Experience Report: Research Result Reproduction in a Grad Student Course

Dirk Beyer and Stefan Winter

LMU Munich, Munich, Germany

In earlier years, the unavailability of research artifacts has been a main obstacle for reproducing research results in various scientific disciplines. In computer science, we have been in the lucky situation that (a) research artifacts *can* be made publicly available with modest effort (compared to other disciplines) and (b) the possibility of a replication crisis has been identified early enough to implement counter-measures. Since its initiation at FSE 2011, *artifact evaluation* as an incentive for the publication of research artifacts has become widely adopted in the computer science landscape.

With numerous research artifacts available, research result reproducibility is facing a new challenge: The lacking incentive to actually use these artifacts for reproducing research results. Although some venues explicitly encourage reproduction and replication reports as submissions to their conference tracks, the number of published result reproductions and replications remains low. This comes at no surprise. Replicating, and even reproducing, research results entails enormous efforts that could also be spent on further advancing the frontiers of research, which is still more highly rewarded in terms of academic "performance" metrics, such as citation counts. In fact, the main benefit of a reproduction study is to establish trust in *someone else's work* and probably improves that other work's citation count, whereas scenarios under which a reproduction or replication attracts citations are rare.

To improve on this situation, we have launched a pilot program to involve CS graduate students (enrolled in a Master's program at LMU Munich) in the reproduction of research results. Just as artifact evaluations are commonly conducted by junior research scientists in the field, reproducibility studies can serve as an attractive pathway to first experiences with scientific publications and at the same time provide a hands-on exposure to cutting-edge research. In this talk we report on our experiences with attempting a total of 64 result reproductions by 13 students on the basis of published research artifacts. Overall, we find that

- graduate students can conduct research reproductions using research artifacts, are highly motivated to do so, and can accomplish the task in a reasonable amount of time for most research artifacts
- certain technology that has become standard in modern CS research (e.g., container-based isolation) has not permeated graduate study program curricula, which should be addressed if we want the next generation of research scientists to continue improving on research reproducibility

 – the outcome of the reproduction attempts highlight a number of open problems in the artifact evaluation or publication culture that may have remained uncovered without the reproduction attempts that we conducted

We hope that these insights encourage colleagues at other institutions to join our initiative against lacking reproducibility results in computer science research by using research artifacts that have become widely available in recent years.