

Replication in Computer Science: The Artifact Evaluation Committee Experience

Emery Berger
*College of Information
and Computer Sciences*

UMassAmherst

The Problem

For many years, some of us [...] have been concerned by the insufficient respect paid to the artifacts that back papers. We find it especially ironic that areas that are so centered on software, models, and specifications would not want to evaluate them as part of the paper review process, as well as archive them with the final paper.

Not examining artifacts enables everything from mere sloppiness to, in extreme cases, dishonesty. More subtly, it also imposes a subtle penalty on people who take the trouble to vigorously implement and test their ideas.

Shriram Krishnamurthi et al.

Artifact Evaluation Committee

- Began in 2011 in one conference in Software Engineering community (FSE)
 - artifact-eval.org
- Separate committee to evaluate *artifacts*
- Now spread to other SE conferences and all major Programming Languages conferences

What's an Artifact?

- Dataset
- Program
- Test suites
- etc.

Artifact Evaluation Committee

Review Criteria

- Consistent with the paper
 - Does the artifact substantiate and help to reproduce the claims in the paper?
- Complete
 - What is the fraction of the results that can be reproduced?
- Well documented
 - Does the artifact describe and demonstrate how to apply the presented method to a new input?
- Easy to reuse
 - How easy is it to reuse the provided artifact?

Artifact Evaluation Committee Process

- Committee separate from Program Committee
 - Receives papers *after* accepted, reviews *before* camera-ready
 - “Chinese Wall” between PC & AEC
 - Constituted from recommended PhD students
- Review content
 - Summary and contributions of the paper
 - Artifact packaging and reproducibility
 - Artifact implementation and usability
 - Overall assessment (*more next slide*)
 - On what platform/how was the artifact evaluated



Charlie Curtsinger



Emma Tosch



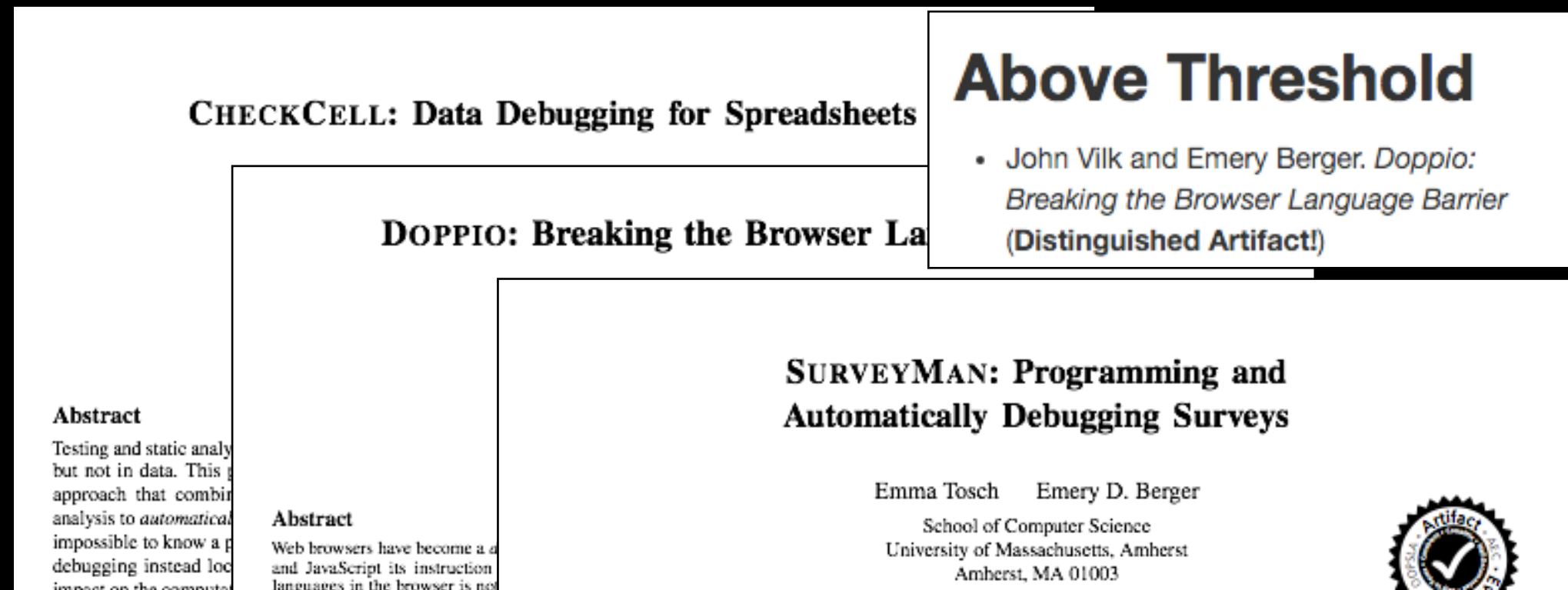
John Vilk



Dan Barowy

Artifact Evaluation Committee Process

- Assessment
 - Significantly exceeded expectations
 - Exceeded expectations
 - Met expectations
 - Fell below expectations
 - Significantly fell below expectations



Artifact Evaluation Committee Questions

- Is artifact evaluation appropriate for your area?
 - Journal vs. conference model...
- Should artifact evaluation be *required*?
 - For accepted papers?
 - For *all* papers?
 - Note: industry issues
- Artifacts are officially private: should they be public?
- Should artifact evaluation “have teeth”?
 - Should an accepted paper be rejected if artifact is below the bar / found to be dishonest?
- How does all this affect incentive systems?
 - Open source, public data, etc.
 - Links: artifact-eval.org, <http://evaluate.inf.usi.ch/artifacts>

