

A photograph of the iconic clock tower at Vanderbilt University, a tall red brick structure with two clock faces, set against a backdrop of autumn foliage and a cloudy sky. The tower is the central focus of the background image.

# Software Engineering: Research Project Selection and Presentations

---

Yu Huang

Vanderbilt University

[yu.huang@vanderbilt.edu](mailto:yu.huang@vanderbilt.edu)

# How to write a paper/research proposal?

- How can I get a research **idea**?
  - From your own experience
    - What do you want to change?
    - What is important to you?
  - ***No, I just cannot come up with any idea from my own experience.***
  - Final trick: read a paper, lift the assumption(s) in it

ICSE, 2019

## TSE, 2012 GenProg: A Generic Method for Automatic Software Repair

Claire Le Goues, ThanhVu Nguyen, Stephanie Forrest, *Senior Member, IEEE*, and Westley Weimer

**Abstract**—This paper describes GenProg, an automated method for repairing defects in off-the-shelf, legacy programs without formal specifications, program annotations, or special coding practices. GenProg uses an extended form of genetic programming to evolve a program variant that retains required functionality but is not susceptible to a given defect, using existing test suites to encode both the defect and required functionality. Structural differencing algorithms and delta debugging reduce the difference between this variant and the original program to a minimal repair. We describe the algorithm and report experimental results of its success on 16 programs totaling 1.25 M lines of C code and 120K lines of module code, spanning eight classes of defects, in 357 seconds, on average. We analyze the generated repairs qualitatively and quantitatively to demonstrate that the process efficiently produces evolved programs that repair the defect, are not fragile input memorizations, and do not lead to serious degradation in functionality.

**Index Terms**—Automatic programming, corrections, testing and debugging.

## Harnessing Evolution for Multi-Hunk Program Repair

Seemanta Saha<sup>†</sup>  
University of California Santa Barbara  
Email: seemantasaha@cs.ucsb.edu

Ripon K. Saha, Mukul R. Prasad  
Fujitsu Laboratories of America, Inc.  
{rsaha, mukul}@us.fujitsu.com

# SE Project Selection

- Overview of what you have learned
- Think about what you have been working on
  - Your PhD work (if you are PhD students)
  - What you want to try (if you are not PhD students)
- How to combine them?
  - Using SE to improve X vs. Using X to improve SE

# SE Project Selection

- What is the purpose of the project?
  - **Learn new knowledge?**
    - Empirical studies – Who cares? So what?
  - **Develop new techniques?**
    - Tools, algorithms, models – Who cares? So what?
- Types
  - Human subject research
  - System research

# If you need data

- Construct your own dataset
  - GitHub
  - From your domain
- Use an existing dataset

# If you need data: what I can provide you

- Human data
  - fMRI data (including resting state, demographic info)
    - Code writing ~30
    - Code reading ~30
    - Code review ~30
    - Data structure manipulation and 3D object rotation ~30
  - fNIRS data
    - Data structure manipulation and 3D object rotation ~30
  - Eye tracking data

# Eye-tracking dataset



|   | # Devs. | Task Type                            | PLs  | Task Size  |
|---|---------|--------------------------------------|--|--|
| 1 | 27      | Code Review and Code Summarization   | Java   | 162 unique Java methods  |
| 2 | 49      | Code Comprehension                   | C++, RACKET                                    | 24 unique C++ tasks and 12 unique RACKET tasks                 |
| 3 | 30      | Debugging                            | C++  | 10 unique C++ tasks  |
| 4 | 37      | Code Review                          | C, C++   | 60 C/C++ tasks   |
| 5 | 22      | Code Comprehension and Summarization | Java, XML, HTML, Scala, YAML, JavaScript, etc. | Multi-language Projects with 3500 files and 80k+ lines of code |

# If you need models to start with: what I can provide you

- All the models/tools in a paper that I am in
- Eye trackers, UIs, analysis pipeline
- Ask me if there is such a tool you can use
  - E.g., webcam eye tracking, static analyzer, dynamic analyzer, APR tools
- Just post on Piazza



# Preliminary results

- Empirical study
  - Selected dataset, tools, etc. – what are you going to study?
  - What are your hypotheses?
    - Analyze one question in your proposal
    - Or demonstrate the “tendency” “facts” in partial data
  - If human study:
    - Collect data from a small set of participants as “pilot results” (e.g.,  $n < 5$ )
    - Pilot study does not require IRB

# Research Project: Examples for Discussion

- Developing testing framework for AR/VR
  - Why traditional SE testing doesn't work?
  - What are the unique challenges to test AR/VR?
  - What can you do to overcome these challenges?
  - What dataset can you use?
  - How would you implement this idea?

# Research Project: Examples for Discussion

- Investigating Human-AI Interactions in CS education
  - Which level of education? CS1?
  - Which process do you want to target?
  - What are your hypotheses? Do you want to include some education theory?
  - How are you going to measure it?

# Research Project: Examples for Discussion

- Develop an AI model for SE tasks
  - Which SE task?
  - What model? Why is it good/novel for this task? Do people care about such a model?
  - Which dataset?
  - What is the domain knowledge being used?

# Research themes (from Lecture 1)

- How to combine SE with X?
  - X can be your own research direction
    - PL
    - HCI
    - Hardware
    - Security
    - Cyber Physical System
    - Social Science
    - Psychology
    - Neuroscience
    - Education
    - ...

# How to make a good presentation?

- Two philosophies
  - **After listening to my presentation...**
    - **My audience do not need to read the paper**
    - **My audience really want to read the paper**

# How to make a good presentation?

- Two philosophies
  - **After listening to my presentation...**
    - **My audience do not need to read the paper**
    - **My audience really want to read the paper**
- The first thing you need to know:
  - **Who are your audience??**
    - Practice: Explain X to your ...
      1. Grandparents
      2. CS researchers but not in your domain
      3. CS researchers in your domain

# How to make a good presentation?

- The first thing you need to know:
  - **Who are your audience??**
    - Practice: Explain X to your ...
      1. Grandparents
      2. CS researchers but not in your domain
      3. CS researchers in your domain
- Common mistakes
  - Assuming your audience already know the background
  - No establishment on the motivation: **Who cares? Why should I care?**
  - Structure is not clear (use outline, but I don't use it before Motivation)
  - No transitions between slides and sections
  - Spending too much time on "greek letters"
  - No highlight on the results: If your audience **can only remember one thing**, what's it?
  - No **slide numbers...!!!**
  - Using a **"Thank You"-only (or "Q&A" only)** slide for Q&A