

Desenvolvimento de recursos para pré-processamento de *tweets*

Leonardo Pereira dos Santos¹, Larissa Astrogildo de Freitas^{1*}
*Orientador

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) -
Campus Osório. Osório, RS, Brasil.

Os usuários têm publicado cada vez mais suas opiniões na *web*, através de plataformas como redes sociais, blogs, fóruns e *sites* para avaliação de produtos e serviços. Com isso, Mineração de Opinião, também chamada de Análise de Sentimentos, busca identificar o sentimento, percepção ou ação do público em relação ao alvo desta opinião. A pesquisa Análise de Sentimentos em Textos Escritos na Língua Portuguesa tem como objetivo o desenvolvimento de sistemas e recursos referentes à Mineração de Opinião em textos publicados em redes sociais. O tema abordado nesta pesquisa é o processo de *impeachment* de Dilma Rousseff, com foco em comentários opinativos feitos no Twitter, por usuários na região de Osório/RS. Para o pré-processamento destes textos, são realizadas três etapas: captura de *tweets*, que consiste na coleta de *tweets*; limpeza de *tweets*, que consiste na correção dos textos informais e na adaptação deste para uma versão mais próxima da norma culta da Língua Portuguesa; e anotação de *tweets*, que consiste na identificação do sujeito do discurso e na classificação da opinião (positiva, negativa ou neutra). Para a realização de cada uma das etapas, foram construídas ferramentas específicas. A Ferramenta de Captura de *Tweets*, responsável pela coleta dos *tweets*, é capaz de filtrar *tweets* a partir de especificações de datas, idioma, região e termos de busca. A Ferramenta de Limpeza de Textos, responsável pela limpeza dos *tweets*, é capaz de fazer substituições de termos abreviados e do *internetês* por correspondências ortográficas, a partir de um dicionário de *internetês* e abreviações. A Ferramenta de Anotação Manual Automatizada (FAMA), responsável pela anotação, é capaz de identificar o sujeito do discurso a partir de um banco de dados contendo nomes de deputados federais, senadores, ministros, partidos políticos, vice-presidente e presidente, a fim de classificar a opinião contida nos textos. Ao final da pesquisa, será desenvolvido um *website*, no qual serão disponibilizados esses recursos que deverão realizar de forma automatizada as três etapas do pré-processamento de texto, para que posteriormente, seja realizado o restante do processo de análise de sentimento.

Palavras-chave: Mineração de Opinião. Twitter. Pré-processamento.

Trabalho executado com recursos do Edital nº 012/2015/Programa Institucional de Bolsas de Iniciação Científica e/ou Tecnológica (PROBICT) do IFRS, da Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação.