UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

RODRIGO PAULA DA SILVA

FILTRO ADAPTATIVO PARA REMOÇÃO DE FALSOS POSITIVOS EM ALERTAS DE SEGURANÇA OBTIDOS DO TWITTER

CAMPO MOURÃO

2023

RODRIGO PAULA DA SILVA

FILTRO ADAPTATIVO PARA REMOÇÃO DE FALSOS POSITIVOS EM ALERTAS DE SEGURANÇA OBTIDOS DO TWITTER

Adaptive filter for removing false positives in security alerts obtained from Twitter

Trabalho de conclusão de curso de graduação apresentado como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação, Universidade Tecnológica Federal do Paraná (UTFPR).
Orientador(a): Prof. Dr. Rodrigo Campiolo

CIDADE (SEDE DO CURSO/CAMPUS)

ANO DA ENTREGA

RODRIGO PAULA DA SILVA

FILTRO ADAPTATIVO PARA REMOÇÃO DE FALSOS POSITIVOS EM ALERTAS DE SEGURANÇA OBTIDOS DO TWITTER

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Ciência da Computação, Universidade Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 19 de Junho de 2023

Rodrigo Campiolo
Doutorado em Ciência da Computação
Link para o currículo Lattes: http://lattes.cnpq.br/2822469089227391
Universidade Tecnológica Federal do Paraná

Luiz Arthur Feitosa dos Santos
Doutorado em Ciência da Computação
Link para o currículo Lattes: http://lattes.cnpq.br/3725232561617394
Universidade Tecnológica Federal do Paraná

André Luís Schwerz
Doutorado em Ciência da Computação
Link para o currículo Lattes: http://lattes.cnpq.br/4954414332524750
Universidade Tecnológica Federal do Paraná

CAMPO MOURÃO

2023

RESUMO

SILVA, Rodrigo Paula da. **Filtro adaptativo para remoção de falsos positivos em alertas de segurança obtidos do Twitter**. 2023. Trabalho de conclusão de curso (Bacharelado em Ciência da Computação) – Universidade Tecnológica Federal do Paraná, Campo Mourão, 2023.

As mídias sociais são um meio moderno e usual para compartilhar informações sobre diversos assuntos. Dentre as mídias sociais mais utilizadas se destacam o Facebook, Instagram, Whatsapp e o Twitter. Há várias pesquisas no sentido de avaliar o uso do Twitter como fonte de alertas de segurança, inclusive há sistemas que o utilizam para obter alertas antecipados. Apesar dos esforços em obter alertas desta fonte, ainda sim ocorrem mensagens não relacionadas a alertas de segurança, consideradas falsos positivos. O objetivo deste trabalho consiste em desenvolver um filtro adaptativo para reduzir o número de mensagens consideradas falsos positivos, com o intuito de serem utilizadas em um sistema de alertas antecipados. Como metodologia, foi utilizada a API do Twitter para realizar a captura dos *tweets*. Foram desenvolvidos filtros de palavras, URLs, usuários e de similaridade para eliminar *tweets* irrelevantes. Para o filtro de similaridade foi utilizado o índice de Jaccard como métrica de similaridade juntamente com o peso das palavras mais recorrentes em *tweets* descartados. Os *tweets* foram agrupados utilizando também o índice de Jaccard como métrica de similaridade e selecionados via interface Web pelo analista de segurança. Como resultado, o filtro removeu 31,56% de *tweets* irrelevantes e obteve uma taxa de 1,76% de erros. O filtro se mostrou uma alternativa válida para redução de mensagens irrelevantes em sistemas de alertas antecipados utilizando como fonte de dados o Twitter.

Palavras-chave: cibersegurança; mídias sociais; recuperação de informação; sistema de alertas antecipados; similaridade de texto.

( ) Não autorizo a disponibilização de endereço de correio eletrônico para contato.

(X) Autorizo a disponibilização do seguinte correio eletrônico para contato:

rodrigopds73@gmail.com

ABSTRACT

SILVA, Rodrigo Paula da. **Adaptive filter for removing false positives in security alerts obtained from Twitter**. 2023. Trabalho de conclusão de curso (Bacharelado em Ciência da Computação) – Universidade Tecnológica Federal do Paraná, Campo Mourão, 2023. Título original: Filtro adaptativo para remoção de falsos positivos em alertas de segurança obtidos do Twitter.

Social media is a modern and usual way to share information about several subjects. Among the most used social media are Facebook, Instagram, Whatsapp and Twitter. There are several studies in the sense of evaluating the use of Twitter as a source of security alerts, including systems that use it to obtain early warnings. Despite efforts to obtain alerts from this source, messages not related to security alerts still occur, considered false positives. The objective of this work is to develop an adaptive filter to reduce the number of messages considered false positives. As methodology, the Twitter API was used to capture tweets. Word, URL, user and similarity filters were developed to eliminate irrelevant tweets. For the similarity filter, the Jaccard index was used as a similarity metric together with the weight of the most recurrent words in discarded tweets. The tweets were also grouped using the Jaccard index as a similarity metric and selected via the web interface by the security analyst. As result, the filter removed 31.56% of irrelevant tweets and achieved a 1.76% error rate. It was found that the adaptive filter was able to reduce irrelevant messages in early warning systems using Twitter as a data source.

Keywords: cybersecurity; social media; information retrieval; early warning system; text similarity.