Integrating Empirical Software Engineering practice in South America

Gabriela Robiolo

Departamento de Informática de la Facultad de Ingeniería de la Universidad Austral

Mariano Acosta 1901(B1629WWA), Derqui, Pilar, Buenos Aires

grobiolo@austral.edu.ar

Abstract

Empirical software engineering (ESE) is a sub-domain of software engineering which focuses on experiments on software systems. Its main interest lies on devising software experiments, on collecting data from these experiments, and on formulating laws and theories from these data. In South America there is a group of researchers that are involved in this topic and have interests in common. This project propose the integration of their work in order to reply the experimentation done in different countries contributing to the increase of empirical software engineering body of knowledge. At this time several publications have been done with the collaboration of master students.

Key words:

Context

Empirical software engineering (ESE) is a sub-domain of software engineering which focuses on experiments on software systems (software products, processes, and resources). Its main interest lies on devising software experiments, on collecting data from these experiments, and on formulating laws and theories from these data. Proponents of experimental software engineering advocate that experimentation is an important method towards contributing to the accumulation of knowledge in software engineering. The methods most frequently used include experiments, a variety of case studies, surveys, and statistical analyses.

Basili (Boehm et al. 2005), who has made an important legacy in this discipline, has stated that -just like physics, medicine, manufacturing, and many other disciplines- software engineering requires the same high level approach to enlarge the knowledge of the discipline: the cycle of model building, experimentation and teaming. He believes that people working in the software engineering area cannot rely solely on observation followed by logical thought. He thinks that software engineering is a laboratory science, which involves an experimental component which tests or disproves theories, and explores new domains. He has proposed experimenting with different techniques to see how and when they really work, to understand their limits, and to understand how to improve them.

Basili (Basili et al. 2007) advocates that the software engineering discipline needs research that will help establish a scientific and engineering basis for the software engineering field. To this end, researchers should be able to build, analyze and evaluate models of the software processes and products, as well as models of the various aspects of the environment in which software is built. Once the models have been obtained, it is especially important to study the interactions of such models: the goal is to develop the conceptual scientific foundations of software engineering, upon which future researchers can build.

It is important to note that one of the challenges the profession is facing is having access to industry-based laboratories that will allow researchers to observe, build and analyze models. Moreover, since each such

laboratory will only provide local -rather than global- models, many experimental laboratories at multiple levels will be necessary to help us generate the basic models and metrics to be used both in business and in science.

On the other hand, the processing and production models generated by researchers should later on be tailored to fit the needs of organizations, using the data collected within each organization, and these models should be flexible enough to continually evolve, based upon the organization's evolving experiences.

Most researchers and practitioners interested in ESE usually participate at ACM / IEEE International Symposia on Empirical Software Engineering and Measurement (ESEM), which have become the most important international event where to present research results related to ESE. These include the discussion of: i) the strengths and weaknesses of software engineering technologies and methods, from an empirical viewpoint; ii) the design and analysis of empirical studies, ranging from controlled experiments to field studies and from quantitative to qualitative studies; iii) the use of data and measurement to understand, evaluate, and model software engineering phenomena. The symposia encourage the presentation of both novel work and replication studies.

As regards publications, the Springer Journal of Empirical Software Engineering provides a forum for applied software engineering research with a strong empirical component. Its goal is to promote research on relevant problems, following the scientific method and to report the best practice in industry. It is therefore a venue where to publish empirical results which are relevant to both researchers and practitioners. Reported empirical studies usually involve the collection and analysis of data and experience that can be used to characterize, evaluate and reveal relationships among software development deliverables, practices, and technologies. Both original and replicated studies can be reported, varying from controlled experiments to field studies, from data intensive to qualitative studies. Preference is given to studies that can be replicated or expanded upon.

In South America, there are some research groups that are interested in this field and have been working on it for some time:

Brazil: led by Guilherme H. Travassos, who started the Experimental Software Engineering Group at COPPE/UFRJ in the context of Software Engineering research in 2000. Currently, ESE Group is staffed by a team of 17 members, 1 professor with dedication, 2 researchers, 4 Ph.D. students, 7 master's students, 3 undergraduate students and one staff member. During its existence they had published more than 190 papers involving experimentation and related issues, particularly in verification, validation and testing; knowledge management, experimentation environments, web engineering, e-science and software quality.

Argentina: led by Gabriela Robiolo, who started the SE line in UA in 2004. Currently, the group is made up of two Full Professors, one researcher who is a Ph.D. student, two graduate students and two undergraduate students. They have published more than 20 papers in the field, principally on measures in software design and requirements and early effort estimations and software re-engineeringe.

Uruguay: led by Martín Solari, who started replicating software engineering experiments in 2005. Currently he do research and coordinate the Centro de Investigación e Innovación en Ingeniería de Software (CIS) at Universidad ORT Uruguay. CIS has three full-time Ph.D. researchers and four master students. They have published several papers in the field of software process improvement, software testing, knowledge management, packaging and replication of software engineering experiments.

Chile; led by Jorge E. Diaz Villegas, executive director at Centro de Estudios de Ingeniería de Software at Universidad de la Frontera(CEIS UFRO). Currently, the group is made up of three professors, three researchers who are students the graduate

alicante University, four software engineer and six scholarship students. The group have 22 papers (2011-2012) on Software Engineering topics, mainly on Requirements Engineering. CEIS UFRO have developed a strong relationship with local software small business that has allowed formulating empirical-based proposals. They have developed a effort estimation method for web software projects based on more than 3 years managing and tracking software projects.

General and specific objectives of the proposed research work

General objective: to contribute to increase the body of knowledge of empirical software engineering, especially that on testing and effort estimation.

Specific objectives:

a. Replication in Argentina, Uruguay, Chile and Brazil of the empirical experience of the applicability of size measures to reinforce the expert judgment in the context of effort estimation of Agile projects. Experts estimate the effort of a user story based on their previous experience. This experience is usually based on analogies with similar user stories previously developed. These analogies are made intuitively or by using historical records. The empirical experience will measure the size of a set of historical stories with COSMIC and use the historical stories to estimate the new stories.

b. Replication in Argentina, Uruguay, Chile and Brazil of the empirical experience of comparing the simplicity for practitioners of using size and complexity functional measures. Three small web projects will be measured by practitioners using Function Points, COSMIC and Paths. A measure is simpler if it yields a smaller mean measurement time and if it displays a smaller standard deviation measurement value, and it is easier to understand if it displays a smaller normalized standard deviation measurement value.

c. Measurement of the functional complexity and functional size measure of some big industrial projects, in order to stress the accuracy of the effort estimation methods more frequently used by estimators. Paths, a functional complexity measure which may improve the accuracy of effort estimation, will be used to measure functional complexity. However, as this measure has only been employed in case studies comprising small academic and industrial projects so far, it is necessary to work with bigger and more diverse projects in order to be able to generalize the results.

d. Automatization of the measurement of the functional complexity measure in code. Paths will be used to measure complexity in use cases. However, in order to trace the "paths" from requirements to Code, it is necessary to measure them in the source code. To fulfill this objective, an automated tool will have to be developed.

e. Identification of objective variables, different from functional size and complexity measures, which may contribute significantly to the accuracy of effort estimation. An objective variable is the opposite of a subjective one, whose value is defined by the experience of an expert. Empirical experimentation with objective variables will be conducted in order to test their contribution to effort estimation.

f. Replication in Argentina, Uruguay and Chile of the survey on "the applicability and relevance of the Activity and Procedures for the Verification and Validation of software".This survey aims to investigate on the state of the practice on verification and software validation in the industry. The questions seek to capture information about the applicability and the degree of importance that the activities and procedures for V & V have within organizations.

f. Replication in Argentina, Uruguay, Chile and Brazil of an empirical study of the current state of software estimation in software

companies. The main objective of this study is to characterize the software project estimation process from two perspectives: the software process and the estimation methods, in order to characterize the methods used in the industry.

Related work

One major problem that arises when trying to integrate study results into a common body of knowledge is the heterogeneity of the reporting styles (Shull et al. 2008a). In order to avoid this problem, these authors proposed using a checklist to report the results obtained from controlled experiments, which is expected to support a systematic, standardized presentation of empirical research. This would improve reporting, in order to support readers in (1) finding the information they are looking for, (2) understanding how an experiment is conducted, and (3) assessing the validity of its results. These authors obtained a unified set of guidelines for reporting experiments in software engineering, which was used to produce the empirical reports previously published by the researchers of this project.

Replications play a key role in Empirical Software Engineering by allowing the community to build knowledge about which results or observations hold under which conditions (Shull et al. 2008b). The authors focused on exact replications, in which the procedures of an experiment are followed as closely as possible, and on conceptual replications, in which the same research question is evaluated by using a different experimental procedure. They discussed the role played by each type of replication in terms of its goals, benefits, and limitations. They also highlighted the importance of producing adequate documentation for an experiment (original or replication), in order to allow for replication. This paper defined a framework to guide our replication work.

Another paper that will guide this project is (Juristo and Vega 2011). Aimed at researchers running non-exact replications, their paper has defined a process with four phases: replication definition and planning, replication operation and analysis, replication interpretation, and analysis of the replication's contribution.

Moreover, the aggregation of evidence will be a challenge to tackle. The required steps (Basili et al.2007) are to identify the information needed, to collect appropriate studies, and to (objectively) aggregate (i.e., summarize) their results. To be able to make informed decisions on introducing, changing, or evolving technologies and processes in practice, as well as in research, these decisions have to be based on aggregated trustable (i.e., corroborated) evidence and statements. The benefits of such an approach include reducing the risk of introducing / changing technologies (from the industrial point of view), and the fact that it is possible to identify evidence gaps (from the research point of view).

Another important concern that is included in this project is software effort estimation, and the techniques and measures involved in it. Jorgensen and Shepped (2007) in "A Systematic Review of Software Development Cost Estimation Studies," analyzed 304 articles published in 76 publications. They pointed out that regression-based techniques predominated in 49% of the articles, and Function Points came second, with 22%. However, it is noteworthy that both showed a declining interest. On the other hand, the use of techniques for classification and regression trees represented 5% of the total, Bayesian networks 2%, with increasing interest, and neural network 7%, and a constant interest.

Jorgensen (Jorgensen, 2004) states that the estimate made by experts is the most frequently applied, and that there are situations in which you can expect this type of estimate to be more accurate than formal estimation models. At the same time, he states that the use of a combination of the estimation models and expert judgment is better than pure expert estimation.

Work already done by researchers

This project was previously defined as a three years project and was accepted in the "8º Concurso Interno de Proyectos de Investigación Científica 2012 and 2014" to be done in 2012- 2015".

Several contributions have been made during the project which are described following: a. *Functional size measure*: We investigate whether it is possible to take into account only subsets of Base Functional Components so as to obtain functional size measures that simplify Function Points with the same effort estimation accuracy as the original Function Points measure. b. *Paths*: a new empirical study was developed in order to obtained more evidence about the simplicity of Paths in comparison with Functions Points and COSMIC. The results have not been published yet. c. *Activities and procedures for verification and software validation*: a survey in Argentina was done. d. *Estimations methods used in the industry*: it was documented and validated an estimation method proposed by the CEIS UFRO and were analyzed in an empirical study if the use of historical data improves the expert estimation methods.
The six publications were done in the context of the first and second year of the project at national international conferences and Journals.
Also, two master's Thesis were finished.

Relevance and originality of the proposal

This project aims at integrating the varied work which is being performed in the context of empirical software engineering in order to enhance the strengths of each group: the significant experience in empirical research of the Brazilian group, the experience in measures and effort estimation of the Argentinean group, the experience in aggregating experimentation of the Uruguayan team and the important data bank developed together with the software industry by the Chilean group.

In fact, in order to be successful in empirical research, research replication becomes a critical aspect if generalized results are expected to be obtained. However, it is not easy to aggregate the results obtained in different contexts, so the experience of working on this topic is valuable.
A drawback of empirical research in the context of software engineering is the shortage of source data that fulfill the requirements of the research needs. Net-working will increase the availability of source data to make high level empirical software engineering publications possible.

References

Basili, V., Rombach, D., Schneider, K., Kitchenham, B., Pfahl, D. and Selby, R. (2007) Empirical Software Engineering Issues. Critical Assessment and Future Directions, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol. 4336.

Boehm, Barry; Rombach, Hans Dieter; and Zelkowitz, Marvin V., "Foundations of Empirical Software Engineering: The Legacy of Victor R. Basili" (2005). *About Harlan D. Mills.*

ISO 2003 ISO/IEC 20926: 2003, Software engineering – IFPUG 4.1 Unadjusted functional size measurement method – Counting Practices Manual, International Organization for Standardization, Geneve.

COSMIC 2009 – Common Software Measurement International Consortium, 2009. The COSMIC Functional Size Measurement Method - version 3.0.1 Measurement Manual (The COSMIC Implementation Guide for ISO/IEC 19761: 2003), May 2009.

Jorgensen, M. y Shepperd. 2007. A systematic review of software development cost estimation studies, IEEE Transactions on Software Engineering, Vol. 33, No. 1. p. 3-53, January.

Jorgensen, M. 2004. A review of studies on expert estimation of software development effort. Journal on System and Software, Vol. 70, No. 1-2, 37-60.

Juristo, N. and Vegas, S (2011) The role of non-exact replications in software engineering experiments. *Empirical Softw. Engg.* 16, 3 (June 2011), 295-324

Shull, F., Singer, J. and Sjøberg, D. (2008a) Guide to Advanced Empirical Software Engineering, 2008, Springer London.

Shull, F., Carver J., Vegas, S. and Juristo, N. (2008b). The role of replications in Empirical Software Engineering Empir Software Eng 13:211–218.