

What and how to change about
testing framework that is not
giving actionable results?

Raimond Sinivee, @RaimondSinivee

04.06.2015 @ntd2015

Agenda

- Terminology
- Little bit about the problem
- Patterns

What makes UI tests so fragile?

Terminology

1. Tools, scripts, harness, framework, helpers but not automation
2. There are no best practices, there are heuristics

What makes it fishy?

- More inaccurate results than correct ones
- Solution needs constant work to keep it running
- Performance is not what is supposed to be

Levels of difficulty



(Getting your toes wet)



(Like a fish in the sea)



(Deep dive)

Why, oh why - The Test Pyramid

- Amount of tests
- Lines of Code – 10 vs. 100000000
- Time to execute – ms vs. min
- Time to understand failures – sec vs. hours

What kind of other differences
come to your mind?

Pattern 1: Forgiveness

- Negative
 - Introducing retries – lost performance
 - Increasing timeouts – lost performance
 - Reducing asserts – lost test coverage
- Positive
 - Correct results – increased credibility

Pattern 2: What's the maximum possible success rate?

- A tool can't be more stable than the technology, software or infrastructure which it relies on
- Count in the dependencies SLA

100% success rate == no new info*

- Failed means that it could not find problems
- Failed means that I can't tell anything based on these results

* Except when running tests after refactoring

Pass/fail vs. success rate

- Make a scientific experiment
- Probabilistic results by repeating tests
- Trends over time instead of a moment

Pattern 3: Results parsing

- Log parsing service
- Rules for known errors
- Logs parsed based on rules
- Error ratio in reproducibility in bugs
- Don't need to check known failures logs*

* Sometimes it's still good to double check the tool

Lack of tools support remains

- Usual CI tools don't support other than forgiveness
- Used in house built tool

Summary

- A tool can't be more successful than what it depends on
- Use and present probabilistic results instead running after 100% accurate pass/fail
- Parse and analyze results with tools to scale down the amount of results to human level

- Tool purpose is to find problems and provide relevant and trustworthy information