SBFT Tool Competition 2023 - Java Test Case Generation Track

Gunel Jahangirova

King's College London, United Kingdom

Valerio Terragni

University of Auckland, New Zealand

14th of May, 2023

History of Java Tool Competition

Round	Year	Venue	Tools	#CUTs / #Projects	#Participants (+ baseline)	Statistical Tests
1	2013	ICST	Cobertura + Javalanche	77 / 5	2	✗
2	2014	FITTEST	JaCoCo + PITest	63 / 9	4	✗
3	2015	SBST	JaCoCo + PITest	63 / 9	8	✗
4	2016	SBST	Defects4J	68 / 5	4	✗
5	2017	SBST	JaCoCo + Our Env.	69 / 8	2 (+ 2)	✓
6	2018	SBST	JaCoCo + Our Env.	59 / 7	2 (+ 2)	✓ + combined analysis
7	2019	SBST	JaCoCo + Our Env.	69 / 8	2 (+ 2)	✓ + combined analysis
8	2020	SBST	JaCoCo + Our Env. + Docker	69 / 8	1 (+ 1)	✓ + combined analysis
9	2021	SBST	JaCoCo + Juge + Docker	98 / 6	4 (+ 1)	✓ + combined analysis
10	2022	SBST	JaCoCo + Juge + Docker	65 / 4	5 (+ 1)	✓ + combined analysis

The 11th edition: what is new?

5 Participants

UtBot-Concolic

UtBot-Fuzzer

Kex-Concolic

Kex-Symbolic

100 CUTs
from 5 projects

Human study to measure test case understandability

Benchmark Selection

5 Subject Systems

Apache Commons Collections

JSoup

Ta4J

Spatial4J

Threeten-extra

1,145 classes

Filtering

1 at least 2 branches

2 at least one method with cyclomatic complexity > 5

3 is testable: test cases generated by Randoop with in 10 sec

298 classes

Sampling

Random

100 classes

Calculating Structural Coverage Criteria

Classes under test

100 classes

Search budgets

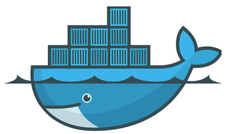
30 seconds

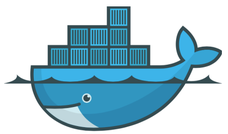
2 minutes

Repetitions

10 repetitions

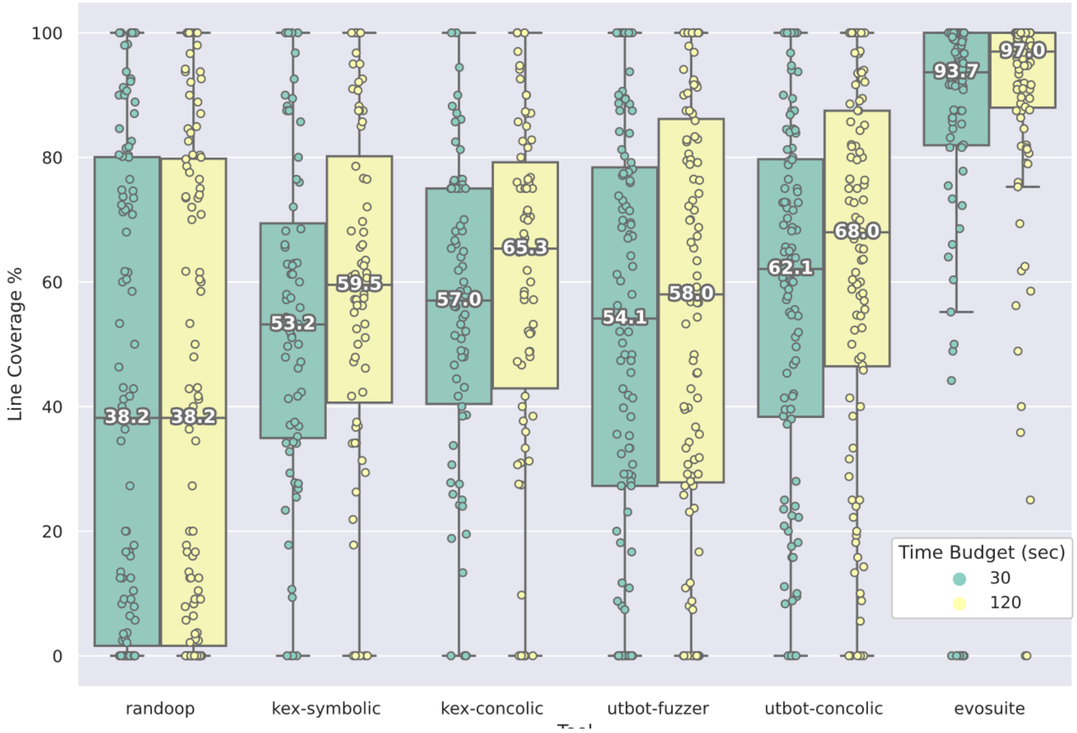
Execution environment: JUGE

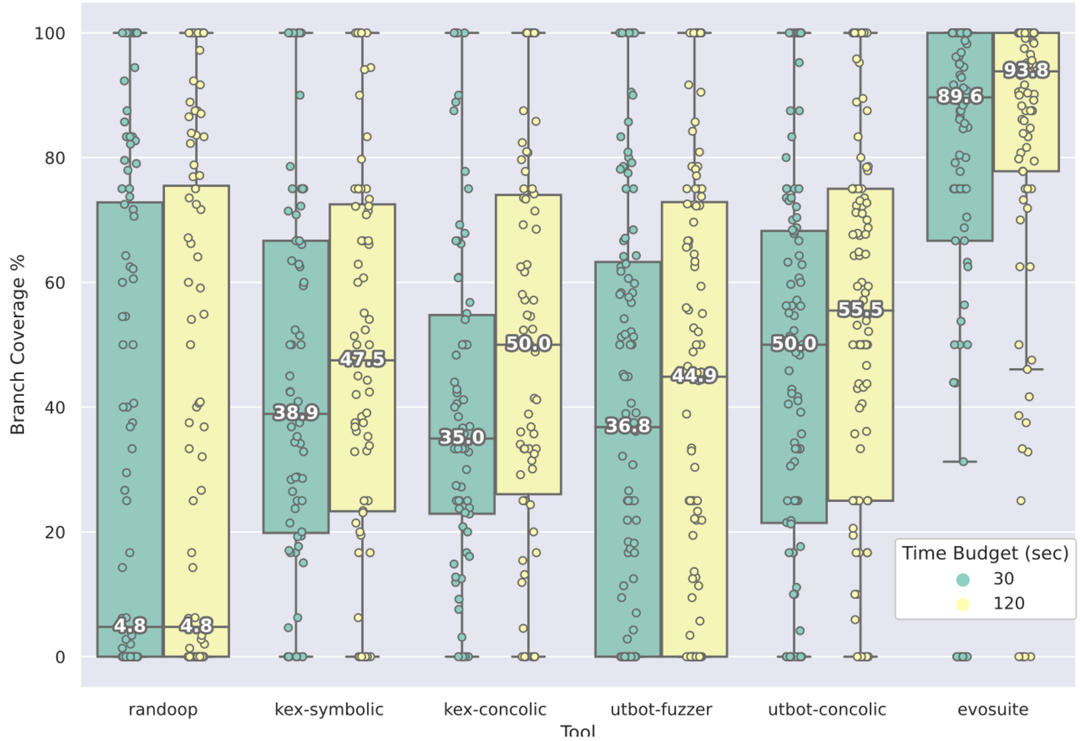




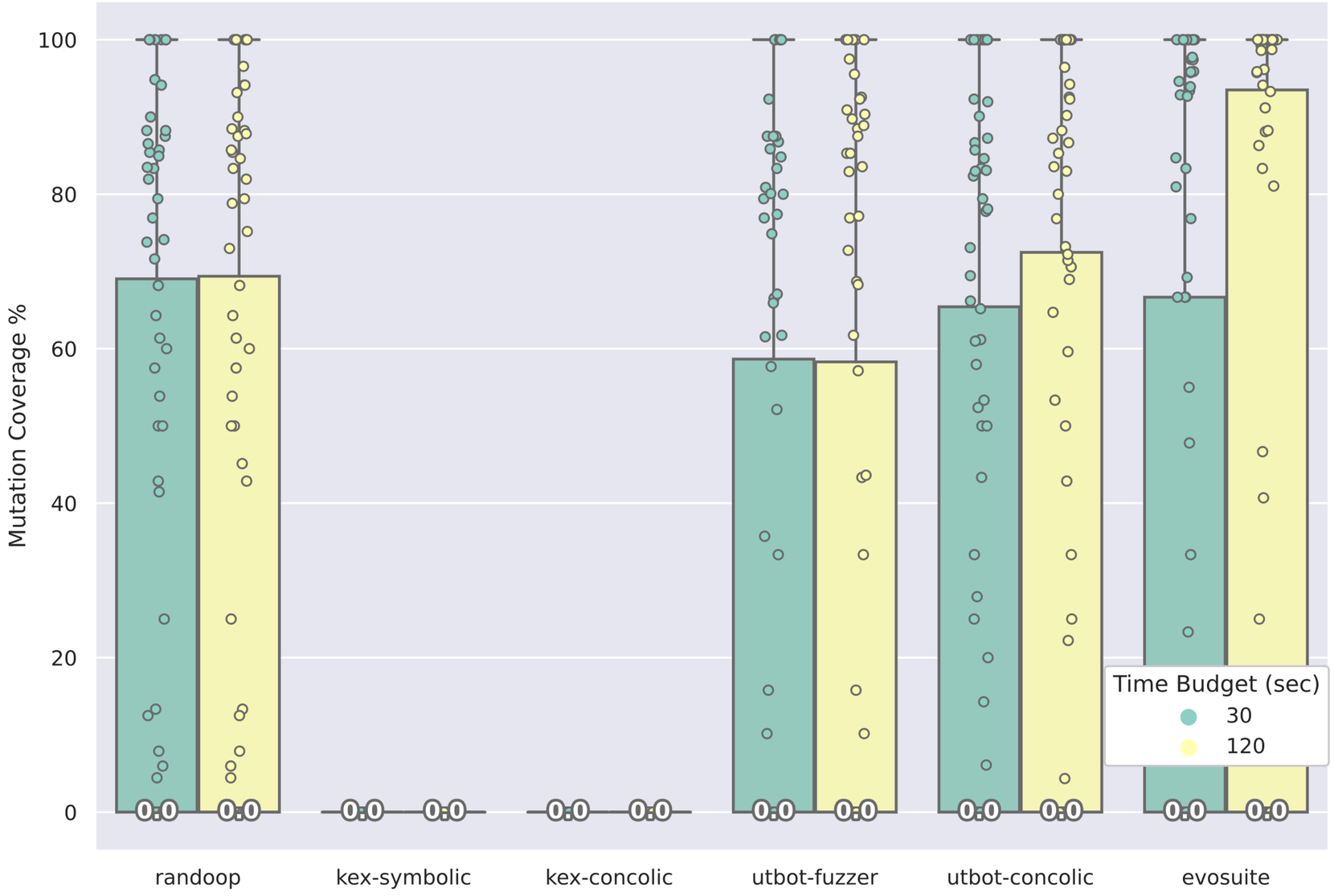
https://github.com/JUnitContest/junitcontest

Results: Line and Branch Coverage





Results: Mutation Coverage



Test Case Understandability: Subject Selection

CUT Selection	Test Case Selection
LMap	Random
Months	
DUtils	

The Task for Human Participants

1. Three Java classes their source codes provided
2. Five test cases each from one of the five competing tools
3. Questions asking the participant to rank the test cases for each class in terms of understandability from the most understandable to the least understandable
4. Questions asking the participant to describe in natural text the behaviour of the test case they have rated the most understandable
5. Questions asking the participant to explain why the test case they ranked the least understandable is hard to understand

1 hour task

Participant Recruitment

Requirements: Knowledge of Java and JUnit

Payment: 10 GBP

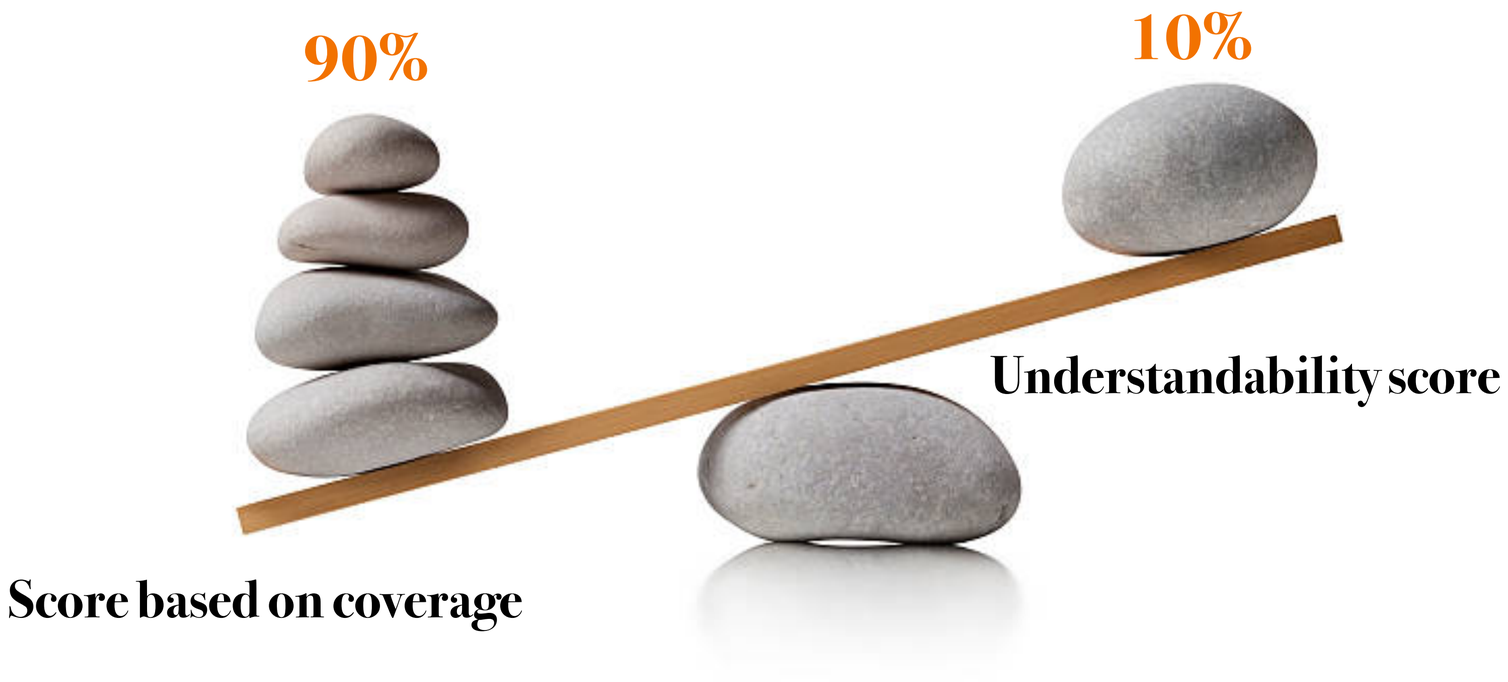
Overall hired: 30 participants

Data kept from: 13 participants

Understandability Results

Tool	LMap	Months	DUtils	Average
EvoSuite	2.38	2.08	2.23	2.23
Utbot-concolic	1.92	2.23	2.23	2.13
Utbot-fuzzer	3.54	2.77	2.69	3.00
Kex-symbolic	3.62	4.00	4.23	3.95
Kex-concolic	3.54	3.92	3.62	3.69

Final Score



Final Results

Tool	CoverageR	UnderstandabilityR	OverallR
EvoSuite	1.79	2.23	1.83
Utbot-concolic	2.61	2.13	2.56
Utbot-fuzzer	3.76	3.00	3.68
Kex-symbolic	4.995	3.95	4.89
Kex-concolic	3.95	3.69	3.92

The Future Ahead

1. Fixing some issues with JaCoCo
2. Improving the understandability study
   - 2.1 More meaningful criteria to select test cases
   - 2.2 Stricter criteria to select participants

The Future Ahead

1. Fixing some issues with JaCoCo
2. Improving the understandability study
   - 2.1 More meaningful criteria to select test cases
   - 2.2 Stricter criteria to select participants

SBFT 2024 - Java Competition Tool Chairs

Gunel Jahangirova

King’s College London, United Kingdom

Valerio Terragni

University of Auckland, New Zealand