Report: Stage 1 Experimentation and research

In this section, we evaluate the performance of a grammar based test suite by comparing it against a baseline: a manually constrcuted suite. We evaluate the performance of each test suite using the following indicators :

1. Time taken to construct the test suite
2. Code Coverage

Code coverage

Instruction Coverage: this metric measures the % of java bytecode instruction executed.

Branch Coverage: also known as decision coverage, this metric measured the % of branch (typically if/switch statments) exercised by the test suite. Exceptions are not included.

Experimental Setup

The test suites were constructed using a blackbox approach
For the baseline, the test were constructed by simply looking at the api
For the grammar suite, The context free grammars were written up then used to construct the test suites

<library api here>
<Grammar here>

Test Data

For the purpose of this investigation , libraries of various sizes and complexities were chosen

Library	Number of Lines	Number of Methods	Number of Instructions (java bytecode)
Strman	358	148	2049
TrieSET (princeton)	77	16	416
StringWriter	32	13	119

Results

Suite	Library	Time	Instruction Coverage/%	Branch Coverage/%
Baseline	Strman	10 hours	86%	67%
	TrieSET	1hour 20 min	96%	88%
	StringWriter	4 hours	95%	75%
Grammar	Strman	5 hours	91%	84%
	TrieSET	1 hour	98%	95%
	StringWriter	1 hour 30min	98%	75%

Commentary:

took less time to construct suites using grammar for all libraries; took roughly 15hours and 20 min for basline and 7hours and 30min (almost half the time of the baseline)

Instruction coverage increased in all cases for the grammar suite
Branch coverage increased in all cases for grammar suite, more significant for larger libraries (strman)