



CoMRAT: Commit Message Rationale Analysis Tool

Mouna Dhaouadi, Bentley Oakes, Michalis Famelis

Rationale in Commit Messages

Context :

For a decision, software developer has
rationale: 'Why'

Rationale is **useful** information to :

- understand the system
- learn from mistakes
- reuse solutions
- avoid conflicts

Research Question (PhD) :

1. How can we **understand** developers' rationale in **commit messages**?
2. How can we **help** developer's **write better commit messages**?

Rationale challenges:

- Implicit or unrecorded
- Ambiguous language
- Non-native writers
- Need technical understanding

The mm/slob.c module

Resulting dataset:

% preview of the labelled dataset:

Number of commits: 146

Number of sentences: 833

Distribution

Decision only sentences: 233

Rationale only sentences: 162

Decision & Rationale sentences: 252

No Decision and No Rationale sentences: 186

Word Clouds



Rationale Presence

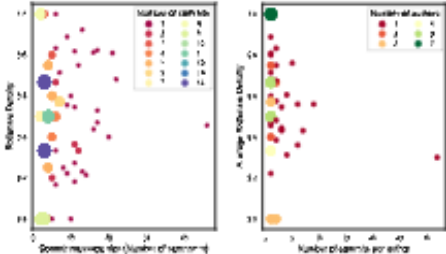
Total Number of commits: 146

Number of commits that contain rationale: 124

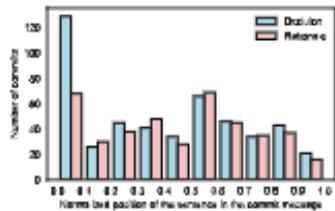
Rationale Percentage: 84.93%

Average Rationale Density: 0.56

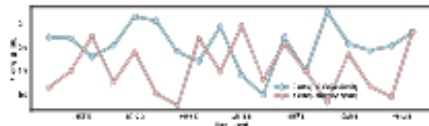
Rationale Factors



Commit Message Structure



Rationale Evolution



CoMRAT: Commit Message Rationale Analysis Tool

<https://rationale-analyses-tool.streamlit.app/>

Module Analyzer (Retrospective Analysis)

- **Presence** of rationale
- **Factors** impacting rationale
- **Evolution** of rationale over time
- **Structure** of commit messages

Commit Message Analyzer (Prospective Analysis)

Commit Message Rationale Analyses

Enter your commit message

mm: make function get_attribute() public.

I changed the visibility of function get_attribute().
Now, bug number #333 is fixed.
Debbuged-by: Mouna

Start Commit Analysis

	message_preprocessed	Decision	Rationale
0	mm: make function get_attribute() public.	1	0
0	I changed the visibility of function get_attribute()	1	0
0	Now, bug number #333 is fixed.	0	1

Number of sentences: 3

Your commit message's rationale density: 0.33

Your commit message's decision density: 0.67

⚠ Low Rationale! Please add justification for your changes.

Approach

Rationale Dataset and Analysis for the Commit Messages of the Linux Kernel Out-of-Memory Killer (ICPC'24)



Dataset Creation:

- 404 commit messages
- 3 labels: “Decision”, “Rationale”, “Supporting Facts”

Rationale Analysis:

- **Presence** of rationale
- **Factors** impacting rationale
- **Evolution** of rationale over time
- **Structure** of commit messages

Automated Extraction and Analysis of Developer Rationale in Open Source Software (FSE'25)

• Generalizable classifiers:

- Two BiLSTM classifiers [Decision, Rationale]
 - Trained on OOM-Killer dataset
 - Validated on:
 - Two other Linux modules
 - Five other open-source projects



Future Work: **A Survey**

About: Rationale in Commit Messages

Target group: Software developers and project maintainers

Duration: 10 minutes



<https://sites.google.com/view/rationale-in-commits/survey>