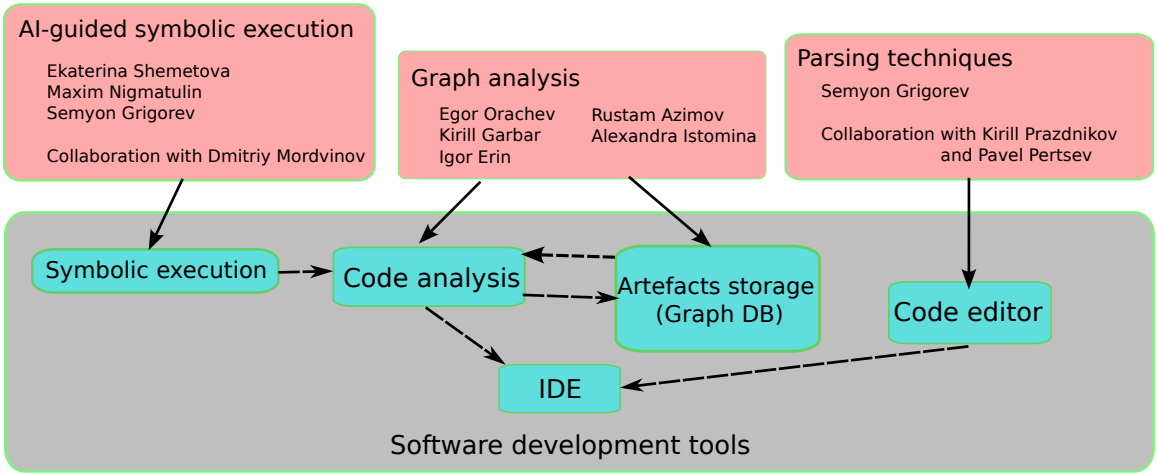
Formal Language Driven Data Analysis Research Group Report

Semyon Grigorev

Saint Petersburg State University

April 26, 2023

Research Landscape



AI-Guided Symbolic Execution

- ✔ Basic infrastructure for training developed and implemented
  - Wrapper for SVM to convert it to server
  - Python client — AI agent to training
  - Basic manipulation with neural networks
- ✔ Basic dataset for train and validation/test
- ✔ First attempts to train AI agent: workflow works fine (but agent too stupid to learn)
- ⚙ Dataset extension
- ⚙ GNN improvement and pretraining
- ⚙ Performance tuning
- ⌛ First version of AI agent which guide SVM on par with algorithmic strategies

Parsing Techniques

- ✔ Partial parsing to improve highlighting speed for huge files

File	Size	Parsing time (ms)			
		Web		Desktop	
		Partial	Full	Partial	Full
EUC_TU_OLD.java	2302Kb	1386	11802	417	2271
JavaParser.java	428Kb	666	6225	86	1175
TestBigObj.java	1539Kb	2324	3256	356	664
INDIFY_Test.Java	927Kb	1162	7756	206	1925

Parsing Techniques

- ✔ Partial parsing to improve highlighting speed for huge files

File	Size	Parsing time (ms)			
		Web		Desktop	
		Partial	Full	Partial	Full
EUC_TU_OLD.java	2302Kb	1386	11802	417	2271
JavaParser.java	428Kb	666	6225	86	1175
TestBigObj.java	1539Kb	2324	3256	356	664
INDIFY_Test.Java	927Kb	1162	7756	206	1925

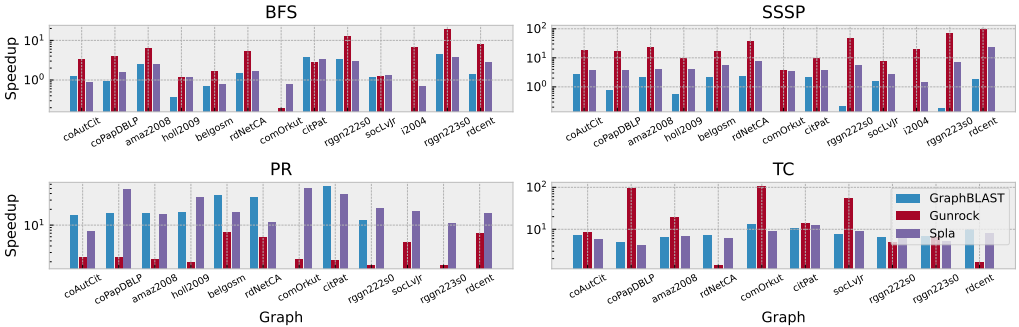
- ⚙ Naïve incremental parsing
- ⚙ Error recovery mechanism
- ⌛ Advanced incremental parsing

Graph Analysis

- ✔ Datalog-based static code analysis prototype implemented
- ⚙ Datalog-based static code analysis evaluation

Graph Analysis

- ✔ Datalog-based static code analysis prototype implemented
- ⚙ Datalog-based static code analysis evaluation
- ✔ Spla — vendor-agnostic sparse linear algebra for graph analysis on GPGPU
  - OpenCL for GPU
  - Intel, AND, Nvidia GPGs evaluated



- ⌛ Performance tuning, more algorithms, . . .

New Members (from June)

- One in AI-based symbolic execution
- One in graph analysis