

Observatório de notícias sobre trânsito de São Paulo em mídias sociais usando mineração de textos

Leticia Lauria Lopes, Vinicius Simões Zamarrenho

Prof. Dr. Leandro Augusto Silva

Introdução

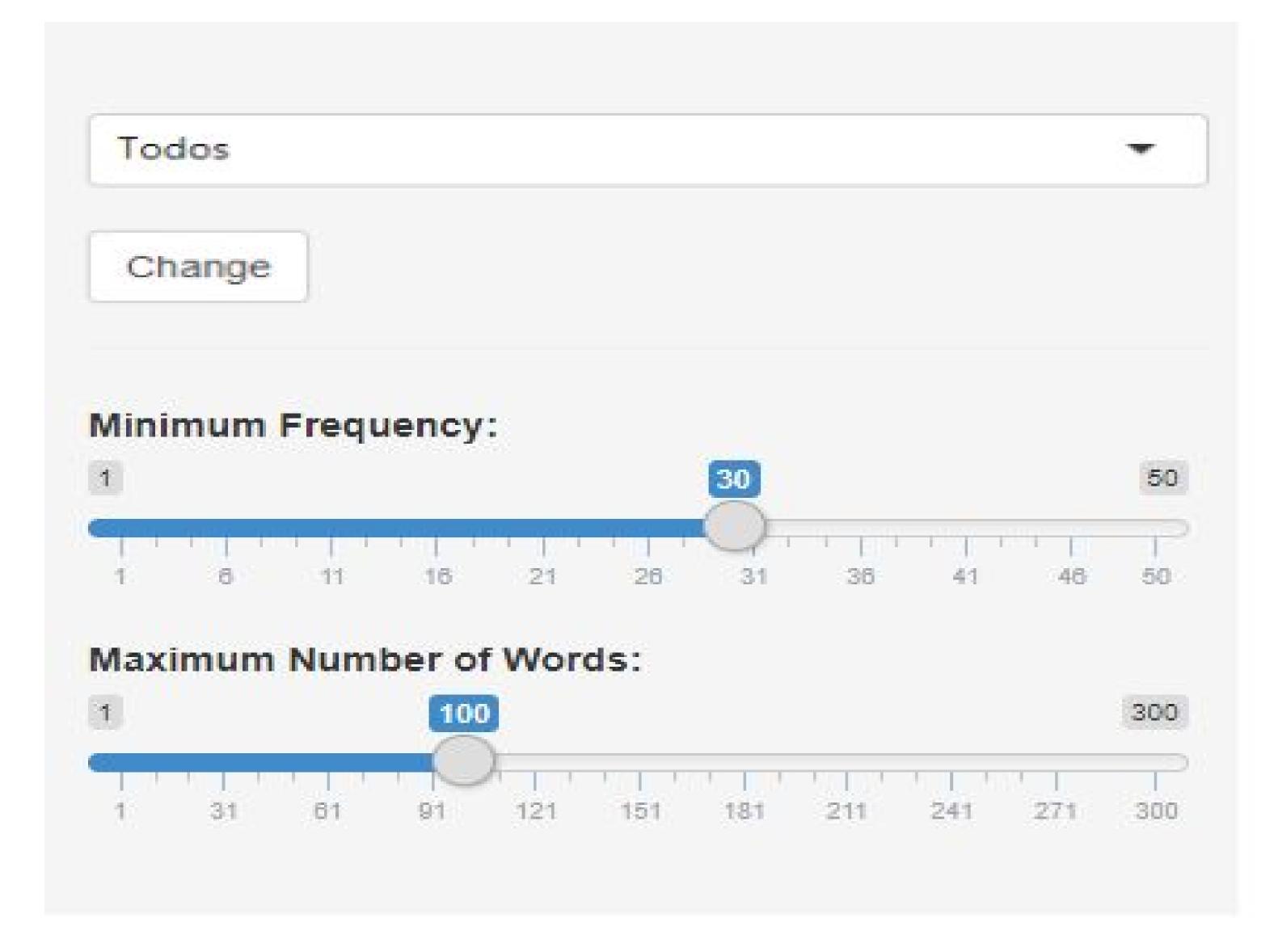
Este trabalho propõe a criação de um de sistema de coleta de informações de tráfego da cidade de São Paulo com o objetivo de apresentar notícias em tempo real sobre o trânsito em mídias sociais das principais responsáveis pelas rotas de tráfego da cidade - CPTM, Metrô, CET e SPTRANS.

Metodologia

O sistema foi desenvolvido com o uso da linguagem R. Inicialmente se faz a coleta de tweets referentes às informações de notícias das principais empresas relacionadas ao tráfego de São Paulo: SPTRANS, CPTM, METRO e CET, que lidam, respectivamente, com os ônibus, trens, metrôs e vias públicas da cidade. Após a coleta de informações, é reproduzida uma visualização dos dados por meio de nuvem de palavras, usando R e a ferramenta Shiny, exibindo as palavras mais encontradas sobre o tráfego.

Resultados

A Figura 1 apresenta um protótipo do sistema, já ilustrando o resultado final das nuvens de palavras agrupadas em cada tráfego específico. Isto significa que para chegar a esta etapa os tweets foram coletados e processados para se chegar a uma contagem das palavras mais frequentes.



A coleta é feita em intervalos regulares de cinco minutos. O sistema possui interatividade com os filtros de quantidade de palavras, frequência mínima e origem dos dados, podendo ser selecionado apenas um dos twitter específicos, um conjunto deles ou todos os quatro, para que o usuário possa ter um melhor planejamento de suas rotas.

Conclusões Preliminares e Trabalhos Futuros

Esta pesquisa apresenta como proposta uma aplicação interativa e simples de acesso à informação e apoio no planejamento de trajetos ao denunciar imprevistos como fechamento de ruas por manifestações ou acidentes nos trilhos, que causam atrasos nas viagens.

O trabalho, entretanto, ainda precisa ter os filtros de pesquisa aprimorados, uma vez que sua eficácia dificilmente é evidenciada quando não há uma situação crítica, como as mencionadas acima, devido aos temas genéricos de postagem, e sendo prejudicada por informações com ruídos, como nomes de usuário e palavras genéricas.



Figura 1: Protótipo com o exemplo de nuvem de palavras com filtros de número máximo de palavras, mínima frequência de cada palavra, e a possibilidade de escolher uma nuvem de palavras de um twitter específico ou um conjunto deles.

Referências

Silva, L.A., Peres, S.M. e Boscarioli, C. "Introdução a mineração de dados: com aplicações em R". Elsevier, 2016.

Morais, E.A.M. e Ambrósio, A.P.L. "Mineração de Textos". 2007. In: Relatório Técnico - Universidade Federal de Goiás, Goiás, 2007. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf Acesso em: 15 ago. 2017

