

### MINERAÇÃO DE REPOSITÓRIOS DE SOFTWARE COM A BIBLIOTECA RIT PARA INVESTIGAR A INFLUÊNCIA DOS COMENTÁRIOS NO PROCESSO DE CONTROLE DE MUDANÇAS

Fábio Trindade Ramos (fabio.ramos@ufv.br) Gláucia Braga e Silva (glaucia@ufv.br) Matheus Kauan Souza (matheus.k.souza@ufv.br)  
Comunicações em Issue Tracking, Mineração de Textos, Similaridade de Cossenos  
Área Temática: Ciência da Computação - Engenharia de Software  
Categoria: Pesquisa

#### Introdução

A comunicação constitui um dos principais gargalos para se garantir bons resultados em projetos de software. Nesse contexto, sistemas de *issue tracking*, presentes em repositórios de *software*, como o GitHub, desempenham um papel fundamental ao armazenar informações acerca das comunicações. Esses sistemas permitem que desenvolvedores relatem uma *issue* (problema, melhoria ou novo requisito) e discutam sobre sua solução.

#### Objetivos

Este trabalho objetiva avaliar o comportamento e a influência da relevância temática dos comentários, calculada pela Biblioteca RIT, em repositórios *open-source* do GitHub com alta atividade em *issue tracking*. O estudo foi guiado pelas seguintes questões de pesquisa:

- Q01: Existe relação entre a relevância temática dos comentários e o tempo de resolução?
- Q02: Existe diferença no tempo de resolução entre as *issues* que foram criadas, comentadas ou atribuídas por/para top-devs?

#### Materiais e Métodos

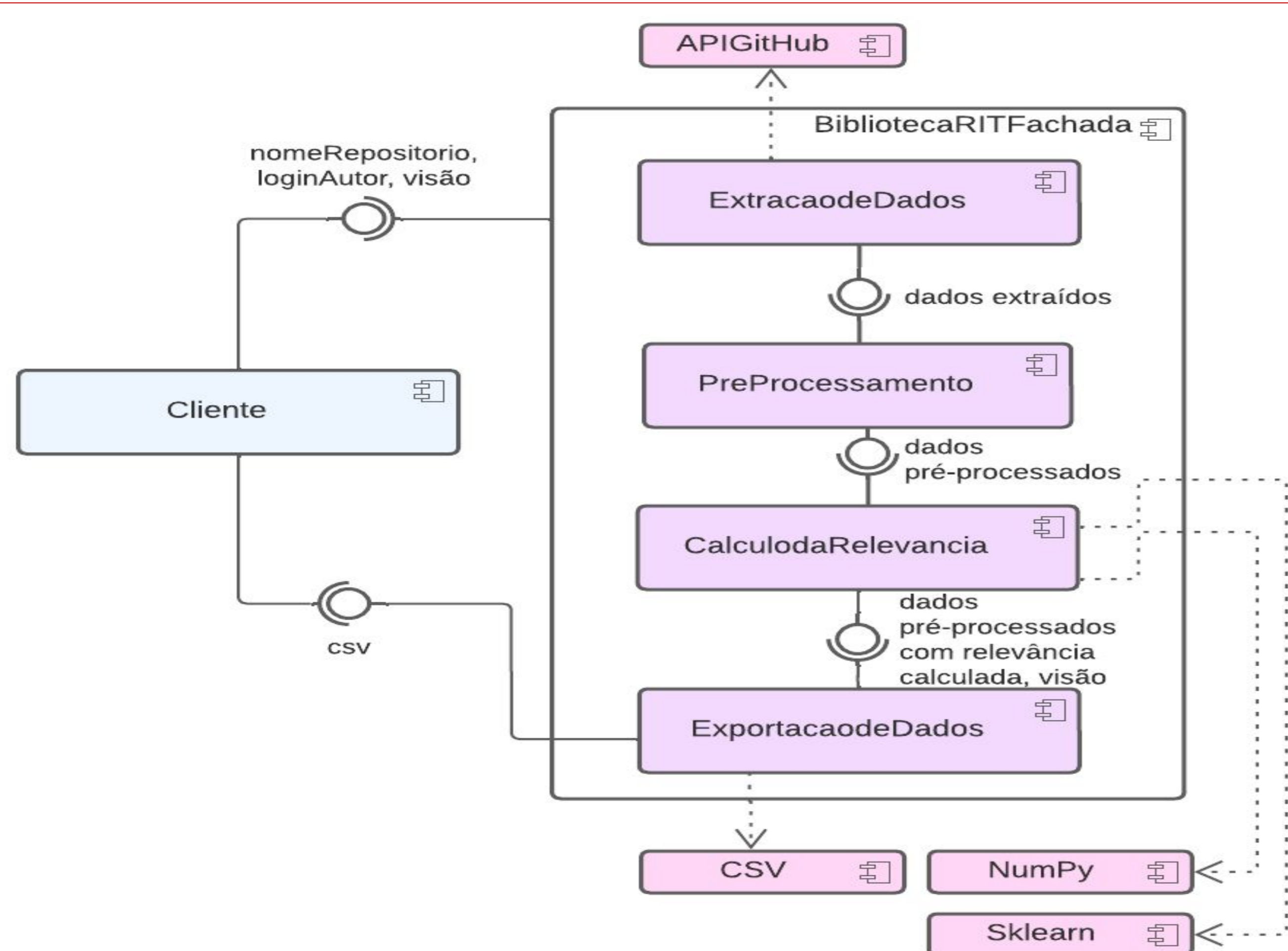


Figura 1 - Diagrama UML de Componentes da Biblioteca RIT

- API REST do GitHub;
- Ruby on Rails;
- Métricas: Relevância temática; Tempo de resolução; Presença de um top-dev em uma *issue* ao criá-la, comentá-la ou ao ser atribuído a ela.

#### Resultados e Discussões

	Issues	PRs	Total
Repositórios	-	-	116
Issues totais (incluindo PRs)	21.859	38.705	60.564
Comentários	107.130	121.299	228.439
Desenvolvedores	16.300	8.488	20.899

Figura 2 - Visão Geral dos Dados Obtidos

- Bom desempenho da RIT no cálculo da relevância temática de milhares de comentários em menos de 12 horas.
- *Issues* criadas por desenvolvedores que fazem comentários mais relevantes (top-devs) têm resolução mais rápida. Já *issues* atribuídas a eles tiveram maiores tempos de resolução.
- A análise da relevância temática não revelou uma relação direta com o tempo de resolução → requer análises adicionais, que considerem outros parâmetros (ex. complexidade e prioridade das *issues*)

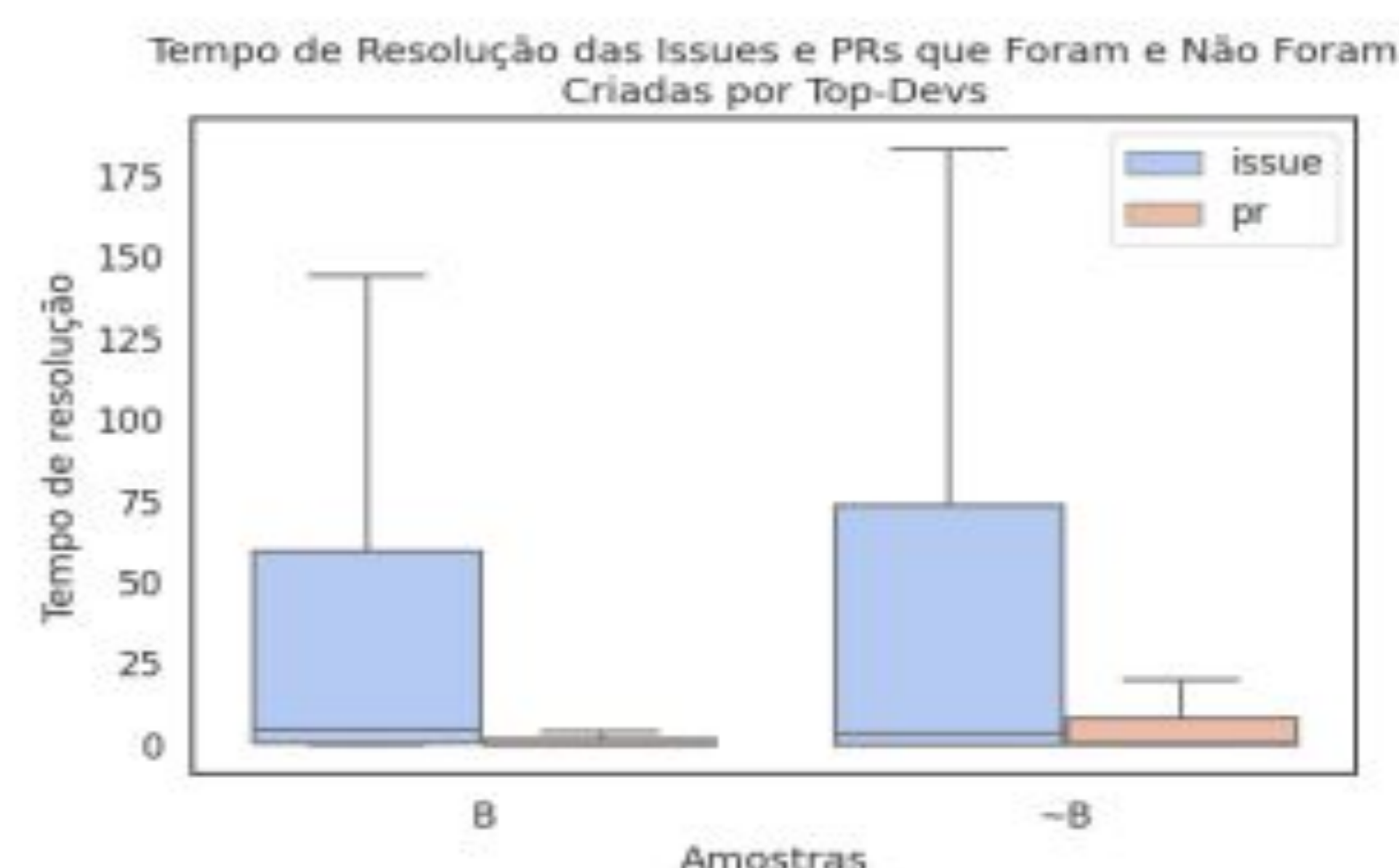


Figura 3 - BoxPlot Sem Outliers do Tempo de Resolução das Issues criadas e não criadas por top-devs

#### Conclusões

Os achados do estudo revelaram alguns indicadores de sobrecarga nos top-devs, uma vez que *issues* atribuídas a eles tiveram maiores tempos de resolução, levando em consideração que apesar de serem poucos, esses mesmos desenvolvedores estavam presentes em boa parte das *issues*. Além disso, a biblioteca, após as melhorias, mostrou-se poderosa para minerar um grande volume de dados.

#### Agradecimentos

Agradecimentos à UFV, ao Nupessc e ao CNPq, pela bolsa concedida.