

Overview

Motivation:

- Software repositories evolve, and new bug reports emerges.
- Current information retrieval-based bug localization (IRBL) tools usually need to re-construct the model to adapt to changes in bug reports and codebases.
- Some simple IR models (e.g., VSM [1]) can be updated incrementally while no such work is about advance models (e.g. rVSM in BugLocator [2]).
- No open-source bug localization tool adopts incremental update strategies.

Contribution:

We present **IncBL**, the first open-source bug localization tool updating models incrementally. It reduces running time by 77.79% on average while maintaining a competitive level of accuracy. **IncBL** has been integrated as a GitHub App and can also be deployed locally.

Pipeline and method:

IncBL is based on **BugLocator** [2]. The workflow is as follows:

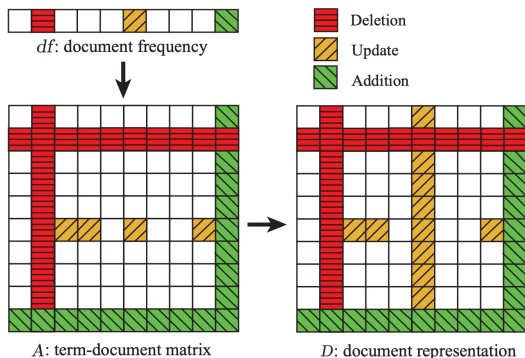
- **Step 1.** Processing code files: pre-processing and vectorizing texts using Vector Space Mode (VSM).
- **Step 2.** Processing bug reports: pre-processing texts and computing the similarity (*SimiScore*) between new and past fixed bug reports.
- **Step 3.** Localizing buggy files: computing the similarity between code files and bug reports, then combining it with *SimiScore* to find relevant buggy files.

Incremental Updates

The VSMs can be incrementally updated by **deletion/update/addition operations in term-document matrix and document frequency matrix**. The new *df* and *idf* can be computed by:

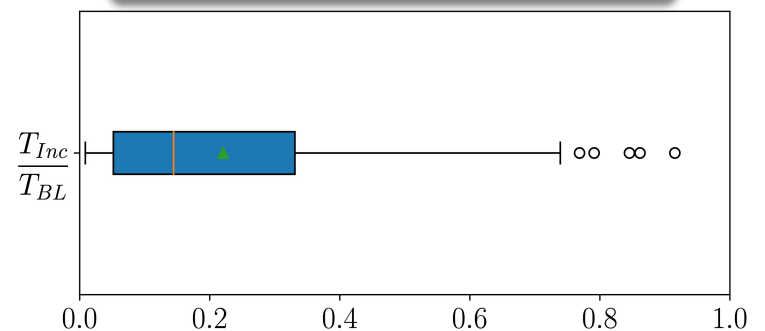
$$df^{new}(w) = df^{old}(w) + [sign(A_m^{new}(w)) - sign(A_m^{old}(w))]$$

$$idf^{new}(w) = idf^{old}(w) + \log\left(\frac{M + \Delta M}{M}\right)$$



This figure illustrates how term-document matrix (*A*), document frequency (*df*) and document representation (*D*) can be updated incrementally.

Experiment Results

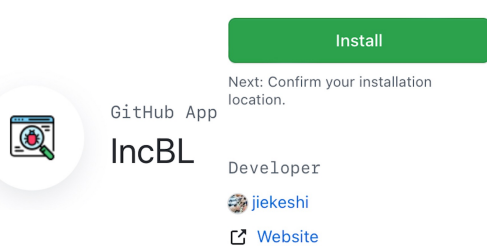


This box-plot shows the ratio of IncBL running time over the running time of BugLocator.

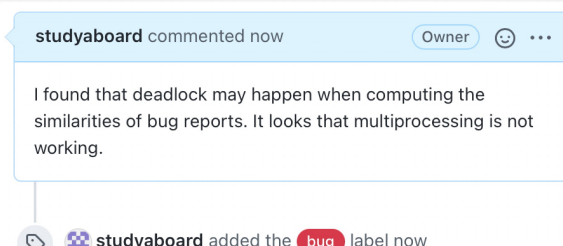
	BugLocator	IncBL
Running Time	1x	4.50x (on average)
MRR	0.328	0.331

Evaluation results on Bugzbook [3] dataset.

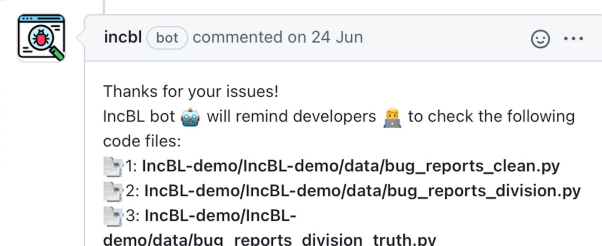
Usage Instruction



Step 1. Install GitHub App.



Step 2. Issue a bug report.



Step 3. Get lists of potentially buggy files.

Reference

- [1] Rao et al. "An incremental update framework for efficient retrieval from software libraries for bug localization." *WCRE 2013*.
- [2] Zhou et al. "Where should the bugs be fixed? more accurate information retrieval-based bug localization based on bug reports." *ICSE 2012*.
- [3] Akbar et al. "A large-scale comparative evaluation of IR-based tools for bug localization." *MSR 2020*.