Available online at www.sciencedirect.com
ScienceDirect
Advanced Engineering Informatics 21 (2007) 133–151
ADVANCED ENGINEERING INFORMATICS
www.elsevier.com/locate/aei

Scientific research ontology to support systematic review in software engineering

Jorge Calmon de Almeida Biolchini, Paula Gomes Mian, Ana Candida Cruz Natali, Tayana Uchôa Conte, Guilherme Horta Travassos *

COPPE/UFRJ – Systems Engineering and Computer Science Program, Cx. Postal 68.511, CEP 21945-970 Rio de Janeiro, Brazil

Received 6 November 2006; accepted 12 November 2006

Abstract

The term systematic review is used to refer to a specific methodology of research, developed in order to gather and evaluate the available evidence pertaining to a focused topic. It represents a secondary study that depends on primary study results to be accomplished. Several primary studies have been conducted in the field of Software Engineering in the last years, determining an increasing improvement in methodology. However, in most cases software is built with technologies and processes for which developers have insufficient evidence to confirm their suitability, limits, qualities, costs, and inherent risks. Conducting systematic reviews in Software Engineering consists in a major methodological tool to scientifically improve the validity of assertions that can be made in the field and, as a consequence, the reliability degree of the methods that are employed for developing software technologies and supporting software processes. This paper aims at discussing the significance of experimental studies, particularly systematic reviews, and their use in supporting software processes. A template designed to support systematic reviews in Software Engineering is presented, and the development of ontologies to describe knowledge regarding such experimental studies is also introduced.

Keywords: Experimental software engineering; Systematic review; Experimental study; Ontology; Scientific method

1. Introduction

Experimentation in Software Engineering has been growing along the last years. Sjøberg et al. [1] describe the increase of investigation based on experimentation in this area of research. The increasing importance of research activity in the Software Engineering (SE) field, addressed to produce knowledge that can be based on sound scientific methodology, has become one of the major challenges to strengthen the foundations of SE as a discipline towards its full maturity [2]. This desideratum is not only related to the academic world, as far as the industrial one has been getting benefits from the scientific method to validate its software technologies [3] and to improve software processes [4]. In contrast to the authority-based approach, which relies on the value ascribed to the enunciator of a certain assertion, and also contrasting with the rationalist approach in the scientific discourse, which derives conclusions from aprioristic assumptions, experimental research is based on the systematic process of factual apprehension and organization of the universe under scrutiny. In the SE field it aims at developing an evidence basis for the scientific understanding and intervention on the processes involved in the development of software technologies, in order to support the engineering methods and technology employed in their planning, design, construction, implementation, validation, maintenance and software processes improvement.

The experimental paradigm in science is based on the systematic observation and experimentation over the elements that interact in the universe at issue [5]. As any other

* Corresponding author. Tel.: + 55 21 25628712.
E-mail addresses: jorgebio@cos.ufrj.br (J.C. de Almeida Biolchini), pgmian@cos.ufrj.br (P.G. Mian), anatali@cos.ufrj.br (A.C.C. Natali), tayana@cos.ufrj.br (T.U. Conte), ght@cos.ufrj.br (G.H. Travassos).

1474-0346/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aei.2006.11.006

research field based on experimentation, Experimental Software Engineering comprises two kinds of investigation: Primary and Secondary studies.

Primary studies are specific designs of methodological processes, addressed to directly evaluate the hypothesis formulated by the researcher, by testing it under well-established conditions of observational or experimental methodological control. Different types of experimental studies can be used in the Software Engineering context [6]. Some proposals to support the accomplishing of these studies can be found in the technical literature [5,7]. Researchers have applied primary studies to build software engineering knowledge [8] and to support the processes regarding the engineering of software technologies, mainly those related to technology evaluation [9]. Nowadays, experimentation represents an important approach to support CMMI's maturity levels 4 and 5 activities [4].

Secondary studies are those intended to produce systematic comparisons or assembling between individual investigations, scientifically selected from within a set of primary studies, as well as to allow the generation of generalizations from them. From the methodological viewpoint, this type of study is designated as systematic review [10], also known as Research Synthesis [14]. Systematic reviews can support the building of an evidence based SE body of knowledge [61]. Besides, this type of study also represents an important tool for the Software Engineer when developing new software technologies. Systematic reviews (SR) have been used to support the initial phases of the engineering of software technologies. Mafra et al. [11] have extended the methodology proposed by Shull et al. [9] for introducing software processes into the industry by including SR as one of the methodological steps. This extended process has been used to concretely develop a checklist-based technique [12] to inspect software architecture models and a reading technique [13] to inspect software requirements documents exploring the object-oriented designer perspective.

Systematic review is primarily an information-based process throughout which each methodological step must be carefully and systematically designed and controlled (by using a formal research protocol), in order to guarantee the necessary consistency and robustness of the obtained results and conclusions [14].

Conducting systematic reviews in Software Engineering consists in a major methodological tool to scientifically improve the validity of the assertions that can be made in the field and, as a consequence, the reliability degree of the methods that are employed for developing software technologies.

Despite its importance, some issues on using systematic review methodology, extracted from our first SR accomplishments [15], showed us that conducting systematic reviews is not a simple task. A systematic review process uses specific concepts and terms that may be unknown to researchers used to conduct ad hoc literature reviews. Besides, systematic reviews require an additional conduction effort. The review must be planned before execution and the whole process must be documented, including intermediary results. These issues pointed us to the need of applying research efforts to develop SR planning and execution methodologies, in order to guide researchers to perform the systematic review process in the Software Engineering field.

However, the systematic review process core part is represented by the precise and explicit formulation of the research question, which must be addressed to the bulk of information to be derived from primary studies. In order to efficiently extract data that will constitute the analytic basis for generating conclusions through the SR, the aforementioned research question must be well structured and its elements explicitly defined.

The core research question of the systematic review is a multidimensional information structure that must be related to all the subsequent stages of the SR conduction process. Consistency among different researchers for defining this conceptual structure during the systematic review planning phase, and correspondingly guaranteeing the obtainment of similar and comparable results, could be improved by having the support of a formalized common terminology of the involved concepts, represented by an ontology. Although ontologies could bring some limitations such as being hard to evolve, not being able to completely satisfy every single group or individual [16], and the present lack of ontology representation language translators [17], we believe that SR research could benefit from using them, due to their applicability in conceptual formalization and knowledge representation.

Ontologies can be used to enable multiple target applications or humans to have access to heterogeneous sources of information that are expressed by using diverse vocabularies or inaccessible formats. An ontology of a given domain can provide a vocabulary for specifying requirements for one or more target applications. In fact, ontology is used as a basis for software specification and development, allowing knowledge reuse. Also, ontologies are applied to searching an information repository for desired resources, improving precision and reducing the overall amount of time spent in searching [18]. Examples of using ontologies to support different software engineering processes are, among others, the works of Kitchenham et al. [19], Falbo et al. [20], and Villela et al. [21].

The development of the scientific research ontology can provide the professional, who is carrying out a systematic review, hierarchies of conceptual classes, which correspond to the set of component categories pertaining to the research question information structure. These hierarchies can represent a useful support for better defining the conceptual elements and relations that are needed in order to compose the research question, methodologically formulated at the planning phase of the SR process [22].

A bottleneck that can be identified regarding the conduction of systematic reviews in Software Engineering is the need for standardization of the concepts and terms of the experimentation and SE fields, and also their conceptual relationships. However, it could also be a source of potential motivation for enhancing experimental investigation in the field. This issue points to the need of developing a scientific research ontology, which can function as a useful terminological model for improving the rigor of investigation in SE, particularly related to the systematic review research.

Terminological models, such as taxonomies, can be helpful tools as well in order to better define the terminological elements that might be included in the queries, during the execution phase of the SR process, in order to achieve the best possible performance of the correspondent information retrieval activity.

This paper aims at discussing the significance of systematic reviews, from their basic concepts and historical origins to their application in the Software Engineering domain, and at stressing the importance and utility of supporting these processes through ontologies.

As a formal, explicit specification of a shared conceptualization [23], an ontology can be a useful tool to support the conduction of systematic reviews in Software Engineering. Its utility ranges from providing semantic and terminological support to the researcher, during the planning and execution phases of a systematic review in the SE field, up to supplying a formal conceptual framework to help examining the results of the studies under analysis.

It can also be an important means to contribute for improving the conceptual structuring of the Software Engineering field. As a consequence, this improvement can also become an instrument to help producing more solid and consistent experimental studies in Software Engineering.

This text has been written based on our own experience in conducting systematic reviews in different topics within the Software Engineering field. Our main interest in dealing with systematic reviews is regarding the assumption that systematic reviews can support software processes by formalizing some important decision making, relative to evidencing knowledge and supporting software processes applied to the development of new software technologies. It primordially intends to share the results of these experiences and faced issues, including a Systematic Review Conduction Process and a Systematic Review Protocol Template, aiming to facilitate the accomplishment of such kind of study [22]. The systematic review process is presented in this paper, as well as some relevant parts of the Protocol Template as they concern to systematic reviews' conceptual and terminological issues. Besides the description of the SR processes and the template to support systematic reviews in Software Engineering, an ontology to describe knowledge regarding such experimental studies is also introduced.

2. Systematic reviews

In contrast to the ad hoc process of literature review, unsystematically conducted whenever one starts a particular investigation, a systematic review is developed, as the term denotes, in a formal and systematic way. This means that the research conduction process of a systematic type of review follows a very well defined and strict sequence of methodological steps, according to an aprioristically developed protocol. This instrument is constructed around a central issue, which represents the core of the investigation, and which is expressed by using specific concepts and terms, that must be addressed towards information related to a specific, pre-defined, focused, and structured research question. The methodological steps, the strategies to retrieve the evidence, and the focus of the question are explicitly defined, so that other professionals can reproduce the same protocol and also be able to judge about the adequacy of the chosen standards for the case.

Synonyms of this methodology that are to be found in the literature include the following terms: overview, research review, research synthesis, research integration, systematic overview, systematic research synthesis, integrative research review, and integrative review.

The type of acceptable evidence to be gathered in a systematic review is stated beforehand. The retrieved evidence is thoroughly reviewed, comparable to other types of evidence previously and elsewhere retrieved.

The evidence data are normalized in such a way as to make results from different studies comparable, in terms of their magnitude of effect, even when they are presented in diverse ways but related to compatible concepts. It is then possible, e.g., to compare studies which evidence is expressed by absolute risk reduction with others where it is expressed by relative risk.

Besides comparing results of individual studies, different kinds of syntheses can be done. The election mode allows the researcher to look for each study separately and counting them as "votes" about the question focus. For instance, in a specific SR conducted in the medicine field, the researcher could find that, among 35 valid studies, 29 showed a positive result, while 5 showed no result, and one study showed a negative result. Internal comparison of studies, based on their specific parameters, can show contrasts and other kinds of differences that may elucidate distinct aspects of the question. In the same example, one could find that the negative effect must be due to a different dosage scheme, while the five studies that showed no result were conducted in subjects that had a different age distribution in comparison to the 29 positive ones.

Another type of research synthesis is known as meta-analysis, where the original individual studies are treated as if they were parts of one larger study, by having their data pooled together in one single and final result that summarizes the whole evidence. By selecting studies that are compatible in their quality level, and by taking strict care with their specific details, this methodological procedure can produce evidence as well as reveal aspects that the original studies are not individually able to elucidate. For instance, meta-analysis may prove that the results are statistically significant when small studies give inconclusive results with large confidence intervals. Besides that, when conflicting results arise from different individual studies, meta-analysis may reconcile the data in a synthetic result, while each individual study can then be weighted and compared with it, so that other kinds of conclusions might be derived from these discrepancies.

Ideally, a meta-analysis should be performed as part of a systematic review, usually as its final step. All meta-analyses should ideally start with an unbiased systematic review, which incorporates articles that are chosen by using predetermined inclusion criteria [24,25]. If the data extracted from these studies meet certain requirements (the most important being a high level of homogeneity of effect measures across studies), then the data can be combined using meta-analysis. However, if the effect measures are found to be heterogeneous, then it is still acceptable to present the work as a systematic review and not perform meta-analysis, or use statistical methods that can account for the heterogeneity. Indeed, there are situations when meta-analysis is clearly inappropriate [26].

2.1. Systematic and unsystematic reviews: differences, advantages, and disadvantages

A literature review is usually an initial step in any research and development enterprise. From the scientific methodology viewpoint, it is in fact a recommended and necessary step for the professional to endeavor whenever starting a research project. Since science is a cooperative social activity and scientific knowledge is the result of a cumulative process of this cooperation, the literature review is the means by which the researcher can perform a mapping of the existing and previously developed knowledge and initiatives in the field. The review can provide material to be used by the researcher in the work that is being designed, and locate it in relation to the different regions of the field and approaches to the issue in focus. It also permits both an analysis of the previous findings, techniques, ideas and ways to explore the topics in question, as well as about their relevance in relation to the issues of interest, and a synthesis and summarization of this information. It can help planning the new research, avoiding unnecessary duplication of effort and error, and orient the investigation process. Due to the growth of scientific production, the role of literature reviews has been proportionally increasing, and "their importance grows as a direct function of the number of documents on a topic" [14]. Due to its important role in the scientific enterprise, general rules for performing literature overviews have been developed, in order to warrant the investigator good quality of information from the covered material.

The systematic review consists in a specific scientific methodology that goes one step further than the simple overview. It aims at integrating empirical research in order to create generalizations. This integrative enterprise involves specific objectives, which allows the researcher to critically analyze the collected data, to resolve conflicts detected in the literature material, and to identify issues for planning future investigation. Due to these particular aims, the systematic review is not considered to be a phase of a research enterprise, a role that is performed by the usual literature review. As a matter of fact, the integrative review is a different methodological procedure of research in its own, comprising distinct investigation aims as well as specific methodological features, requirements, and procedures. From the epistemological perspective, it represents a different approach to the relevant issues in a research area that opens up a new field of possibilities for generating new types of knowledge in a scientific domain.

In practice, the distinction between ordinary review articles and systematic review ones can be done by comparing their underlying semantic structures, as evidenced by the types of content to be found in their respective abstracts as well as in the titles of the respective article sections. In the medical field, for instance, a simple overview article refers in its abstract to key points about the subject, without discussing or emphasizing the methodology of the review itself. The article sections include titles that refer to topics that are very similar to the sections that are usually found in a textbook chapter, such as the natural history of the disease referred to its different phases of evolution and expression, the characteristics of the symptoms and signs and the differential diagnosis with other diseases, causal mechanisms or hypotheses of the disease, the goals of the treatment, the types of drugs that might be used or recommended, other types of intervention, and so on [27]. In contrast, a systematic review article abstract contains a specific pattern of sections, such as background, purposes, data sources, study selection, data extraction, data synthesis, discussion and conclusion. The article sections expand this same abstract section structure, including in its titles terms such as 'methods', 'data synthesis', 'efficacy', 'discussion', as well as describing and discussing the methodology of the research review itself. It also presents considerations about the specific requirements that were aprioristically defined and made explicit in order to include (or exclude) the primary studies in (from) the review material. It also includes tables containing quantitative information, such as the data extracted from the individual studies, weighted results of each study to account for the relative size of the study, a row entitled 'total', and sometimes individual study numbers reassessed as a new, aggregated pool of patients.

Despite the importance of the literature reviews, even when they are conducted according to their corresponding 'good practice' rules, they suffer from lack of scientific rigor in performing its different steps. The unsystematic conduction of this type of review might introduce, as it usually does, some research biases in different stages of the review process, ranging from question formulation, through data collection, data evaluation, analysis, interpretation, summarization, and presentation. The development of a systematic approach of research review aims to establish a more formal and controlled process of conducting this type of investigation, avoiding the introduction of the biases of the unsystematic review. In addition to this central aspect, the systematic review does not consist on a simple rearrangement of the already known or published data. It is at the same time a new type of methodological approach for doing research, with an integrative purpose. Therefore it emphasizes the discovery of general principles, in a higher level of conceptual abstraction in the research field, it stresses the diagnosis and analysis of the relative external inconsistencies when comparing individual studies with contrasting results between themselves, as well as it helps to illuminate new aspects and issues in the field and guide future research lines and possibilities. For the classical approach of literature review, variation among studies tends to represent a source of noise, a disturbing factor for interpretation and judgment. For the systematic review methodology, on the contrary, variety is a stimulating factor for understanding the whole scenario of the particular issue that is under investigation, allowing the researcher to moderate the relative influences of the different individual studies, by viewing them as probabilistically distinct possibilities of result.

Like any other scientific methodology, the integrative and systematic review presents its potentials and also its limitations [28,29]. When compared to primary research, the unique contributions of research synthesis include the improvements in precision of the data and the reliability of the information, as well as the three aforementioned ones: testing hypotheses that possibly have never been, or could never be tested in primary studies; using consistent and explicit rules for an evidence-based process of moderating influences of primary studies; and, in a recursive way in relation to cumulative scientific knowledge, addressing questions about the research enterprise itself, such as trends of issues, concepts, methods, or results over time, as well as questions about the field research contexts in a broader sphere.

The main limitations of research integration are related to the nature of review-generated evidence and of post hoc hypothesis testing. The first one can happen when the researcher compares the results of primary studies that used different procedures to test the same hypothesis. Because the antecedent variable is not aprioristically controlled, such as randomly assigning the issue of interest to the different types of study procedure, confounding factors are liable to exist and consequently interfere in the integrated results. The consequence is that, in this situation, causal inferences are not possible to be done with the same degree of confidence. At the same time, it can provide the researcher with some good suggestions related to the future orientation of new primary studies.

The second limitation can derive from the fact that, in many cases, the researcher has in advance a reasonable knowledge about the empirical evidence related to the issue of interest. If the hypothesis stated for a specific systematic review is derived from the same data that will be integrated through this methodology, the researcher cannot use this same evidence to test the hypothesis so generated. In order to avoid a vicious circularity of the evidence base, the data used to generate the review hypothesis, and the ones used to test it, must be independent between themselves.

These limitations reinforce the idea that primary studies and systematic review are complementary processes of knowledge production. The second methodological approach cannot be considered to be a substitute for primary evidence production, in a competing way. On the contrary, the enhancement of precision and reliability provided by the systematic review process helps to improve and to better direct future primary research, through a positive feedback relationship between them.

2.2. The origins of systematic review

Early works to integrate results can be traced back from the beginning of the 20th century [14]. Pearson, in 1904, calculates the average of results of the correlation between inoculation for typhoid fever and mortality in order to better estimate this type of effect and to compare it with that of the inoculation for other kinds of disease.

In the 1930s, methods for combining estimates are developed in other fields of research, such as the physical sciences [30] and in the statistical sciences [31–34], and later applied to other fields, such as agriculture, with few methodological unfolding in the following decades. The use of synthesis techniques gains momentum in the 1970s, when new methodological proposals for integrating research results are developed as well as several applications are mainly developed in the social sciences.

In the methodology sphere, Feldman [35] describes steps in the literature reviewing process; Light and Smith [36] develop a methodological treatment of the variations in the outcomes of studies; Taveggia [37] describes common problems in literature reviews and proposes to treat "contradictory findings" of individual researches in a particular topic as "positive and negative" details of a probabilistic "distribution of findings" rather than "inconsistencies"; Glass [38] defines the term meta-analysis as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings".

In the application field, the main studies include the fields of clinical psychology [39], industrial/organizational psychology [40], social psychology [41], education [42], and cardiology [43].

The field of research synthesis spreads from the social sciences to medicine in the 1980s, while books devoted to this methodology are published [44–47], which are followed by others in the same decade. At this time, the research synthesis methodological approach becomes a new and independent specialty and achieves legitimacy as a field of research. The research review and the meta-analytic approaches are integrated, new methods and techniques are developed, and a more rigorous level of methodology is achieved. Research synthesis also spreads to the social policy domain, to help the decision-making process, where Light and Pillemer [48] emphasize the relevance of uniting numeric data and narrative information for the effectiveness of result interpretation and communication.

Since the late 1980s, systematic research synthesis and meta-analysis reach an especially distinctive methodological status in the health sciences domain [49]. From then on, health policy agencies and legislation have fostered and relied on this methodology as a fundamental requirement to develop, publish, and recommend guidelines on clinical practice in the various medical specialties and application areas.

2.3. Examples of systematic reviews in medicine

In this section, we present a sample of some titles of systematic review studies, developed in the medical area, that illustrate the specific nature of the results that are obtained by using this type of methodology.

The type of evidence that can be derived from primary studies differs according to the research design that is used for conducting them. The degree of confidence that is possible to obtain from an investigation experiment directly depends on the degree of experimental control that the researcher can exert over the object that is under study [50].

In the medical field, with the purpose of ascribing different values for the quality and scientific reliability of studies, a scale of distinct types of study design has been developed, in order to serve as a reference of the different confidence degrees of evidence that can be produced. Control measures of the experiment, such as blindness of either or both the patients and the clinicians that directly assess them, as well as comparison of subgroups inside the study, and also random assignment of the patients to the different subgroups, contribute to increase the reliability degree in the evidence that can be produced.

Therefore, the following evidence-grading reference system has been developed in the health field in order to help professionals to judge the quality of articles reporting scientific studies (Table 1).

Table 1
Level and source of evidence [51]

Level of evidence	Source of evidence
1a	Systematic review of randomized controlled trials
1b	Individual randomized controlled trial
1c	"All or none" case series
2a	Systematic review of retrospective cohort studies
2b	Individual retrospective cohort study; or low quality individual randomized controlled trial
2c	"Outcomes" research
3a	Systematic review of case–control studies
3b	Individual case–control study
4	Case series; or low quality cohort studies; or low quality case–control studies
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"

The analysis of the syntactic structure of the study titles listed below shows that they contain two basic information units, one referring to the technology that is being studied, and the second one standing for the target problem. Other information units might sometimes be present, such as the target population, the obtained result, or the primary study methodology.

This type of knowledge corresponds to what is known as foreground knowledge, specifically addressed for providing support to decision making processes, in contrast to the background knowledge type, which is acquired by the professional during one's training and educational process.

The italicized words and the curly brackets' remarks in the study titles below were introduced by the authors, in order to evidence and make it explicit the difference between the linguistic parts of the titles' text. The remarks refer to the aforementioned basic information units.

- Carotid endarterectomy {=Technology} *for* symptomatic carotid stenosis {=Problem} [52].
- Medium-dose aspirin or other antiplatelet drugs {=Technology} *for* patients at high risk {=Population} *of suffering some* occlusive vascular disease {=Problem} *over the next few months or years* [53].
- *Meta-analysis of* exercise testing {=Technology} *to detect* coronary artery disease {=Problem} *in* women {=Population} [54].
- Streptokinase or other "clot-busting" drugs {=Technology} *as* emergency treatment for patients {=Population} *who are suffering an* acute heart attack {=Problem} [55].
- Chronic hepatitis B virus infection: {=Problem} treatment strategies {=Technology} *for the next millennium* [56].
- Lack of significant benefit {=Result} *of* magnesium infusion {=Technology} *in* suspected acute myocardial infarction {=Problem} [57].
- *A systematic review of* randomized controlled trials {=Primary Study Methodology} *of* pharmacological therapy {=Technology} *on* osteoarthritis of the knee {=Problem}, *with an emphasis on trial methodology* [58].
- Hormonal adjuvant treatments {=Technology} *for* early breast cancer {=Problem} [59].
- *A systematic review of* newer pharmacotherapies {=Technology} *for* depression {=Problem} *in* adults: {=Population} *evidence report summary* [60].

3. Systematic reviews in Software Engineering

Several primary studies have been conducted in the field of software engineering in the last years, accompanied by an increasing improvement in methodology. However, in most cases software is built with technologies for which developers have insufficient evidence to confirm their suitability, limits, qualities, costs, and inherent risks. It is difficult to be sure that changing software practices or processes will necessarily be a change for the better. It is possible that research syntheses can provide the mechanisms needed to assist practitioners to adopt appropriate technologies and to avoid inappropriate technologies. Thus, the development of research syntheses in this field is still an area of investigation that remains to be explored and that could well bring many benefits.

In this context, there are few initiatives that question how Software Engineering would benefit from adopting the evidence approach. Kitchenham et al. [61] discuss the possibility of evidence-based Software Engineering by using an analogy with medical practice. Nevertheless in order to obtain evidence that can be generalized, it is necessary to perform systematic reviews. So, Kitchenham [61] evolves the idea of Evidence-Based Software Engineering and proposes a guideline for systematic reviews that is appropriate for software engineering researchers. The guideline has been adapted to reflect the specific problems of Software Engineering research and covers three phases of a systematic review process: planning the review, conducting the review and reporting the review. Although it describes a template for SR in Software Engineering, at a relatively high level, it does not consider the impact of question type on the review procedures, nor does it specify in detail the mechanisms that are needed to undertake meta-analysis.

Like all knowledge areas that have previously employed this research methodology, developing this investigation approach in the software engineering field implies in adapting the conceptual and methodological dimensions of research synthesis to the domain, taking into account its specificities as a scientific knowledge area [62].

Differently to the medical area, Software Engineering has some specificity that would make it difficult for the research synthesis to obtain evidence.

One major difference between medicine and software engineering is that most software engineering methods and techniques must be performed by skilled software practitioners who are aware of the methods and techniques that are being applied. In contrast, although medical practitioners are skilled individuals, the treatments they prescribe (e.g. medicines and other therapeutic remedies) do not necessarily require awareness of their effective presence in order to be skillfully administered by the professional or received by the patient. The reason why professional skill presents a problem in conducting a controlled experiment in the software engineering field is due to the fact that it prevents adequate blinding of practitioners during the study. In medical experiments (particularly drug-based experiments), the gold standard experiment is a double-blind randomized controlled trial (RCT). In a double-blind experimental trial neither the doctor nor the patient knows which treatment the patient is being administered. The reason why double-blinded trials are required is to prevent patient and doctors expectations biasing the results. Such experimental protocols are impossible to be conducted in software engineering experiments, which rely on a subject performing a human-intensive task.

Another difference between software engineering and medicine is that most SE techniques impact part of the lifecycle, in such a way that it makes the individual effect of a technique difficult to be isolated. The target techniques interact with many other development techniques and procedures. In general, it is difficult to determine a linear causal link between a particular technique and a desired project outcome, when the application of the technique and the final outcome are temporally removed from one another, while at the same time there are many other tasks and activities involved in the study that could also affect the final outcome.

And also, differently from software engineering, medical researchers and practitioners look for already published systematic reviews, i.e., papers that have already assembled all relevant reports of a particular topic. Medical researchers have a large amount of technological and scientific infrastructure to support them in this enterprise. There are several organizations (in particular, the international Cochrane Collaboration – www.cochrane.com) that assemble systematic reviews of studies about drugs and medical procedures. To provide a central information source for evidence, the Cochrane Collaboration publishes systematic reviews in successive issues of The Cochrane Database of Systematic Reviews. These reviews are continually revised, both as new experimental results become available and as a result of valid criticisms of the reports.

There is no equivalent to the Cochrane Collaboration in the Software Engineering area. Instead of it, there are many abstracting services that provide access to software engineering articles. Currently, available evidence related to software engineering technologies accordingly Kitchenham et al. is [61]:

- *Fragmented and limited*. Many individual research groups undertake empirical studies. However, because the goals of such works are either to produce individual publications and/or to generate post-graduate theses, it may be little harder to find an overall purpose to such studies. Due to the lack of a research culture that strongly advocates systematic reviews and replication, it is easier for researchers to undertake research in their own areas of interest rather than contribute to a wider research agenda.
- *Not properly integrated*. Currently, there are no agreed standards for systematic reviews. Thus, although most software engineering researchers undertake reviews of the "State of the Art" in their topic of interest, the quality of such reviews is variable, and as a rule they do not lead to published papers. Furthermore, there have been few attempts to apply meta-analytic techniques (meta analysis) to Software Engineering because of the limited number of studies replications.
- *Without agreed standards*. There are no generally accepted guidelines or standard protocols for conducting individual experiments. Kitchenham et al. [61] proposed some preliminary guidelines for formal

experiments and surveys. However, they do not address observational, as well as investigative studies. Furthermore, because they attempt to address several different types of experimental studies, the guidelines are neither as specific nor as detailed as they are found in the medical area.

4. Experiences in applying systematic reviews in Software Engineering

Arguing that research in Software Engineering may benefit from the application of the systematic review methodology; we have conducted some reviews addressing different research topics in the context of SE master and doctoral degree thesis production. Our first initiatives of conducting systematic reviews used the work of Kitchenham [10] and the protocol example by Mendes and Kitchenham [64] as reference material. This set of works proposes a guideline for conducting systematic reviews in SE, derived from existing guidelines used by medical researchers, and adapted to reflect the specific problems of Software Engineering research, besides a concrete example for a SR protocol applied to the SE field.

For instance, we have conducted systematic reviews on web-based systems development processes, reading techniques for quality inspections, software process evaluation models and ubiquitous computing. Among these, the work of [63] aimed at characterizing software processes that are used to develop web applications. The objective was to identify the current state of the art of this particular research topic. The protocol was created following the same previous guidelines to plan systematic reviews proposed by [10] and using the protocol given at [64] as reference. In order to evaluate the developed protocol, specialists were asked to peer-review it throughout its definition. As a result of these evaluations, 5 different protocol versions were generated. The search was performed by using web search engines, and included journals and conference articles. The procedure used to select the studies was reading the articles' full texts. During the systematic review execution, 108 articles were obtained. Among them, 22 were selected.

After this initial experience, and based on the acquired experiences, additional systematic reviews have been accomplished. Some of them resulted in technical papers published elsewhere. Others are represented by their protocols and results, described as technical reports. These secondary studies were prepared to deal with:

- *Software architecture* [65]: the objective of this systematic review was to identify evaluation approaches for software architecture models. The search was executed in the digital libraries and libraries. First, the abstracts of the obtained studies were read in order to filter the ones that were not considered to be relevant. Then, the "Introduction" section of each remaining article was read. Eighty studies were obtained and 54 were selected. The results from the SR have been used to define a set of requirements for the development of a checklist based inspection technique concerned with software architecture models [12].
- *Software process models evaluation approaches* [66]: this systematic review was conducted to identify existing initiatives for review and verification of such models. The search was executed in selected digital libraries. First, the abstracts of the obtained studies were read in order to filter the ones that were not relevant. Then, the full text of each remaining article was read. One hundred and twenty five studies were obtained and just 1 was selected.
- *Reading techniques for inspecting requirement specifications* [67]: the objective of this review was to identify, analyze and evaluate experimental studies regarding reading techniques for inspecting requirement documents. The search was executed in the selected digital libraries. To select the studies, the abstracts of the obtained articles were read. Two hundred and seventy eight studies were obtained and 38 were selected. As mentioned before, the results of this SR have been also used to support the definition of the features needed to build an object-oriented designer perspective based reading technique [68].
- *Ubiquitous computing* [69]: this systematic review was conducted to characterize Ubiquitous Computing, to establish its state of art and to define ubiquitous applications' characteristics. The search was executed in digital libraries. Initially, the abstracts of all obtained articles were read. The researchers selected a smaller set of articles, the full texts of which were read. Five hundred and ninety one articles were obtained and 38 were selected. The results have been used to suggest a classification schema regarding ubiquitous software projects.
- *Software testing techniques* [70]: the objective of this review was to find initiatives of test planning. The search was executed in digital libraries and libraries. First, the abstracts of the obtained studies were read in order to filter the ones that were not relevant. Then, the full text of each remaining article was read. Fifty six studies were obtained and 8 were selected. This review was not performed as formally as the SRs presented above. It has been planed according to the concepts described in this text, but its execution was simplified due to time restrictions. Even so, the results were strong enough to allow the development of Maraká, an infrastructure to support testing planning and control.

The motivations to conduct these systematic reviews were diverse. Some of them intended to collect indicators of some technique existence or use. Others aimed at identifying gaps in the SE research area, at suggesting new works in the field, or at pointing the current context of a research topic. An additional attempt to accomplish a more sophisticated research protocol, including meta-analysis as one of the objectives, can be represented by a systematic review to evaluate the use of estimation models by software organizations [71].

5. Lessons learned and issues

Although the systematic reviews listed above had different contexts and objectives, the issues found during their execution were similar. The main types of difficulties that we faced were: time spent with the initial learning process, necessary to really understand the systematic review execution process; search machines' research restrictions; the systematic review execution effort; and meta-analysis execution.

In the systematic review learning process, the biggest effort is spent during the planning phase, comprising the research protocol definition. Researchers are frequently used to conduct informal reviews, in which the main part of the time is spent in the search execution itself as well as in the article selection. Due to this previously acquired behavior pattern, it is necessary for researchers to become familiar with the new way of proceeding, which emphasizes the utmost importance of the review planning stage. In the systematic review described above [63], it was necessary to contact specialists on systematic review conduction to fully understand the guidelines proposed by [10,64].

Our experience in developing and reviewing research protocols taught us that having a detailed definition of all the information items could be a major help for constructing a more consistent protocol. Besides this relevant resource, the planning stage can be even better developed if the researcher can access some examples or possible alternatives for all the protocols items [15].

Another issue that impacts the systematic review execution is related to the research restriction that is caused by the limitations of the currently accessible search machines at the digital library sources. Some search machines present important limitations that hinder or even forbid the execution of search strings of a more complex nature. For instance, the search strings defined at [63] could not be executed in two of the selected digital sources, due to the corresponding search machines' restrictions, in terms of the combination features that are found in its 'Advanced Search' resource. In this case, the search strings needed to be revised and divided into new less complex expressions. However, in such cases, we ought to be extremely careful if we need to use phrases as parts of the search strings. In the majority of the search machines, isolated terms are treated in a different way as compared to phrases – fixed combinations of terms. If we are using phrases, we should therefore not break them [15]. A similar problem has been described in [68].

We found other problems in another digital source that restricts the number of the characters that can be used in the search string itself. In this case, we had to use a multiple searching procedure. So, in the work of [63] it was necessary to run part of a search, and subsequently, we executed a new complementary search into its result set. However, all these resource-based procedure variations compromise the study repeatability, since one of the basic premises for conducting a systematic review is the utilization of a uniform set of search string elements in all selected sources [15].

Yet another issue about the limitations of the available search machines is related to the fact that, even in the sources where we were able to run the planned search strings, there happened to be a great number of retrieved articles that were not really relevant for the designed research. This fact indicates that there is a need for evaluating the information retrieval precision of the available search machines.

From these difficulties, we learned that it is strongly recommended to perform a pre-search in the selected sources, in order to evaluate each search machine resource itself and the nature of information retrieval recall and precision it can provide. This pre-search can help us to verify and assess our planned search string set in relationship to the resources that are available at each source, in order to better execute the search stage [15].

Another difficulty found was defining synonyms for the terms that compose the search string. In the work of [63], for instance, the researchers realized that different authors use distinct terms to refer to "web application" (web system, web-based system, web software and/or internet applications). Therefore, it may be necessary to ask specialists in the field to review the protocol and to suggest related terms. However, we believe that it could be simplified with the existence of a common terminology for the area shared by all researchers.

Regarding the effort spent in conducting systematic reviews, it is important to notice that the methodological rigor in applying a systematic review is greater than the one that is necessary in applying an ad hoc one. This implies the need of a more explicit and detailed documentation of all the produced results, in all the different phases of the systematic review process: from planning the research and designing the protocol to the data search, collection, retrieval, and analysis stages [15].

Finally, our last consideration is regarding the meta-analysis. As reported in [72], it was not possible to apply formal meta-analysis in those systematic reviews due to information that has not been reported in the articles used to support data extraction. As suggested by those authors, future studies in Software Engineering should include information to support meta-analysis (such as residuals and actual and estimated measurements) besides to be independent and sharing standard protocols.

6. Systematic review conduction process

Based on our experiences and issues raised when executing systematic reviews, on the existing guidelines [10], and on a protocol example [64], we have described a Systematic Review Conduction Process [22,72].

Systematic review conduction can be understood as a three-step approach. The main phases composing the systematic review process, as shown in Fig. 1, are Review Planning, Review Execution, and Result Analysis.

During the *planning phase*, research objectives are listed and a review protocol is defined. Such protocol specifies the central research question and the methods that will be used to execute the review. The *execution stage* involves primary studies identification, selection and evaluation in accordance with the inclusion and exclusion criteria established in the review protocol. Once studies have been selected, data from the articles are extracted and synthesized during the *result analysis phase*. While each one of these phases is executed, their results must be packaged. Therefore, *systematic review packaging* is performed throughout the whole process. There are two checkpoints in the proposed systematic review process. Before executing the systematic review, it is necessary to guarantee that planning is suitable. The protocol must be evaluated and if problems are found, the researcher must return to the planning stage in order to review the protocol. Similarly, if problems regarding web search engines are found during the execution phase, the systematic review may be re-executed [22].

To support the execution of this process, a review protocol template has been formalized [22,72]. This template was based on the systematic review protocols developed in the medical area, on the guidelines for SR in SE proposed by [10] and on the protocol example found in [64] together with our experiences on SR.

The objective of the template (Fig. 2) is to serve as a guideline to Software Engineering researchers when conducting the systematic review. Therefore, the template lead researchers through each step of the systematic review process, previously presented, by clearly defining the content of each protocol section. A complete description of this template can be found in [22,15]. However, some parts of the Protocol Template, relevant for better understanding the ontology concepts, are described below:

1. **Question formularization:** the research objectives must be clearly defined, filling the items:
   1.1. **Question focus:** defines the systematic review focus of interest, i.e., the review research objectives. Here, the researcher must decide what he/she expects to be answered in the SR end.
   1.2. **Question quality and amplitude:** this section aims at defining the research question syntax (the context in which the review is applied and the question the study must answer) and its semantics specificity (or question range) described by the remaining items of this section:

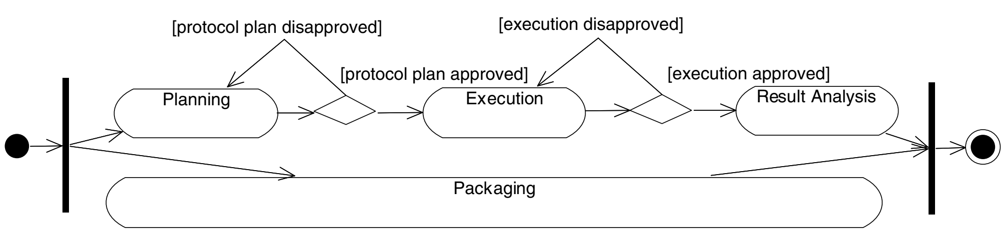

Fig. 1. Systematic review conduction process.

1. Question Formularization
   1.1. Question Focus
   1.2. Question Quality and Amplitude
      - Problem
      - Question.
      - Keywords and Synonyms
      - Intervention
      - Control
      - Effect
      - Outcome Measure
      - Population.
      - Application
      - Experimental Design
2. Sources Selection
   2.1. Sources Selection Criteria Definition
   2.2. Studies Languages
   2.3. Sources Identification
      - Sources Search Methods
      - Search String
      - Sources List
   2.4. Sources Selection after Evaluation
   2.5. References Checking
3. Studies Selection
   3.1. Studies Definition
      - Studies Inclusion and Exclusion Criteria Definition
      - Studies Types Definition
   3.2. Procedures for Studies Selection
   3.3. Selection Execution
      - Initial Studies Selection
      - Studies Quality Evaluation
      - Selection Review
4. Information Extraction
   4.1. Information Inclusion and Exclusion Criteria Definition
   4.2. Data Extraction Forms
   4.3. Extraction Execution
      - Objective Results Extraction
         i) Study Identification
         ii) Study Methodology
         iii) Study Results
         iv) Study Problems
      - Subjective Results Extraction
         i) Information through Authors
         ii) General Impressions and Abstractions
   4.4. Resolution of divergences among reviewers
5. Results Summarization
   5.1. Results Statistical Calculus
   5.2. Results Presentation in Tables
   5.3. Sensitivity Analysis
   5.4. Plotting
   5.5. Final Comments
      - Number of Studies
      - Search, Selection and Extraction Bias
      - Publication Bias
      - Inter-Reviewers Variation.
      - Results Application
      - Recommendations

Fig. 2. Systematic review protocol template.

- **Problem:** defines the systematic review target, describing briefly the research context.
- **Question:** research question to be answered by the systematic review. It is important to highlight that, if the systematic review context is too wide, it may be necessary to decompose the research question in secondary questions to narrow the research target.
- **Keywords and synonyms:** list of the main terms that compose the research question. These terms will be used during the review execution (in case the search by keywords is chosen as study selection methodology).
- **Intervention:** what is going to be observed in the context of the planned SR.
- **Control:** baseline or initial data set that the researcher already possesses.
- **Effect:** types of results expected in the end of the systematic review.
- **Outcome measure:** metrics used to measure the effect.
- **Population:** population group that will be observed in relation to the intervention.
- **Application:** roles, professional types or application areas that will benefit from the systematic review results.
- **Experimental design:** describes how meta-analysis will be conducted, defining which statistical analysis methods will be applied on the collected data to interpret the results.

In the following section, we present the basis of a scientific research ontology development, which can function as a useful terminological model for improving the rigor of investigation in SE, particularly related to systematic review research.

7. Scientific research ontology for Experimental Software Engineering

As an explicit specification of the conceptualization, the scientific research ontology can be useful to guarantee the terminological homogeneity of the concepts that are to be used by different researchers, contributing to a higher consistency between the retrieved information and the consequent results.

Such ontology can also be a means to help assessing the present status of the different taxonomies that are used by the distinct SE digital libraries and article databases, as well as their evolution in time. It can function as a reference instrument to the distinct library taxonomies, for evaluating and enhancing the degree of their terminological precision, consistency, and extension, as well as increasing the degree of commensurability between them.

At the same time the scientific research ontology in SE can provide elements to help supporting the development of information extraction tools from the scientific texts in the field, by identifying, sorting, aggregating, and associating the related items.

It can also provide support to explicitly present to the professional the conceptual relations between these items, according to the existent background knowledge in the area, as represented in the ontology [73].

In order to accomplish this knowledge representation task, the research ontology must fulfill some basic requirements.

From the structural point of view, the organization of the conceptual structures need to include the two major types of branching hierarchy, namely taxonomies and meronomies (*meros* in greek means 'part'), respectively organized according to the *is_a* and the *part_of* kinds of conceptual relationships.

Other types of associative relations between the concepts must also be included, in order to be able to represent additional kinds of relevant knowledge structures.

From the semantic viewpoint, the information contents must contemplate the major categories of concepts pertaining to the different knowledge domains that are involved in the conduction of systematic reviews in Software Engineering. These conceptual entities are comprised by the following domains of the Scientific Research ontology: Experimental Method, Primary Research, Research Synthesis, and Software Engineering. The taxonymic relationships between these specific domains can be seen in the following diagram (Fig. 3), which presents a taxonomic hierarchy of domains in the field of scientific knowledge. The graphical notation that is used for the hierarchical diagrams has been adapted from the work of [74,75], and is applied to the conceptual meta-level of the models. Boxes are used to represent the concepts, and circles are used for the conceptual relations.

Information elements and structures about the scientific method itself, as well as about the experimental method of research in particular, constitute important parts of the knowledge to be represented in the ontology. Therefore, terms such as Experimental Unit, Population, Universe, Sample, Experimental Unit, Problem, Hypothesis, and Variable are some that might be included to represent fundamental concepts in the Experimental Method ontology. The two following graphs as displayed below (Figs. 4 and 5) illustrate the first three levels of the taxonomic hierarchy [76] of the present status of the knowledge representation of the Experimental Method. The conceptual entity Experimental Element is the highest level hypernym of the Experimental Method ontology, and subsumes the concepts in the lower levels of the hierarchy. For better visual clarity of the conceptual representation, each graph separately displays the second level subcategories.

Figs. 4 and 5 show the subcategories of the conceptual entities Unit of Study and Object Formulation. Fig. 6 displays the fourth level of the taxonomic hierarchy of the knowledge representation of the Experimental Method, showing the sibling subcategories of the Variable concept.

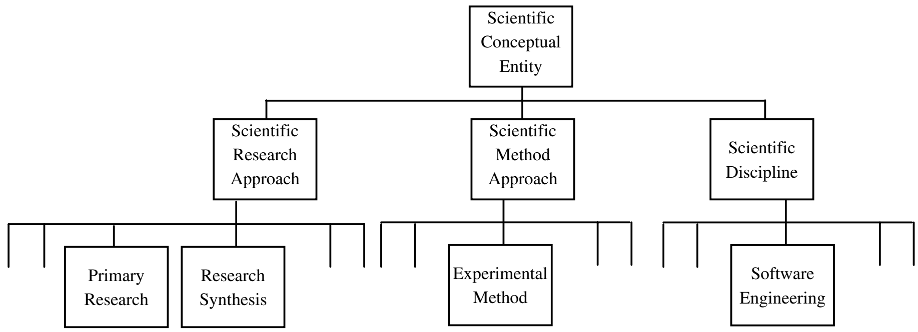

Fig. 3. Taxonymic relationships between domains.

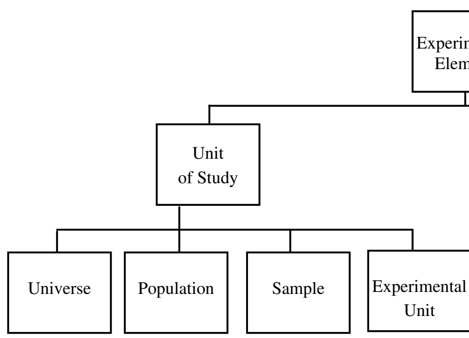

Fig. 4. Unit of study subcategories, in the Experimental Method ontology.

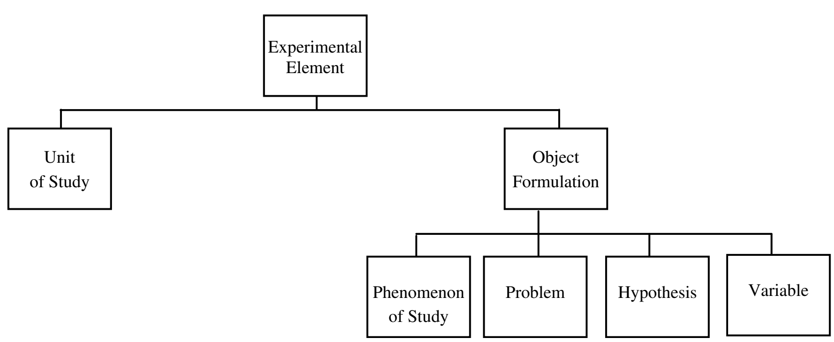

Fig. 5. Object formulation subcategories, in the Experimental Method ontology.

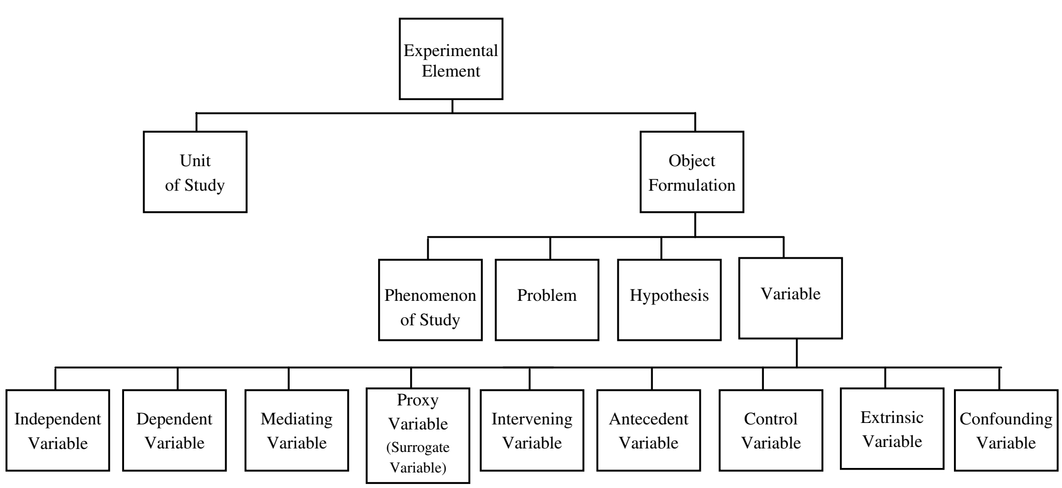

Fig. 6. Variable subcategories, in the Experimental Method ontology.

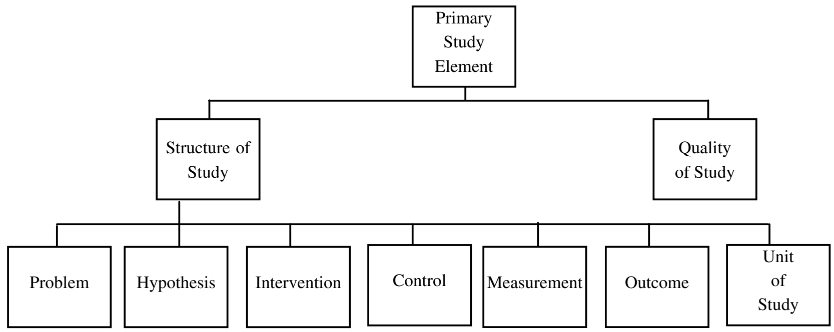

Fig. 7. Primary study structure, in the Primary Research ontology.

The concepts that are present in the Primary Research knowledge domain constitute a major parcel of this universe, since they constitute the material of Primary Studies upon which the Research Syntheses are conducted.

The two following graphs (Figs. 7 and 8) illustrate the first three levels of the taxonomic hierarchy of the present status of the Primary Research knowledge representation. The conceptual entity Primary Study Element is the highest level hypernym of the Primary Research ontology, and subsumes the concepts in the lower levels of the hierarchy. For better visual clarity of the conceptual representation, each graph separately displays the second level subcategories.

Figs. 7 and 8 show the subcategories of the conceptual entities Structure of Study and Quality of Study.

As aforementioned, besides containing the *taxonymic* type of relationships between concepts, the *is_a* type, giving origin to the correspondent taxonomic structures as displayed above, the ontology also contains the *meronymic* type of conceptual relationship, the *has* type, composing the resultant meronomic structures [77]. These can be formally represented both by frame and first order logics as well as using description logics.

Some examples of this second type of conceptual relationships, that can be found in the Primary Study ontology, informally presented in natural language, are:

- Hypothesis *has* Concepts.
- Hypothesis *has* Terms.
- Hypothesis *has* Formulation.

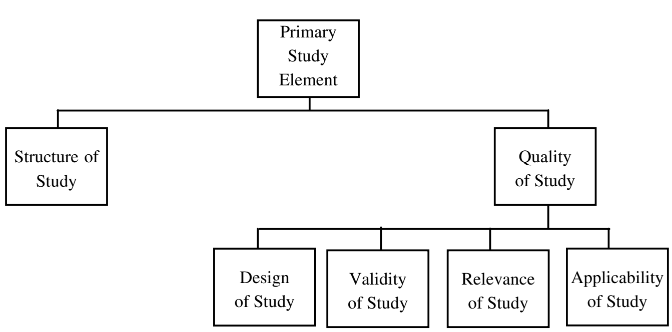

Fig. 8. Primary study quality, in the Primary Research ontology.

Regarding this last conceptual relationship, it is important to notice two major points. First, the Hypothesis Formulation conceptual structure is formally linked and dependent on the conceptual elements of the two other slots of the Hypothesis category. This is necessarily so, from the scientific methodological perspective. The second remark is that the Hypothesis Formulation conceptual structure is also directly linked with both sub-items of the first section of the Systematic Review Protocol Template, as presented above, in Section 6 of this paper.

Therefore, these conceptual links can be ontologically represented, and consequently formally defined, so that it can enhance both the robustness and the applicability of the proposed Primary Study ontology and better support the planning and conduction of systematic reviews in the field of Software Engineering.

It is noteworthy to say that, from the semantic structure point of view, the Hypothesis Formulation is a complex sentence, composed by three basic elements, which respectively correspond to the Independent Variable, the *influence* associative type of conceptual relation, and the Dependent Variable, as disposed in the following syntax:

- Independent Variable *influences* Dependent Variable

According to the complexity of the causal network relationships, both terms for the variables can have different concurrent attributes, such as: (a) be single or multiple; (b) be generic or specific; (c) be a function or a non-function between the same kind of variables.

As a consequence, these syntax variants can make the Hypothesis Formulation result in a more descriptive type of scientific hypothesis or alternatively in a more prescriptive one.

Therefore we can have different resulting formats for the hypothesis formulation syntax, such as the following examples:

- What Condition *influences* Phenomenon A of Population B?
- Does Characteristics A (B, C, ...) of the Phenomenon D *influence* Phenomenon E of Population F?
- Does Doing A (B, C, ...) in/to the Phenomenon D *influence* Phenomenon E of Population F?

Such as the different levels of the taxonomic structures, that allow representing nested taxonymic links between different levels of concept specificity, some nested meronymic conceptual relationships can also be represented. An example of this type of concept specificity is the following:

- Intervention *has* intervention effect.
- Intervention effect *has* magnitude.
- Intervention effect *has* precision.
- Intervention effect *has* reproducibility.

Meronomic hierarchies can also be graphically displayed for visually representing the correspondent conceptual structures and relationships. The following graphs (Figs. 9 and 10) show this type of hierarchical organization, respectively for the concepts Measurement and Outcome, in the Primary Research ontology. As a consequence of the combination of both conceptual hierarchy types, the resulting schematic representations correspond to hybrid kinds of ontologies [78]: Fig. 9 shows the hierarchical organization for the concept Measurement. Fig. 10 displays both types of conceptual structures for the category Outcome, in such a way that hybridity is presented at the same hierarchical level.

At the same time, terms and concepts referring to the four experiment types in Software Engineering might as well be represented. The Guide to the Software Engineering Body of Knowledge (SWEBOK) [79] provides an important source of conceptual and terminological entities, which constitute the ISO recommended standards of contents that should be present in the Software Engineering field as a discipline. Current initiatives in developing ontologies, based on this material, point to the relevant role of such standards for providing a basis to improve the scientific commensurability of the SE field.

The knowledge domain of the secondary studies also corresponds to a major parcel of the elements to be represented, since they constitute the major conceptual categories for the professional to define, during the planning, executing, and analyzing phases of the systematic review approach.

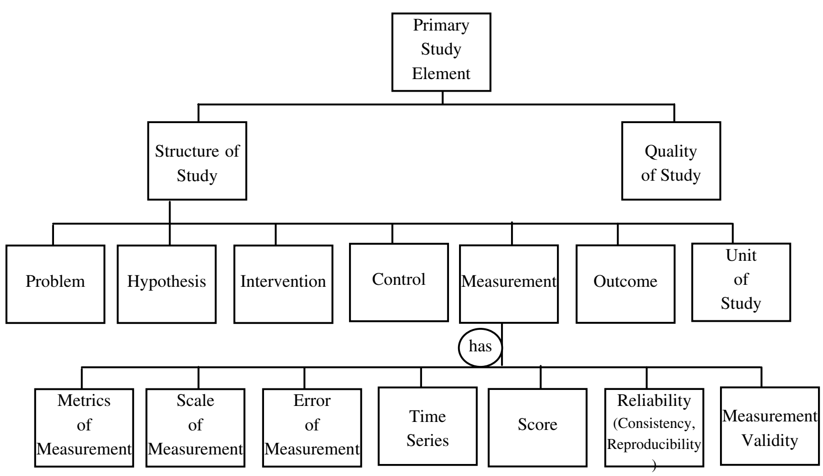

Fig. 9. Ontological hybridism represented by meronymic relations of the measurement concept subcategories plus taxonymic relations of its supercategories.

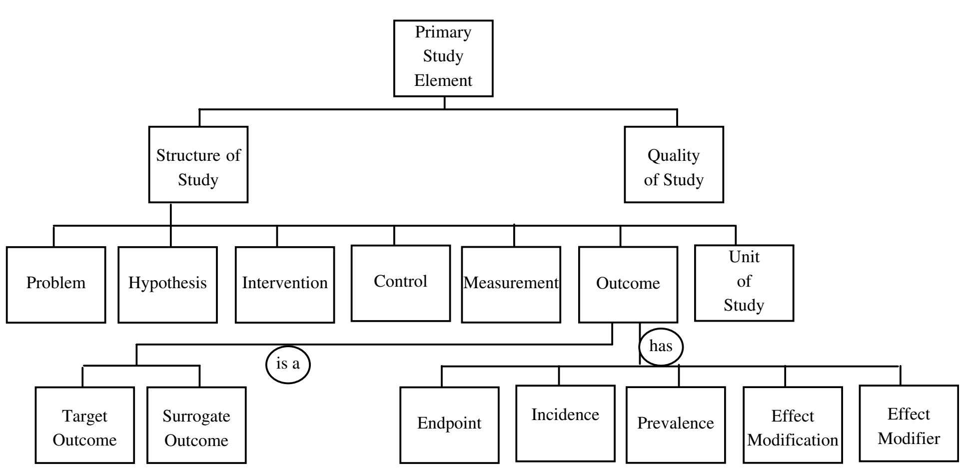

Fig. 10. Ontological hybridism represented by meronymic plus taxonymic relations of the outcome concept in the same level of categorization (subcategories).

In this sense, terms such as Internal Validity, Setting Item, Heterogeneity, Eligibility Criterion, Exclusion Criterion, Formal Methods of Retrieval, Free-Text Search, Inter-rater Reliability, and Coding Rule are included in the Research Synthesis ontology.

The three following graphs as displayed below illustrate the first three levels of the taxonomic hierarchy of the present status of the knowledge representation of Research Synthesis. The conceptual entity Secondary Study Element is the highest level hypernym of the Research Synthesis ontology, and subsumes the concepts in the lower levels of the hierarchy. For better visual clarity of the conceptual representation, each graph (Figs. 11–13) separately displays the second level subcategories.

Figs. 11–13 show, respectively, the subcategories of the conceptual entities Structure of Study, Methodological Procedure and Quality of Study.

It is worthwhile to mention that there exist natural superimpositions between some of the conceptual categories of the Primary Research ontology and the Research Synthesis one, since this second scientific research methodological approach, the systematic review methodology, is based on the first one and corresponds to a higher level of abstraction and generalization of its experimentally obtained data.

These natural ontological correspondences between both knowledge representation models point to some of the taxonomic links that can be established between them through these specific conceptual nodes.

Other kinds of conceptual relationships can be formally defined as well, in an inter-ontological space [80], based not only on other taxonymic relations that are more implicit, but also on using meronymic and other kinds of associative conceptual relations between nodes.

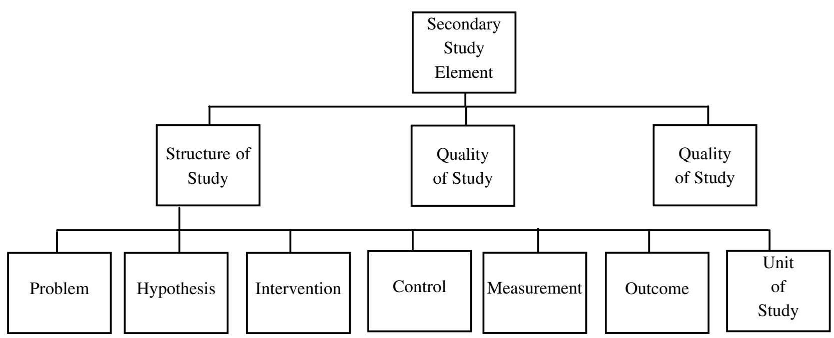

Fig. 11. Secondary study structure, in the Research Synthesis ontology.

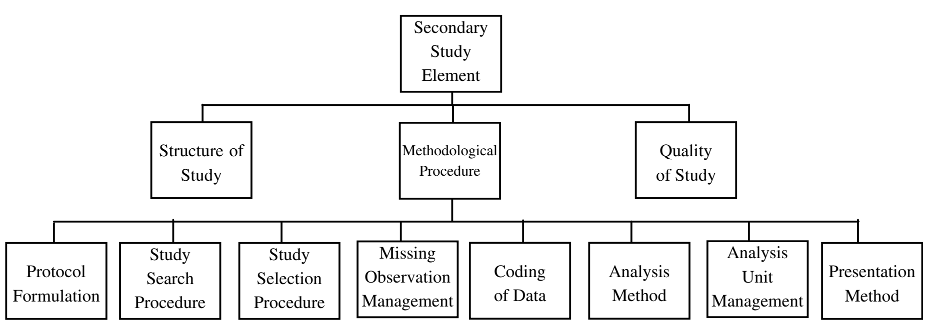

Fig. 12. Methodological procedure subcategories, in the Research Synthesis ontology.

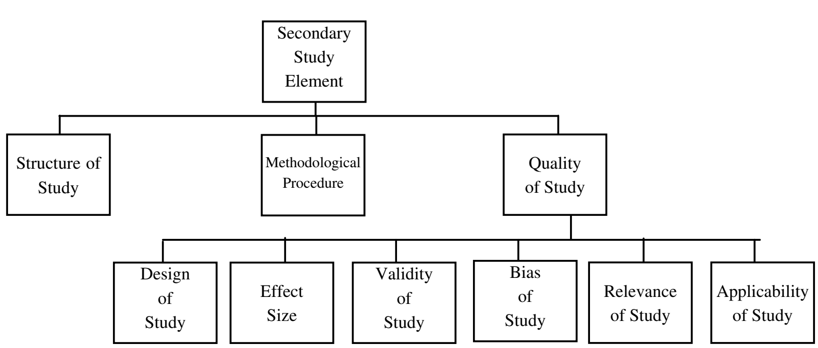

Fig. 13. Secondary study quality, in the Research Synthesis ontology.

This knowledge engineering methodological process contributes to add value to the ontology models that are being developed, both for scientific research in general and for the Software Engineering field in particular. Through the incorporation of these multidimensional semantic contents in the knowledge representation models, these elements and structures can be used to enhance the intelligence properties of knowledge-based systems.

8. On going and future work

In this paper, the importance of primary and secondary experimental studies have been discussed, highlighting their basic concepts and presenting how these concepts can be organized aiming at supporting the use of scientific methodology in Software Engineering. In addition, in Section 4 there are indications on how systematic reviews can support engineering process to build new software technologies. The Scientific Research Ontology for Software Engineering presented in this work is presently under development, and it is based on a conception of ontology building life cycle that aims at contriving evolving prototypes that are submitted to progressive changes at each stage of construction [81]. Delimiting the scope of application by the SR generated knowledge might allow the ontology to undergo a more controlled experimental evaluation process.

As the knowledge acquisition and the documentation processes, the evaluation process is also performed as a relevant supporting activity throughout the whole ontology building life cycle. These processes are conducted in parallel to the ontology development activities, and depend on each phase of the construction process as well as on the corresponding prototype that has been contrived. The development of definitions of formal axioms and rules is an important stage in this process. It aims at guaranteeing a higher degree of consistency and robustness of the conceptualization and must be enhanced.

This Scientific Research Ontology has been built to organize knowledge that is being generated from the conduction of systematic reviews by our research group on Experimental Software Engineering at COPPE/UFRJ (http://www.cos.ufrj.br/~ese). The primary and secondary studies' domain we are concerned with is Software Engineering. Therefore, a Scientific Research Ontology should also include knowledge about this domain. For that reason, a *Software Engineering Ontology* must be integrated to the proposed Scientific Research Ontology. This SE Ontology may be composed of many already described sub-ontologies, such as *Software Process Ontology* [20], *Enterprise Ontology* [21], *Software Maintenance Ontology* [19], or even by aggregating new ones regarding, for instance, to *Web Applications*. To illustrate this integration, Fig. 14 shows the relation between Scientific Research Approach Ontology and one of Software Engineering's sub-ontologies (Software Process Ontology). This relation implies that every scientific approach is guided by a process.

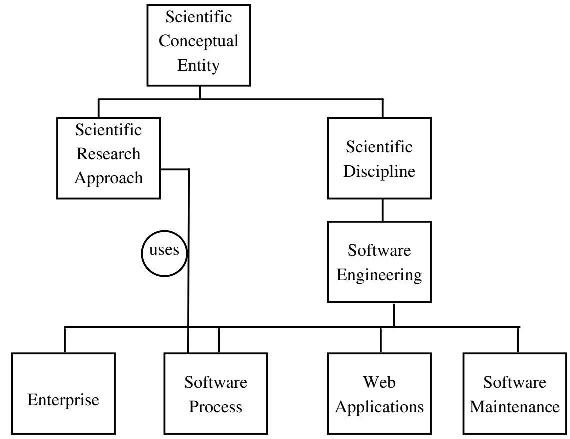

Fig. 14. Ontologies integration.

To support using all these ideas, one of our Experimental Software Engineering group's ongoing research work concerns the definition and building of a computerized infrastructure to support the management of knowledge involved in the experimentation process in Software Engineering. This infrastructure, called eSEE (*experimental **S**oftware **E**ngineering **E**nvironment*), is able to instantiate software engineering environments in order to support the definition, planning, execution and packaging of experimental studies, as well as the management of knowledge produced throughout the experimentation process [82].

We believe that eSEE could support the SR process and packaging of secondary study results. Therefore, besides supporting primary studies, the eSEE infrastructure also is being built to support secondary ones. In order to do so, the Systematic Review Conduction Process [22,72] will compose the eSEE's experimental processes' repository in order to allow the instantiation of experimental environments to support the execution of systematic reviews.

An infrastructure such as eSEE would also benefit from the definition of an ontology to organize knowledge about experimental software engineering. This ontology would be the core to knowledge retrieval, to identify the general characteristics of the experimental study types and common knowledge regarding SE experimental studies, as well as to allow the communication among users and tools, opening the opportunity to explore e-science concepts into the Software Engineering experimentation domain. The scientific research ontology to support systematic review described in this article represents an initial step towards a wider *Experimental Software Engineering Ontology*.

Acknowledgements

The authors acknowledge CAPES, CNPq and FAPEAM for the financial support to this work. The authors especially thank professors Barbara Kitchenham and Emilia Mendes for sending the initial materials regarding this area. We recognize the contributions of Arilo Claudio Dias Neto, Jobson Luiz Massolar da Silva, Rafael Barcelos, Reinaldo Cabral, Rodrigo de Oliveira Spinola and Somulo Mafra, from the Experimental Software Engineering Group (http://www.cos.ufrj.br/~ese), to providing the data regarding their secondary studies.

This work has been developed under the CNPq Grant 472135/2004-0. (Experimental Software Engineering Project.)

References

[1] D. Sjøberg, J. Hannay, O. Hansen, V. Kampenes, A. Karahasanović, N. Liborg, A. Rekdal, A survey of controlled experiments in software engineering, IEEE Transactions on Software Engineering 31 (9) (2005) 733–753.

[2] D. Rodríguez, Empirical software engineering research: epistemological and ontological foundations, in: First Workshop on Ontology, Conceptualizations and Epistemology for Software and Systems Engineering (ONTOSE), 2005.

[3] M.V. Zelkowitz, D.R. Wallace, D.W. Binkley, Experimental validation of new software technology, in: Natalia Juristo, Ana M. Moreno (Eds.), Lecture Notes on Empirical Software Engineering, Series on Software Engineering and Knowledge Engineering, World Scientific, 2003, pp. 229–263, ISBN 981-02-4914-4 (Chapter 6).

[4] M.B. Chrissis et al., CMMI: Guidelines for Process Integration and Product Improvement, Addison-Wesley, 2003.

[5] N. Juristo, A. Moreno, Basics of Software Engineering Experimentation, first ed., Kluwer Academic Press, 2001.

[6] C. Wohlin, M. Höst, K. Henningsson, Empirical research methods in web and software engineering, in: Emilia Mendes, Nile Mosley (Eds.), Web Engineering, Springer, 2006, pp. 409–429, ISBN 3-540-28196-7 (Chapter 13).

[7] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, A. Wesslén, Experimentation in Software Engineering – An Introduction, Kluwer Academic Publishers, 2000.

[8] V. Basili, F. Shull, F. Lanubile, Building knowledge through families of experiments, IEEE Transactions on Software Engineering 25 (4) (1999) 456–473.

[9] F. Shull, J. Carver, G. Travassos, An empirical methodology for introducing software processes, in: Proceedings of Joint 8th European Software Engineering Conference (ESEC) and 9th ACM SIGSOFT Foundations of Software Engineering (FSE-9). Vienna, Austria, 2001, pp. 288–296.

[10] B. Kitchenham, Procedures for performing systematic reviews, Joint Technical Report Software Engineering Group, Department of Computer Science, Keele University, United King and Empirical Software Engineering, National ICT Australia Ltd., Australia, 2004.

[11] S.N. Mafra, R.F. Barcelos, G.H. Travassos, Applying an evidence based methodology on the definition of new software technologies, in: SBES 2006: XX Brazilian Symposium on Software Engineering. Brazilian Computer Society, Brazil, 2006 (in Portuguese).

[12] R.F. Barcelos, G.H. Travassos, ArqCheck: a checklist based approach to inspect software architecture models, in: SBQS2006: V Brazilian Symposium on Software Quality, Brazilian Computer Society, Brazil, 2006 (in Portuguese).

[13] S.N. Mafra, Perspective based reading: the object oriented project view, Master Dissertation, PESC, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2006.

[14] H. Cooper, L.V. Hedges (Eds.), The Handbook of Research Synthesis, Russell Sage Foundation, New York, 1994.

[15] P. Mian, T. Conte, A. Natali, J. Biolchini, E. Mendes, G. Travassos, Lessons learned on applying systematic reviews to software engineering, in: Proceedings of the 3rd International Workshop Guidelines for Empirical Work in the Workshop Series on Empirical Software Engineering (WSESE2005). Finland, 2005.

[16] D.E. O'Leary, Impediments in the use of explicit ontologies for KBS development, International Journal of Human–Computer Studies 46 (2–3) (1997).

[17] B. Swartout, R. Patil, K. Knight, T. Russ, Toward distributed use of large-scale ontologies, Ontological Engineering – Working Notes, Stanford, CA, March, 1997.

[18] R. Jasper, M.A. Uschold, Framework for understanding and classifying ontology applications in: Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW'99), Alberta, Canada, 1999.

[19] B. Kitchenham, G. Travassos, A. Mayrhauser, et al., Towards an ontology of software maintenance, Journal of Software Maintenance: Research and Practice 11 (6) (1999) 365–389.

[20] R. Falbo, C. Menezes, A.R. Rocha, Using ontologies to improve knowledge integration in software engineering environments, Proceedings of the International Conference on Information Systems Analysis and Synthesis (1998) 1/8–8/8.

[21] K. Villela, G. Santos, L. Schnaider, A.R. Rocha, G.H. Travassos, The use of an enterprise ontology to support knowledge management in software development environments, Brazilian Computer Society Journal 2 vol. II (2005).

[22] J. Biolchini, P.G. Mian, A.C. Natali, G.H. Travassos, Systematic review in software engineering: relevance and utility, Technical Report, PESC – COPPE/UFRJ, Brazil. Available from: http://cronos.cos.ufrj.br/publicacoes/reltec/es67905.pdf, 2005.

[23] T.R. Gruber, What is an Ontology? Stanford: Knowledge Systems Laboratory, Stanford University. http://www-ksl.stanford.edu/kst/what-is-an-ontology.html (accessed by 14.07.05), 2001.

[24] M. Egger, G.D. Smith, D.G. Altman (Eds.), Systematic Reviews in Health Care. Meta-analysis in Context, BMJ Publishing Group, London, 2001, pp. 47–369.

[25] M. Clarke, Oxman, A.D. (Eds.), Cochrane Reviewers' Handbook 4.2.0 [updated March 2003]. In: The Cochrane Library, Issue 2. Oxford: Update Software. Available from: <http://www.cochrane.dk/cochrane/handbook/hbook.htm>, 2003.

[26] M. Pai, M. Mcculloch, J.D. Gorman, N. Pai, W. Enanoria, G. Kennedy, P. Tharyan, J.M. Colford Jr., Systematic reviews and meta-analyses: an illustrated, step-by-step guide, The National Medical Journal of India 17 (2) (2004) 86–95.

[27] R. Gross, Decisions and Evidence in Medical Practice, Mosby, Inc., St. Louis, 2001.

[28] A.R. Feinstein, Meta-analysis: statistical alchemy for the 21st century, Journal of Clinical Epidemiology 48 (1) (1995) 71–79.

[29] A. Liberati, Meta-analysis: statistical alchemy for the 21st century: discussion. A plea for a more balanced view of meta-analysis and systematic overviews of the effect of health care interventions, Journal of Clinical Epidemiology 48 (1) (1995) 81–86.

[30] R.T. Birge, The calculation of errors by the method of least squares, Physical Review 40 (1932) 207–227.

[31] L.H.C. Tippett, The Methods of Statistics, Williams&Norgate, London, 1931.

[32] R.A. Fisher, Statistical Methods for Research Workers, fourth ed., Oliver&Boyd, London, 1932.

[33] W.G. Cochran, Problems arising in the analysis of a series of similar experiments, Journal of the Royal Statistical Society 4 (1937) 102–118.

[34] F. Yates, W.G. Cochran, The analysis of groups of experiments, Journal of Agricultural Science 28 (1938) 556–580.

[35] K.A. Feldman, Using the work of others: some observations on reviewing and integrating, Sociology of Education 4 (1971) 86–102.

[36] R.J. Light, P.V. Smith, Accumulating evidence: procedures for resolving contradictions among research studies, Harvard Educational Review 41 (1971) 429–471.

[37] T.C. Taveggia, Resolving research controversy through empirical cumulation: toward reliable sociological knowledge, Sociological Methods and Research 2 (1974) 395–407.

[38] G.V. Glass, Primary, secondary, and meta-analysis, Educational Researcher 5 (1976) 3–8.

[39] M.L. Smith, G.V. Glass, Meta-analysis of psychotherapy outcome studies, American Psychologist 32 (1977) 752–760.
[40] F.L. Schmidt, J.E. Hunter, Development of a general solution to the problem of validity generalization, Journal of Applied Psychology 62 (1997) 529–540.
[41] R. Rosenthal, D.B. Rubin, Interpersonal expectancy effects: the first 345 studies, Behavioral and Brain Sciences 3 (1978) 377–386.
[42] G.V. Glass, M.L. Smith, Meta-analysis of research on the relationship of class size and achievement, Educational Evaluation and Policy Analysis 1 (1978) 2–16.
[43] I. Chalmers, M. Enkin, M.J.N.C. Keirse, Effective Care in Pregnancy and Childbirth, Oxford University Press, Oxford, 1989.
[44] G.V. Glass, B. McGaw, M.L. Smith, Meta-analysis in Social Research, Sage, Beverly Hills, 1981.
[45] J.E. Hunter, F.L. Schmidt, G.B. Jackson, Meta-analysis: Cumulating Research Findings Across Studies, Sage, Beverly Hills, 1982.
[46] R. Rosenthal, Meta-analytic Procedures for Social Research, Sage, Beverly Hills, 1984.
[47] L.V. Hedges, I. Olkin, Statistical Methods for Meta-analysis, Academic Press, Orlando, 1985.
[48] R.J. Light, D.B. Pillemer, Summing Up: The Science of Reviewing Research, Harvard University Press, Cambridge, 1984.
[49] S. Piantadosi, Clinical Trials – A Methodological Perspective, John Wiley & Sons Inc., New York, 1997.
[50] K.S. Warren, F. Mosteller (Eds.), Doing More Good Than Harm – The Evaluation of Health Care InterventionsAnnals of the New York Academy of Sciences 703 (1993) 18–32.
[51] D.L. Sackett, S.E. Straus, W.S. Richardson, W.M.C. Rosenberg, R.B. Haynes, Evidence-Based Medicine, Churchill Livingstone, Edinburgh, 2000.
[52] C.S. Cina et al., Carotid endarterectomy for symptomatic carotid stenosis, Cochrane Database of Systematic Reviews 2: CD001081, 2000.
[53] ISIS-2 (Second International Study of Infarct Survival) Collaborative Group, Randomized trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Lancet ii: (1988), pp. 349–360.
[54] Y. Kwok et al., Meta-analysis of exercise testing to detect coronary artery disease in women, American Journal of Cardiology 83 (5) (1999) 660–666.
[55] Early Breast Cancer Trialists' Collaborative Group, Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomized trials of more than 1000 patients, Lancet 343 (1994) 311–322.
[56] A.H. Malik, W.M. Lee, Chronic hepatitis B virus infection: treatment strategies for the next millennium, Annals of Internal Medicine 132 (2000) 723–731.
[57] ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group, Protocol for a large, simple study of the effects of oral mononitrate, of oral captopril, and of intravenous magnesium, American Journal of Cardiology 68 (1991) 87D–100D.
[58] T.E. Towheed, M.C. Hochberg, A systematic review of randomized controlled trials of pharmacological therapy on osteoarthritis of the knee, with an emphasis on trial methodology, Seminars in Arthritis and Rheumatism 26 (5) (1997) 755–770.
[59] Early Breast Cancer Trialists' Collaborative Group, Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomized trials involving 31,000 recurrences and 24,000 deaths among 75,000 women, Lancet 339 (1992) 1-15 (Part I) and 71-85 (Part II).
[60] J.W. Williams et al., A systematic review of newer pharmacotherapies for depression in adults: evidence report summary, Annals of Internal Medicine 132 (2000) 743–756.
[61] B.A. Kitchenham, T. Dybå, M. Jørgensen, Evidence-based software engineering, Proceedings of ICSE (2004) 273–281, IEEE Computer Society Press.
[62] A. Endres, D. Rombach, A Handbook of Software and Systems Engineering: Empirical Observations, Laws, and Theories, The Fraunhofer IESE Series on Software Engineering, Addison-Wesley, 2003.
[63] T. Conte, E. Mendes, G.H. Travassos, Web application development processes: a systematic review, in: Proceedings of WEBMEDIA – Brazilian Symposium on Multimedia Systems and Web, Brazil, 2005 (in Portuguese).
[64] E. Mendes, B. Kitchenham, Protocol for systematic review. http://www.cs.auckland.ac.nz/emilia/srspp.pdf (accessed 10.05.05), 2004.
[65] R. F. Barcelos, G.H. Travassos, Evaluation approaches for software architectural documents: a systematic review, in: Proceedings of the Ibero–American Workshop on Requirements Engineering and Software Environments (IDEAS 2006), Argentina, 2006.
[66] R. Cabral, G.H. Travassos, Inspection applied to software process models, Technical Report (Course COS722), PESC – COPPE/UFRJ, Brazil, 2004 (in Portuguese).
[67] S. Mafra, G.H. Travassos, Software reading techniques: a systematic review, in: XIX Brazilian Symposium on Software Engineering, Brazilian Computer Society, Brazil, 2005 (in Portuguese).
[68] S.N. Mafra, G.H. Travassos, Perspective based reading: OO designer view, in: SBQS2006: V Brazilian Symposium on Software Quality, Brazilian Computer Society, Brazil, 2006 (in Portuguese).
[69] R. Spinola, J. Massolar, G.H. Travassos, Towards a conceptual framework to classify ubiquitous software projects, in: Proceedings of the International Conference on Software Engineering and Knowledge Engineering (SEKE 2006), San Francisco, USA, 2006.
[70] A.C. Dias Neto, G.H. Travassos, Maraká: a computerized infrastructure to support software testing planning and control, in: SBQS2006: V Brazilian Symposium on Software Quality, Brazilian Computer Society, Brazil, 2006 (in Portuguese).
[71] B. Kitchenham, E. Mendes, G.H. Travassos, A systematic review of cross- vs. within-company cost estimation models, in: Proceedings of EASE 2006 – Empirical Assessment in Software Engineering, UK, 2006.
[72] P. Mian, T. Conte, A. Natali, J. Biolchini, G. Travassos, A systematic review process to software engineering, in: Proceedings of the 2nd Experimental Software Engineering Latin American Workshop (ESELAW'05), Brazil, 2005.
[73] D. Milward, M. Bjäreland, W. Hayes, M. Maxwell, L. Öberg, N. Tilford, J. Thomas, R. Hale, S. Knight, J. Barnes, Ontology-based interactive information extraction from scientific abstracts, in: Proceedings of BioLINK SIG Text Mining Workshop, ISMB/ECCB, Glasgow, 2004.
[74] J.F. Sowa, Conceptual Structures – Information Processing in Mind and Machines, Addison-Wesley Publishing Company, Reading, 1984.
[75] J.F. Sowa, Relating diagrams to logic, in: G.W. Mineau, B. Moulin, J.F. Sowa (Eds.), Conceptual Graphs for Knowledge Representation, Springer-Verlag, Berlin, 1993.
[76] G.L. Murphy, M.E. Lassaline, Hierarchical structure in concepts and the basic level of categorization, in: K. Lambert, D. Shanks (Eds.), Knowledge, Concepts, and Categories, Psychology Press, Hove, 1997.
[77] D.A. Cruse, Lexical Semantics, Cambridge University Press, Cambridge, 1997.
[78] C. Habel, S. Pribbenow, G. Simmons, Partonomies and Depictions: A Hybrid Approach, in: J. Glasgow, N.H. Narayanan, B. Chandrasekaran (Eds.), Diagrammatic Reasoning – Cognitive and Computational Perspectives, The AAAI Press, Menlo Park, 1995.
[79] P. Bourque, R. Dupuis, A. Abran, J.W. Moore, L.L. Tripp, The guide to the software engineering body of knowledge, IEEE Software 16 (1999) 35–44.
[80] P. Gärdenfors, Conceptual Spaces – The Geometry of Thought, The MIT Press, Cambridge, 2000.
[81] A. Gómez-Pérez, M. Fernández-López, O. Corcho, Ontological Engineering, Springer-Verlag, London, 2004.
[82] P. Mian, W. Chapetta, P.S. Santos, C.R. Melo Jr., A.C.C. Natali, J. Biolchini, G. Travassos, A.R. Rocha, eSEE: an infrastructure for supporting experimental software engineering, in: Proceedings the 4th IEEE/ACM International Symposium on Empirical Software Engineering (ISESE2005) – Late Breaking Paper, Australia, 2005.

Jorge Biolchini is a medical doctor, presently conducting a post-doctoral research in the Experimental Software Engineering Group of the Systems Engineering and Computer Science Program at COPPE – Federal University of Rio de Janeiro (UFRJ) – Brazil. He received his doctorate degree from UFRJ in 2003, with a fellowship research at the National Center for Biomedical Communications – National Institutes of Health (NIH), in Bethesda, USA. His research interests address to the fields of knowledge representation, ontology development in the medical and computer science domains, e-science, cognition and decision making support methods, and experimental software engineering. Most of the information regarding his research projects and working activities can be found at http://www.cos.ufrj.br/~jorgebio.

Paula Gomes Mian is a doctorate student at COPPE – Federal University of Rio de Janeiro – Brazil. She is a former student from Federal University of Espírito Santo, from where she received her master degree in 2003. Her current research interests include experimental software engineering, ontologies, software engineering environments and process improvement. She has industrial experience in software development and process implementation consulting. She is a member of the Brazilian Computer Society. Most of the information regarding her research projects and working activities can be found at http://www.cos.ufrj.br/~pgmian.

Ana Candida Natali is a doctorate student at COPPE – Federal University of Rio de Janeiro – Brazil. She is a former student from Federal University of Espírito Santo, from where she received her master degree in 2003. Her current research interests include software quality, experimental software engineering, ontologies, software engineering environments and process improvement. She has industrial experience in software development and process implementation consulting. She is a member of the Brazilian Computer Society. Most of the information regarding her research projects and working activities can be found at http://www.cos.ufrj.br/~anatali.

Tayana Conte is an Assistant Professor of Computer Science Department at UFAM – Federal University of Amazonas – Brazil. She is also a doctorate student at COPPE – Federal University of Rio de Janeiro, where she is a member of the ESE (Experimental Software Engineering Group). She received her master degree from Federal University of Minas Gerais in 2001. Her current research interests include software quality, verification and validation, web engineering, web usability and experimental software engineering. She is a member of the Brazilian Computer Society (SBC). Most of the information regarding her research projects and working activities can be found at http://ese.cos.ufrj.br/ese.

Guilherme H. Travassos is an Associate Professor of Software Engineering in the Systems Engineering and Computer Science Program at COPPE – Federal University of Rio de Janeiro – Brazil and a CNPq – Brazilian Research Council researcher. He received his doctorate degree from COPPE/UFRJ in 1994. He has been with the Experimental Software Engineering Group at the University of Maryland – College Park for a post-doctoral position (1998/2000). He leads the Experimental Software Engineering Group at COPPE/UFRJ. His current research interests include experimental software engineering, e-science, software quality, software engineering environments and verification, validation and testing concerned with object-oriented software. He is member of the ISERN, ACM and SBC – Brazilian Computer Society. Most of the information regarding his research projects and working activities can be found at http://www.cos.ufrj.br/~ght.