Aggregation of Empirical Evidence

Marcus Ciolkowski

One of the most important challenges in empirical software engineering today is to better integrate empirical studies with decision support, and to collect appropriate data and experiments. The required steps 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 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 industrial point of view), and that it is possible to identify evidence gaps (from research point of view).

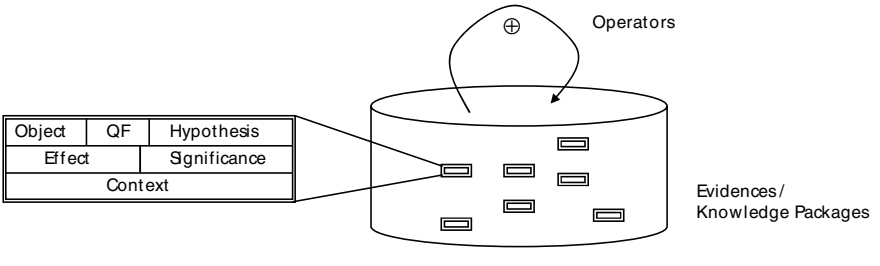
Today, the "state of the art" of aggregation in software engineering is to summarize a set of studies in a tabular form, which is a form of vote counting. Vote counting typically aggregates studies by counting the outcome of significance tests. Intuitively, if a large proportion of studies generate statistically significant results, then the overall effect can be interpreted as being non-zero. Vote counting, however, can be erroneous. In particular, for studies with small effect size and a low number of subjects (i.e., most of the studies in software engineering), vote counting can be strongly biased towards the conclusion that the treatment has no effect. This bias is not reduced as the number of studies increase.

Meta-analysis techniques promise to solve these problems in other fields, such as medicine or psychology. However, attempts at applying these techniques in software engineering have failed so far, as the meta analysis techniques used were not adequate to be applied at current software engineering experiments.

In other words, the problem is that systematic approaches are missing for (1) corroborating statements with (aggregated) trustable evidence, and (2) identifying the need for future studies.

To make progress to solve the stated problem, a framework is needed for goal-oriented aggregation of empirical results, linking empirical results with the goal of aggregation. In other words, the goal is to build corroborated statements from existing evidence.

The framework comprises a model of atomic evidence, knowledge packages (representing aggregated evidence), and aggregation operators to generalize and abstract knowledge.



V. Basili et al. (Eds.): Empirical Software Engineering Issues, LNCS 4336, p. 20, 2007.