Decoding the Issue Resolution Process & Identifying Solution-Related Content in Issues

Antu Saha, Md. Mehedi Hasan Sun, and Oscar Chaparro

Problem Definition & Overview

Problem: While the Issue Resolution Process is well-defined in literature, its practical implementation by devs is unclear

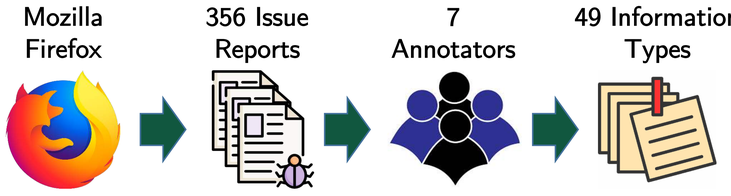
Solution: Qualitatively analyze issue report discussions to

- Understand practical issue-resolution process
- Identify the solution-related contents

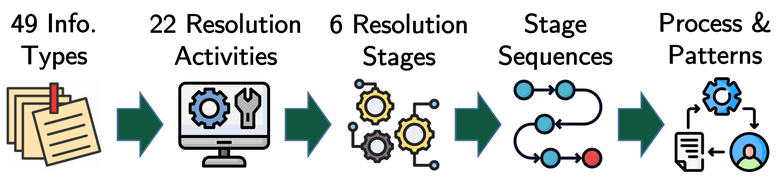
Usefulness: Assess the practical issue resolution process, estimate the developer's effort for issue resolution, and automatic identification of solution contents

Issue Discussion Investigation

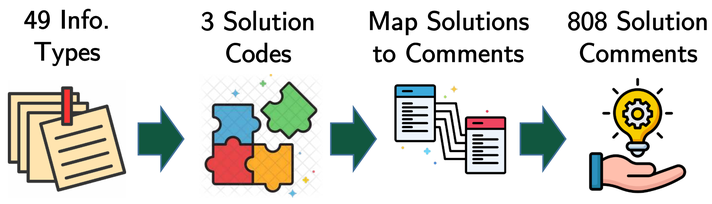
Goal: Qualitatively analyze the developers' discussion to identify the information types discussed in issue reports



Issue Resolution Process & Pattern Derivation



Solution Identification Data Preparation

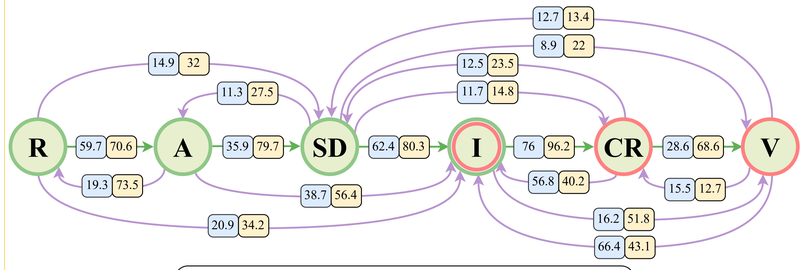


Issue Resolution Process & Patterns

Issue Resolution Stages

Reproduction (R)	Analysis (A)	Solution Design (SD)
47 (13.2%)	134 (37.6%)	150 (42.1%)
Implementation (I)	**Code Review (CR)**	**Verification (V)**
328 (92.1%)	264 (74.2%)	146 (41.0%)

Issue Resolution Process



Issue Resolution Patterns

I,CR,I?	A,I,(I|CR|V)?	(I,(CR|V))+
64 (Simple)	32 (Simple)	28 (Complex)
SD,I,CR,(I|V)?	**A,SD,I,(I|CR|V)+**	**I**
24 (Simple)	22 (Complex)	21 (Simple)

A? → A is optional; (A|B) → either A/B/both stages appear

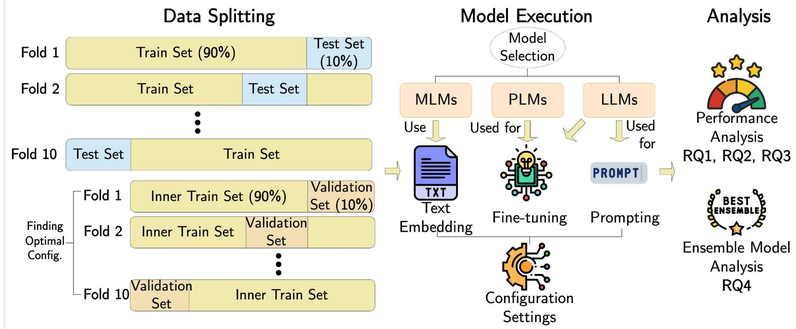
(A,B,...,Z)+ → A, B, ..., and Z appear repetitively

Major Findings

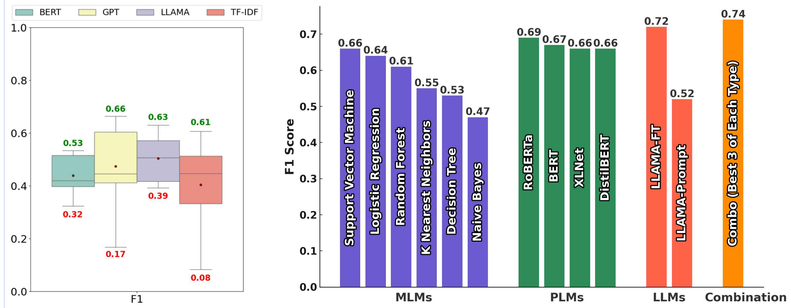
- Firefox's issue resolution follows an iterative process that deviates from the (theoretical) linear process
- Developers go back and forth from one stage to another as needed to solve the issues
- The 47 identified issue resolution patterns indicate that solving issues at Firefox is done in a wide variety of ways
- 18 patterns are recurrently found in 80.6% of Firefox issues

Automatic Solution Identification

Study Methodology



Study Results



Major Findings

- Among six MLMs, SVM achieved the highest classification performance (F1: 0.66)
- RF and KNN showed significant improvements from data balancing (F1: Increased by 25.5% and 282.4%)
- LM embeddings (i.e., GPT, LLAMA, and BERT) outperformed traditional TF-IDF encodings across MLMs
- Among four PLMs, RoBERTa achieved the highest classification performance (F1: 0.69)
- LLM benefitted more from fine-tuning than from prompting (F1: 0.72 vs. 0.52)
- Combining the top models from each model type achieved the best overall performance (F1: 0.74)