



<b>Evento</b>	Salão UFRGS 2020: SIC - XXXII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2020
<b>Local</b>	Virtual
<b>Título</b>	Exploração de Técnicas de Sumarização para Modern Code Review
<b>Autor</b>	GABRIEL PAKULSKI DA SILVA
<b>Orientador</b>	INGRID OLIVEIRA DE NUNES

# **Exploração de Técnicas de Sumarização para Modern Code Review**

Autor: Gabriel Pakulski, Orientadora: Ingrid Nunes

Universidade Federal do Rio Grande do Sul

A Revisão de Código Moderna (do inglês, MCR) tem como principal objetivo aumentar a qualidade no desenvolvimento de software. Essa prática é tipicamente assíncrona, informal e suportada por ferramentas. Nela, um desenvolvedor submete uma alteração feita no código fonte para revisão por outros desenvolvedores. Como parte da revisão, são geradas discussões entre os revisores e o autor acerca da alteração, que podem resultar (ou não) em mudanças no código desenvolvido. Cada uma das discussões pode incluir threads de conversação. Nas discussões da MCR, podem existir informações sobre decisões de projeto do software sendo desenvolvido que ficam difusas em meio a uma grande quantidade de comentários. Essas decisões podem ser colocadas como parte da documentação do software e da alteração realizada, ou como comentários no código. Dessa forma, a informação pode auxiliar na manutenção do software, pois fica acessível a todos os desenvolvedores. Neste projeto, exploramos técnicas de sumarização de texto automáticas, aplicando-as sobre comentários da revisão de código sob duas perspectivas: (1) gerar um sumário dos comentários como um todo; (2) gerar um sumário para threads de conversação relacionadas com esclarecimentos sobre uma dada decisão de projeto. Para atingir os objetivos produzimos scripts para a extração dos dados da plataforma GitHub e para utilizar selecionados algoritmos de sumarização extrativa. Os resultados para o primeiro caso foram textos sucintos, com sentenças por vezes desconexas, que envolviam diversos assuntos. Já para o segundo, foi possível identificar o tópico central de cada thread. Porém, a falta de conexão entre as sentenças também ocorreu. Analisando os comentários dos desenvolvedores, identificamos que existe objetividade na forma de se expressar, fazendo com que a concatenação de sentenças de comentários diferentes apresente falta de coesão. Para obter melhores resultados, como trabalho futuro, sugere-se explorar diferentes formas de pré-processamento do texto, utilizando heurísticas relacionadas a este domínio.