**Week 5: Review and Analyze Conference Venue Review**

Nate Bachmeier

April 7th, 2019

TIM-8101: Principals of Computer Science

Northcentral University

# Describe the types of papers presented/research conducted at the venue

## ISCE: Addresses the How

The International Conference on Software Engineering focuses on software engineering trends and explores the implications of those decisions. This is the premium conference where IEEE journal entries are discussed.

## MSR: Addresses the Which

The Mining Software Repositories focuses on analyzing and extracting metadata from software repositories. Many works center around GitHub and empirical studies of code hygiene.

## FSE: Addresses the What

The European Software Engineering Conference and Symposium on the Foundations of Software Engineering focuses on “recent innovations, trends, experiences and challenges in the field of software engineering (ESEC/FSE, 2018).” There are topic overlaps with ISCE and special attention to the application of software engineering.

## ASE: Addresses the Tedious

The Automated Software Engineering conference focuses on tooling and frameworks for simplifying engineering tasks. A large emphasis is placed on removing redundant operations.

## ISSTA: Addresses the Correctness

The International Symposium on Software Testing and Analysis focuses on software testing and best practices. The target audience is software development engineers in test (SDET) roles.

# Describe each track and types of papers that are submitted to each track

Every venue has a unique set of tracks that are available, but they tend to cluster into the same high-level themes.

## Technical, Research, and Journal-First Papers

The primary goal of this category is to present original and unpublished results that expand the knowledge of software engineering. Journal-First is a subset which provides a forum for recently published works that have not been widely demonstrated.

## Demonstrations, Showcases, Industry Use-cases and Artifacts

The primary goal of this category is to present concrete implementations of ideas that are being used in academia or industry experts. These tracks tend to focus on tooling, frameworks, and data sets built by the community.

## Doctoral Symposiums & New Ideas and Emerging Results (NIER)

The primary goal of this category is to share early investigation results as they pertain to ongoing research. Researchers can use the forum to have a dialog with the community experts and gain feedback on how to expand efforts.

## Challenges, Workshops and SE Education and Training (SEET)

The primary goal of this category is to provide hands on experience with a collection of tools and processes. Researchers will often perform studies using a specified data sets and their existing tools. This helps to have a consistent context and theme between the presentations.

# Describe Materials from ICSE

## What problems do they solve?

The selected materials were Magnus Frodigh’s Keynote, Code Review Comments Matter, and Are Code Examples Reliable.

The first contained a description of the vertical integrations that will be needed once 5G wireless becomes readily available. Frodigh describes Industrial IoT (IIoT) scenarios where a need for sensor networks will drive the need to innovate in data processing and storage technologies.

The other two focused of improving the quality of source code by addressing different aspects of the development process. As the efficiency of these steps are improved so is the productivity of the development team.

## What was their methodology?

The keynote approached the problem as an onion where one layer leads to the next challenge. For instance, he detailed the need for edge computing directly at the factory but also needing to blend into the cloud for cost efficiency. Then he continued to describe how these sensors will create tremendous volumes of data, which drives the need for improved AI algorithms.

The code review comments piece used natural language processing to determine the effectiveness of different language patterns. To determine the quality of online code examples, thousands of posts from Stack Overflow were parsed into an abstract syntax tree and their correctness measured against published best practices.

## What are future improvements?

Analyzing code review comments through natural language provides an effective way discover best practices across a large audience. The researchers approach sounds very specific to the English grammar system. It would be interesting to see how well it can be adapted to Chinese or German.

The analysis of online code examples concluded that 31% of the Java samples had one or more issues. These issues were largely grouped into error handling, control flow, and incorrect API usage. I disagree with their findings, as the intent of online examples is to demonstrate invocation of a function. Adding error handling and conditional execution reduces the clarify of the example. It would however be interesting to see more investigation into the incorrect API usage scenario.

## What are related efforts?

The code review analysis was innovative as previous work has focused on sentiment analysis. While that is useful in many scenarios it does not work in code reviews as business policies explicitly request neutral emotional tone.

The quality of examples on Stack Overflow have been investigated in previous research. What made this approach novel was the parsing of the snippets into a custom syntax pattern. This allowed for many variations of the same code to clustered and compared as a single unit.

## Why are these works important?

Frodigh provides a roadmap that details where network communication has been and where it will be going. Understanding the capabilities that will available in the enables us to plan those future scenarios today.

Engineering time is expensive and needs to be efficiently used by providing the best feedback and the right examples the first time. This is difficult for the developer as they might lack the expertise, however algorithms can detect these issues and flag the issue upfront.

# Describe Materials from MSR

## What problems do they solve?

The selected materials were Public Git Archive, Anatomy of Functionality Deletion, and A Study on Inappropriately Partitioned Commits. The first paper addressed the challenge of finding sufficiently large datasets for academia to perform Big Code analysis. The second paper investigated reasons that mobile applications remove entire features. The final paper looked at analyzing developer tasks when the task is persisted across multiple commits.

## What was their methodology?

The Public Git Archive attempts to make a snapshot of the most popular repositories on GitHub. Each instance of the snapshot is approximately 3TB in size and contains both the code and commit metadata. This is accomplished by listening to public events from GitHub and then scripting the git tool chain.

The researchers inspected commit messages to Android mobile manifest files across several hundred open source repositories. The manifest files were targeted as they declare all components exposed to the UI layer.

The final paper determined the likelihood that two commits are related by first discovering all methods touched by the commit. Then using a graph representation of the source tree, they calculate a distance between the method sets. If the distance is within a threshold then the commits are said to be related.

## What are future improvements?

To understand why a feature was deleted the researchers focused on commit messages text. However, they also reported that nearly 60% of all commits did not contain a message. Instead of focusing on an individual file perhaps the research should focus on the source tree structure itself.

The primary reasons functionality was removed were (1) it didn’t work, (2) negative user feedback, and (3) not compatible with the supported devices. Using telemetry and crash reports could help to empirically drive decisions on when to pull the plug.

## What are related efforts?

Bundling a collection of software repositories into a single artifact is not a new idea and has been released in the past. Those artifacts have focused on niche areas such as only Java code or only mobile applications. The Public Git Archive is unique because it contains the most frequently book marked irrespective of repository contents. That results in a wider range of examples and can be leveraged across a more diverse set of objectives.

## Why are these works important?

The

# Describe Materials from FSE

## What problems do they solve?

## What was their methodology?

## What are future improvements?

## What are related efforts?

## Why are these works important?

# Describe Materials from ASE

## What problems do they solve?

## What was their methodology?

## What are future improvements?

## What are related efforts?

## Why are these works important?

# Describe Materials from ISSTA

## What problems do they solve?

## What was their methodology?

## What are future improvements?

## What are related efforts?

## Why are these works important?