Software Technology*, vol. 52, no. 1, pp. 14–30, Jan. 2010.
- M. Svahnberg, T. Gorschek, R. Feldt, R. Torkar, S. B. Saleem, and M. U. Shafique, "A systematic review on strategic release planning models," *Information and Software Technology*, vol. 52, no. 3, pp. 237–248, Mar. 2010.
- D. Benavides, S. Segura, and A. Ruiz-Cortés, "Automated analysis of feature models 20 years later: A literature review," *Information Systems*, vol. 35, no. 6, pp. 615–636, Sep. 2010.
- S. U. Khan, M. Niazi, and R. Ahmad, "Barriers in the selection of offshore software development outsourcing vendors: An exploratory study using a systematic literature review," *Information and Software Technology*, vol. 53, no. 7, pp. 693–706, Jul. 2011.
- C. Blanco, J. Lasheras, E. Fernández-Medina, R. Valencia-García, and A. Toval, "Basis for an integrated security ontology according to a systematic review of existing proposals," *Computer Standards & Interfaces*, vol. 33, no. 4, pp. 372–388, Jun. 2011.
- B. J. Williams and J. C. Carver, "Characterizing software architecture changes: A systematic review," *Information and Software Technology*, vol. 52, no. 1, pp. 31–51, Jan. 2010.
- P. Mohagheghi, V. Dehlen, and T. Neple, "Definitions and approaches to model quality in model-based software development – A review of literature," *Information and Software Technology*, vol. 51, no. 12, pp. 1646–1669, Dec. 2009.
- M. Turner, B. Kitchenham, P. Brereton, S. Charters, and D. Budgen, "Does the technology acceptance model predict actual use? A systematic literature review," *Information and Software Technology*, vol. 52, no. 5, pp. 463–479, May 2010.
- J. Pardillo and C. Cachero, "Domain-specific language modelling with UML profiles by decoupling abstract and concrete syntaxes," *Journal of Systems and Software*, vol. 83, no. 12, pp. 2591–2606, Dec. 2010.
- T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Information and Software Technology*, vol. 50, no. 9–10, pp. 833–859, Aug. 2008.
- A. S. Jadhav and R. M. Sonar, "Evaluating and selecting software packages: A review," *Information and Software Technology*, vol. 51, no. 3, pp. 555–563, Mar. 2009.
- M. Shepperd and S. MacDonell, "Evaluating prediction systems in software project estimation," *Information and Software Technology*, vol. 54, no. 8, pp. 820–827, Aug. 2012.
- M. Staples and M. Niazi, "Experiences using systematic review guidelines," *Journal of Systems and Software*, vol. 80, no. 9, pp. 1425–1437, Sep. 2007.
- S. U. Khan, M. Niazi, and R. Ahmad, "Factors influencing clients in the selection of offshore software outsourcing vendors: An exploratory study using a systematic

literature review," *Journal of Systems and Software*, vol. 84, no. 4, pp. 686–699, Apr. 2011.

- A. L. Mesquida, A. Mas, E. Amengual, and J. A. Calvo-Manzano, "IT Service Management Process Improvement based on ISO/IEC 15504: A systematic review," *Information and Software Technology*, vol. 54, no. 3, pp. 239–247, Mar. 2012.
- F. O. Bjørnson and T. Dingsøyr, "Knowledge management in software engineering: A systematic review of studied concepts, findings and research methods used," *Information and Software Technology*, vol. 50, no. 11, pp. 1055–1068, Oct. 2008.
- P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, Apr. 2007.
- K. Petersen, "Measuring and predicting software productivity: A systematic map and review," *Information and Software Technology*, vol. 53, no. 4, pp. 317–343, Apr. 2011.
- H. Sharp, N. Baddoo, S. Beecham, T. Hall, and H. Robinson, "Models of motivation in software engineering," *Information and Software Technology*, vol. 51, no. 1, pp. 219–233, Jan. 2009.
- S. Beecham, N. Baddoo, T. Hall, H. Robinson, and H. Sharp, "Motivation in Software Engineering: A systematic literature review," *Information and Software Technology*, vol. 50, no. 9–10, pp. 860–878, Aug. 2008.
- W. Afzal and R. Torkar, "On the application of genetic programming for software engineering predictive modeling: A systematic review," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11984–11997, Sep. 2011.
- J. Nicolás and A. Toval, "On the generation of requirements specifications from software engineering models: A systematic literature review," *Information and Software Technology*, vol. 51, no. 9, pp. 1291–1307, Sep. 2009.
- J. Valaski, A. Malucelli, and S. Reinehr, "Ontologies application in organizational learning: A literature review," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7555–7561, Jun. 2012.
- S. Lane and I. Richardson, "Process models for service-based applications: A systematic literature review," *Information and Software Technology*, vol. 53, no. 5, pp. 424–439, May 2011.
- R. Prikladnicki and J. L. N. Audy, "Process models in the practice of distributed software development: A systematic review of the literature," *Information and Software Technology*, vol. 52, no. 8, pp. 779–791, Aug. 2010.
- A. M. Magdaleno, C. M. L. Werner, and R. M. de Araujo, "Reconciling software development models: A quasi-systematic review," *Journal of Systems and Software*, vol. 85, no. 2, pp. 351–369, Feb. 2012.
- V. Alves, N. Niu, C. Alves, and G. Valença, "Requirements engineering for software product lines: A systematic literature review," *Information and Software Technology*, vol. 52, no. 8, pp. 806–820, Aug. 2010.
- D. S. Cruzes and T. Dybå, "Research synthesis in software engineering: A tertiary study," *Information and Software Technology*, vol. 53, no. 5, pp. 440–455, May 2011.
- J. C. de Almeida Biolchini, P. G. Mian, A. C. C. Natali, T. U. Conte, and G. H. Travassos, "Scientific research ontology to support systematic review in software engineering," *Advanced Engineering Informatics*, vol. 21, no. 2, pp. 133–151, Apr. 2007.
- F. Q. B. da Silva, A. L. M. Santos, S. Soares, A. C. C. França, C. V. F. Monteiro, and F. F. Maciel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," *Information and Software Technology*, vol. 53, no. 9, pp. 899–913, Sep. 2011.

- S. Grimstad, M. Jørgensen, and K. Moløkken-Østvold, “Software effort estimation terminology: The tower of Babel,” *Information and Software Technology*, vol. 48, no. 4, pp. 302–310, Apr. 2006.
- A. Ampatzoglou and I. Stamelos, “Software engineering research for computer games: A systematic review,” *Information and Software Technology*, vol. 52, no. 9, pp. 888–901, Sep. 2010.
- E. Engström and P. Runeson, “Software product line testing – A systematic mapping study,” *Information and Software Technology*, vol. 53, no. 1, pp. 2–13, Jan. 2011.
- S. Barney, K. Petersen, M. Svahnberg, A. Aurum, and H. Barney, “Software quality trade-offs: A systematic map,” *Information and Software Technology*, vol. 54, no. 7, pp. 651–662, Jul. 2012.
- S. G. MacDonell, “Software source code sizing using fuzzy logic modeling,” *Information and Software Technology*, vol. 45, no. 7, pp. 389–404, May 2003.
- J. Wen, S. Li, Z. Lin, Y. Hu, and C. Huang, “Systematic literature review of machine learning based software development effort estimation models,” *Information and Software Technology*, vol. 54, no. 1, pp. 41–59, Jan. 2012.
- B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering – A systematic literature review,” *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, Jan. 2009.
- B. Kitchenham, R. Pretorius, D. Budgen, O. Pearl Brereton, M. Turner, M. Niazi, and S. Linkman, “Systematic literature reviews in software engineering – A tertiary study,” *Information and Software Technology*, vol. 52, no. 8, pp. 792–805, Aug. 2010.
- M. Staples and M. Niazi, “Systematic review of organizational motivations for adopting CMM-based SPI,” *Information and Software Technology*, vol. 50, no. 7–8, pp. 605–620, Jun. 2008.
- M. Palacios, J. García-Fanjul, and J. Tuya, “Testing in Service Oriented Architectures with dynamic binding: A mapping study,” *Information and Software Technology*, vol. 53, no. 3, pp. 171–189, Mar. 2011.
- J. E. Hannay, T. Dybå, E. Arisholm, and D. I. K. Sjøberg, “The effectiveness of pair programming: A meta-analysis,” *Information and Software Technology*, vol. 51, no. 7, pp. 1110–1122, Jul. 2009.
- J. Portillo-Rodríguez, A. Vizcaíno, M. Piattini, and S. Beecham, “Tools used in Global Software Engineering: A systematic mapping review,” *Information and Software Technology*, vol. 54, no. 7, pp. 663–685, Jul. 2012.
- E. Loukis, K. Pazalos, and A. Salagara, “Transforming e-services evaluation data into business analytics using value models,” *Electronic Commerce Research and Applications*, vol. 11, no. 2, pp. 129–141, Mar. 2012.
- A. Fernandez, E. Insfran, and S. Abrahão, “Usability evaluation methods for the web: A systematic mapping study,” *Information and Software Technology*, vol. 53, no. 8, pp. 789–817, Aug. 2011.
- M. Dyer; M. Shepperd; “Systematic Reviews in Evideicne based Software technology and Software Engineering” *Information and Software Technology*, vol. 47, no. 8, pp. 792–805, 2005.
- Cagatay Catal, Banu Diri, “A systematic review of software fault prediction studies”, *Expert Systems with Applications*, Volume 36, Issue 4, May 2009, Pages 7346-7354, ISSN 0957-4174, 10.1016/j.eswa.2008.10.027.

8.5 Primary studies from IEEEXplore

- S. M. Mitchell and C. B. Seaman, “A comparison of software cost, duration, and quality for waterfall vs. iterative and incremental development: A systematic

review," in *3rd International Symposium on Empirical Software Engineering and Measurement, 2009. ESEM 2009*, 2009, pp. 511 –515.

- J. F. Bastos, P. A. da Mota Silveira Neto, E. S. de Almeida, and S. R. de Lemos Meira, "Adopting software product lines: A systematic mapping study," in *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, 2011, pp. 11 –20.
- S. Jalali and C. Wohlin, "Agile Practices in Global Software Engineering - A Systematic Map," in *2010 5th IEEE International Conference on Global Software Engineering (ICGSE)*, 2010, pp. 45 –54.
- S. Kakarla, S. Momotaz, and A. S. Namin, "An Evaluation of Mutation and Data-Flow Testing: A Meta-analysis," in *2011 IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, 2011, pp. 366 – 375.
- M. Sulayman and E. Mendes, "An extended systematic review of software process improvement in small and medium Web companies," in *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, 2011, pp. 134 –143.
- K. A. Sedek, S. Sulaiman, and M. A. Omar, "A systematic literature review of interoperable architecture for e-government portals," in *Software Engineering (MySEC), 2011 5th Malaysian Conference in*, 2011, pp. 82 –87.
- S. P. Shashank, P. Chakka, and D. V. Kumar, "A systematic literature survey of integration testing in component-based software engineering," in *2010 International Conference on Computer and Communication Technology (ICCCT)*, 2010, pp. 562 – 568.
- E. Barreiros, A. Almeida, J. Saraiva, and S. Soares, "A Systematic Mapping Study on Software Engineering Testbeds," in *2011 International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2011, pp. 107 –116.
- C. Blanco, J. Lasheras, R. Valencia-Garcia, E. Fernandez-Medina, A. Toval, and M. Piattini, "A Systematic Review and Comparison of Security Ontologies," in *Third International Conference on Availability, Reliability and Security, 2008. ARES 08*, 2008, pp. 813 –820.
- M. Jorgensen and M. Shepperd, "A Systematic Review of Software Development Cost Estimation Studies," *IEEE Transactions on Software Engineering*, vol. 33, no. 1, pp. 33 –53, Jan. 2007.
- M. Riaz, E. Mendes, and E. Tempero, "A systematic review of software maintainability prediction and metrics," in *3rd International Symposium on Empirical Software Engineering and Measurement, 2009. ESEM 2009*, 2009, pp. 367 –377.
- H. P. Breivold, M. A. Chauhan, and M. A. Babar, "A Systematic Review of Studies of Open Source Software Evolution," in *Software Engineering Conference (APSEC), 2010 17th Asia Pacific*, 2010, pp. 356 –365.
- S. Ali, L. C. Briand, H. Hemmati, and R. K. Panesar-Walawege, "A Systematic Review of the Application and Empirical Investigation of Search-Based Test Case Generation," *IEEE Transactions on Software Engineering*, vol. 36, no. 6, pp. 742 – 762, Dec. 2010.
- T. Tahir and A. Jafar, "A Systematic Review on Software Measurement Programs," in *Frontiers of Information Technology (FIT), 2011*, 2011, pp. 39 –44.
- B. Cornelissen, A. Zaidman, A. van Deursen, L. Moonen, and R. Koschke, "A Systematic Survey of Program Comprehension through Dynamic Analysis," *IEEE Transactions on Software Engineering*, vol. 35, no. 5, pp. 684 –702, Oct. 2009.
- [P81] P. Karpati, G. Sindre, and A. L. Opdahl, "Characterising and Analysing Security Requirements Modelling Initiatives," in *2011 Sixth International Conference on Availability, Reliability and Security (ARES)*, 2011, pp. 710 –715.
- [P82] S. G. MacDonell and M. J. Shepperd, "Comparing Local and Global Software Effort Estimation Models – Reflections on a Systematic Review," in *First*

International Symposium on Empirical Software Engineering and Measurement, 2007. ESEM 2007, 2007, pp. 401 –409.

- S. Beecham, J. Noll, I. Richardson, and N. Ali, "Crafting a Global Teaming Model for Architectural Knowledge," in *2010 5th IEEE International Conference on Global Software Engineering (ICGSE)*, 2010, pp. 55 –63.
- B. A. Kitchenham, E. Mendes, and G. H. Travassos, "Cross versus Within-Company Cost Estimation Studies: A Systematic Review," *IEEE Transactions on Software Engineering*, vol. 33, no. 5, pp. 316 –329, May 2007.
- A. Davis, O. Dieste, A. Hickey, N. Juristo, and A. M. Moreno, "Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review," in *Requirements Engineering, 14th IEEE International Conference*, 2006, pp. 179 –188.
- N. Salleh, E. Mendes, and J. Grundy, "Empirical Studies of Pair Programming for CS/SE Teaching in Higher Education: A Systematic Literature Review," *IEEE Transactions on Software Engineering*, vol. 37, no. 4, pp. 509 –525, Aug. 2011.
- P. Sfetsos and I. Stamelos, "Empirical Studies on Quality in Agile Practices: A Systematic Literature Review," in *Quality of Information and Communications Technology (QUATIC), 2010 Seventh International Conference on the*, 2010, pp. 44 –53.
- M. Unterkalmsteiner, T. Gorschek, A. K. M. M. Islam, C. K. Cheng, R. B. Permadi, and R. Feldt, "Evaluation and Measurement of Software Process Improvement #x02014;A Systematic Literature Review," *IEEE Transactions on Software Engineering*, vol. 38, no. 2, pp. 398 –424, Apr. 2012.
- J. M. Moreno-Rivera and E. Navarro, "Evaluation of SPL Approaches for WebGIS Development: SIGTel, a Case Study," in *2011 44th Hawaii International Conference on System Sciences (HICSS)*, 2011, pp. 1 –10.
- A. Causevic, D. Sundmark, and S. Punnekkat, "Factors Limiting Industrial Adoption of Test Driven Development: A Systematic Review," in *2011 IEEE Fourth International Conference on Software Testing, Verification and Validation (ICST)*, 2011, pp. 337 –346.
- R. B. Svensson, M. Ho&#x0308 andst, and B. Regnell, "Managing Quality Requirements: A Systematic Review," in *2010 36th EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA)*, 2010, pp. 261 –268.
- R. Y. Huang and J. Symonds, "Mobile marketing evolution: Systematic literature review on multi-channel communication and multi-characteristics campaign," in *Enterprise Distributed Object Computing Conference Workshops, 2009. EDOCW 2009. 13th*, 2009, pp. 157 –165.
- A. C. C. Franca, T. B. Gouveia, P. C. F. Santos, C. A. Santana, and F. Q. B. da Silva, "Motivation in software engineering: A systematic review update," in *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, 2011, pp. 154 –163.
- P. Brereton, M. Turner, and R. Kaur, "Pair programming as a teaching tool: a student review of empirical studies," in *22nd Conference on Software Engineering Education and Training, 2009. CSEET '09*, 2009, pp. 240 –247.
- N. Condori-Fernandez, M. Daneva, K. Sikkel, and A. Herrmann, "Practical relevance of experiments in comprehensibility of requirements specifications," in *2011 First International Workshop on Empirical Requirements Engineering (EmpiRE)*, 2011, pp. 21 –28.
- M. J. Monasor, A. Vizcai&#x0301 andno, M. Piattini, and I. Caballero, "Preparing Students and Engineers for Global Software Development: A Systematic Review," in *2010 5th IEEE International Conference on Global Software Engineering (ICGSE)*, 2010, pp. 177 –186.
- H. F. Landim, A. B. Albuquerque, and T. C. Macedo, "Procedures and Conditions that Influence on the Efficiency of Some Agile Practices," in *Quality of Information and*

Communications Technology (QUATIC), 2010 Seventh International Conference on the, 2010, pp. 385 –390.

- E. Hossain, M. A. Babar, H. Paik, and J. Verner, "Risk Identification and Mitigation Processes for Using Scrum in Global Software Development: A Conceptual Framework," in *Software Engineering Conference, 2009. APSEC '09. Asia-Pacific*, 2009, pp. 457 –464.
- G. K. Hanssen, D. Smite, and N. B. Moe, "Signs of Agile Trends in Global Software Engineering Research: A Tertiary Study," in *2011 Sixth IEEE International Conference on Global Software Engineering Workshop (ICGSEW)*, 2011, pp. 17 – 23.
- H. Zhang, B. Kitchenham, and D. Pfahl, "Software Process Simulation Modeling: Facts, Trends and Directions," in *Software Engineering Conference, 2008. APSEC '08. 15th Asia-Pacific*, 2008, pp. 59 –66.
- C. Monteiro, D. F. Arcoverde, F. Q. B. da Silva, and H. S. Ferreira, "Software support for the Fuzzy Front End stage of the innovation process: a systematic literature review," in *2010 IEEE International Conference on Management of Innovation and Technology (ICMIT)*, 2010, pp. 426 –431.
- S. Noponen, H. Sihvonen, J. Salonen, and J. Kurki, "Systematic literature review of virtual role," in *Internet Technology and Secured Transactions (ICITST), 2011 International Conference for*, 2011, pp. 738 –743.
- L. Major, T. Kyriacou, and O. P. Brereton, "Systematic literature review: Teaching novices programming using robots," in *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, 2011, pp. 21 –30.
- O. Dieste and N. Juristo, "Systematic review and aggregation of empirical studies on elicitation techniques," *IEEE Transactions on Software Engineering*, vol. 37, no. 2, pp. 283 –304, Apr. 2011.
- S. Kollanus, "Test-Driven Development - Still a Promising Approach?," in *Quality of Information and Communications Technology (QUATIC), 2010 Seventh International Conference on the*, 2010, pp. 403 –408.
- D. Liu, Q. Wang, and J. Xiao, "The role of software process simulation modeling in software risk management: A systematic review," in *3rd International Symposium on Empirical Software Engineering and Measurement, 2009. ESEM 2009*, 2009, pp. 302 –311.
- K.-J. Stol, M. A. Babar, B. Russo, and B. Fitzgerald, "The use of empirical methods in Open Source Software research: Facts, trends and future directions," in *ICSE Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development, 2009. FLOSS '09*, 2009, pp. 19 –24.
- M. Kalinowski, G. H. Travassos, and D. N. Card, "Towards a Defect Prevention Based Process Improvement Approach," in *Software Engineering and Advanced Applications, 2008. SEAA '08. 34th Euromicro Conference*, 2008, pp. 199 –206.
- T. Silva da Silva, A. Martin, F. Maurer, and M. Silveira, "User-Centered Design and Agile Methods: A Systematic Review," in *Agile Conference (AGILE), 2011*, 2011, pp. 77 –86.
- E. Hossain, M. A. Babar, and H. Paik, "Using Scrum in Global Software Development: A Systematic Literature Review," in *Fourth IEEE International Conference on Global Software Engineering, 2009. ICGSE 2009*, 2009, pp. 175 –184.
- [P111] M. T. Sletholt, J. E. Hannay, D. Pfahl, and H. P. Langtangen, "What Do We Know about Scientific Software Development's Agile Practices?," *Computing in Science Engineering*, vol. 14, no. 2, pp. 24 –37, Apr. 2012.
- [P112] A. Maglyas, U. Nikula, and K. Smolander, "What do we know about software product management? - a systematic mapping study," in *2011 Fifth International Workshop on Software Product Management (IWSPM)*, 2011, pp. 26 –35.

• M. Jorgensen and M. Shepperd, “A Systematic Review of Software Development Cost Estimation Studies,” IEEE Transactions on Software Engineering, vol. 33, no. 1, pp. 33 –53, Jan. 2007

8.6 Primary studies from Springer Link

• M. Razavian and P. Lago, “A Frame of Reference for SOA Migration,” in *Towards a Service-Based Internet*, vol. 6481, E. Di Nitto and R. Yahyapour, Eds. Springer Berlin / Heidelberg, 2010, pp. 150–162.
• D. Wahyudin, R. Ramler, and S. Biffl, “A Framework for Defect Prediction in Specific Software Project Contexts,” in *Software Engineering Techniques*, vol. 4980, Z. Huzar, R. Koci, B. Meyer, B. Walter, and J. Zendulka, Eds. Springer Berlin / Heidelberg, 2011, pp. 261–274.
• A. Teka, N. Condori-Fernandez, and B. Sapkota, “A Systematic Literature Review on Service Description Methods,” in *Requirements Engineering: Foundation for Software Quality*, vol. 7195, B. Regnell and D. Damian, Eds. Springer Berlin / Heidelberg, 2012, pp. 239–255.
• G. Loniewski, E. Insfran, and S. Abrahão, “A Systematic Review of the Use of Requirements Engineering Techniques in Model-Driven Development,” in *Model Driven Engineering Languages and Systems*, vol. 6395, D. Petriu, N. Rouquette, and Ø. Haugen, Eds. Springer Berlin / Heidelberg, 2010, pp. 213–227.
• T. Yue, L. Briand, and Y. Labiche, “A systematic review of transformation approaches between user requirements and analysis models,” *Requirements Engineering*, vol. 16, no. 2, pp. 75–99, 2011.
• Y. Khan, M. Elish, and M. El-Attar, “A Systematic Review on the Impact of CK Metrics on the Functional Correctness of Object-Oriented Classes,” in *Computational Science and Its Applications – ICCSA 2012*, vol. 7336, B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. Rocha, D. Taniar, and B. Apduhan, Eds. Springer Berlin / Heidelberg, 2012, pp. 258–273.
• I. Steinmacher, A. Chaves, and M. Gerosa, “Awareness Support in Global Software Development: A Systematic Review Based on the 3C Collaboration Model,” in *Collaboration and Technology*, vol. 6257, G. Kolfschoten, T. Herrmann, and S. Lukosch, Eds. Springer Berlin / Heidelberg, 2010, pp. 185–201.
• R. Vivian, E. Huzita, G. Lapasini Leal, and A. Chaves Steinmacher, “Context-Awareness on Software Artifacts in Distributed Software Development: A Systematic Review,” in *Collaboration and Technology*, vol. 6969, A. Vivacqua, C. Gutwin, and M. Borges, Eds. Springer Berlin / Heidelberg, 2011, pp. 30–44.
• S. Kollanus, “Critical Issues on Test-Driven Development,” in *Product-Focused Software Process Improvement*, vol. 6759, D. Caivano, M. Oivo, M. Baldassarre, and G. Visaggio, Eds. Springer Berlin / Heidelberg, 2011, pp. 322–336.
• D. Blanes, E. Insfran, and S. Abrahão, “Requirements Engineering in the Development of Multi-Agent Systems: A Systematic Review,” in *Intelligent Data Engineering and Automated Learning - IDEAL 2009*, vol. 5788, E. Corchado and H. Yin, Eds. Springer Berlin / Heidelberg, 2009, pp. 510–517.
• Q. Gu and P. Lago, “Service Identification Methods: A Systematic Literature Review,” in *Towards a Service-Based Internet*, vol. 6481, E. Di Nitto and R. Yahyapour, Eds. Springer Berlin / Heidelberg, 2010, pp. 37–50.
• M. Ivarsson and T. Gorschek, “Technology transfer decision support in requirements engineering research: a systematic review of REj,” *Requirements Engineering*, vol. 14, no. 3, pp. 155–175, 2009.
• U. Hyrkkänen and S. Nenonen, “The Virtual Workplace of a Mobile Employee – How Does Vischer’s Model Function in Identifying Physical, Functional and Psychosocial Fit?,” in *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments*, vol. 6763, J. Jacko, Ed. Springer Berlin / Heidelberg, 2011, pp. 69–75.

• C. Moraga, M. Moraga, C. Calero, and Á. Caro, “Towards the Discovery of Data Quality Attributes for Web Portals,” in *Web Engineering*, vol. 5648, M. Gaedke, M. Grossniklaus, and O. Díaz, Eds. Springer Berlin / Heidelberg, 2009, pp. 251–259.

8.7 Primary studies from ACM

- S. De S. Reis and R. O. Prates, “Applicability of the semiotic inspection method: a systematic literature review,” in *Proceedings of the 10th Brazilian Symposium on on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction*, Porto Alegre, Brazil, Brazil, 2011, pp. 177–186.
- A. Arcuri and L. Briand, “A practical guide for using statistical tests to assess randomized algorithms in software engineering,” in *Proceedings of the 33rd International Conference on Software Engineering*, New York, NY, USA, 2011, pp. 1–10.
- R. Rasmussen, “Electronic whiteboards in emergency medicine: a systematic review,” in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, New York, NY, USA, 2012, pp. 483–492.
- Engström, M. Skoglund, and P. Runeson, “Empirical evaluations of regression test selection techniques: a systematic review,” in *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*, New York, NY, USA, 2008, pp. 22–31.
- R. C. de Boer and R. Farenhorst, “In search of `architectural knowledge’,” in *Proceedings of the 3rd international workshop on Sharing and reusing architectural knowledge*, New York, NY, USA, 2008, pp. 71–78.
- S. R. S. Souza, M. A. S. Brito, R. A. Silva, P. S. L. Souza, and E. Zaluska, “Research in concurrent software testing: a systematic review,” in *Proceedings of the Workshop on Parallel and Distributed Systems: Testing, Analysis, and Debugging*, New York, NY, USA, 2011, pp. 1–5.
- D. S. Cruzes and T. Dyb\aa, “Synthesizing evidence in software engineering research,” in *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, New York, NY, USA, 2010, pp. 1:1–1:10.
- L. Chen, M. Ali Babar, and N. Ali, “Variability management in software product lines: a systematic review,” in *Proceedings of the 13th International Software Product Line Conference*, Pittsburgh, PA, USA, 2009, pp. 81–90.