

GIESTAS, Felipe da Costa. **Desenvolvimento de Tecnologia para extração e anonimização de dados não estruturados em boletins de ocorrência, à luz da Lei Geral de Proteção de Dados**. 2024. 162f. Dissertação (Mestrado em Segurança Pública), Programa de Pós-Graduação em Segurança Pública, Instituto de Filosofia e Ciências Humanas, Universidade Federal do Pará, Belém, Pará, Brasil, 2024.

RESUMO

Introdução/Importância: A pesquisa aborda a análise de boletins de ocorrência da Polícia Civil do Estado do Pará, com foco no desenvolvimento de um programa computacional de anonimização e extração de dados não estruturados. A importância desse estudo reside na necessidade de preservar a privacidade dos indivíduos mencionados nos boletins, ao mesmo tempo em que permite a análise estatística desses dados para fins de pesquisa e investigação. **Objetivo:** Analisar a viabilidade e eficácia da aplicação de uma análise léxica como método de anonimização de dados não estruturados em boletins de ocorrência. Além disso, busca-se avaliar a aplicabilidade prática desse processo em um contexto específico, visando contribuir para o desenvolvimento de ferramentas de processamento de dados forenses. **Método:** Trata-se de uma pesquisa aplicada com abordagem quantitativa, exploratória e descritiva. A coleta de dados foi realizada por meio de um estudo documental junto à Polícia Civil do Estado do Pará, utilizando boletins de ocorrência como fonte primária de informações. O processo de anonimização foi conduzido através do desenvolvimento de um programa computacional baseado em análise léxica para identificação e supressão de informações sensíveis nos boletins. **Resultados:** Os resultados preliminares destacaram a eficácia da análise léxica na identificação e anonimização de informações sensíveis, como nomes, números de documentos e locais específicos, presentes nos boletins de ocorrência. O programa computacional desenvolvido demonstrou sua capacidade de processar os dados de forma automatizada, permitindo a aplicação prática do método de anonimização em larga escala. Essa abordagem facilitou a proteção da privacidade dos indivíduos envolvidos nos boletins, tornando os dados adequados para análise estatística e investigativa sem comprometer a confidencialidade. Além disso, os resultados do estudo piloto na Diretoria Estadual de Combate a Crimes Cibernéticos da Polícia Civil do Estado do Pará, que analisou 5.858 boletins de ocorrência, revelaram insights valiosos para a investigação. A análise proporcionou uma compreensão detalhada dos padrões de crimes cibernéticos, incluindo a relação entre os boletins de ocorrência por meio da chave PIX, representações temporais de prejuízo monetário, frequência de uso de aplicativos e envolvimento de instituições bancárias. Esses resultados evidenciam a eficiência da ferramenta desenvolvida em Python com os frameworks Flet e Pandas, oferecendo uma solução inovadora para otimizar o processo de investigação e compreensão dos crimes cibernéticos. **Conclusão:** Este estudo conclui que a análise léxica é uma abordagem viável e eficiente para a anonimização de dados não estruturados em boletins de ocorrência. A aplicação desse método contribui para a proteção da privacidade dos indivíduos envolvidos e abre caminho para o desenvolvimento de ferramentas forenses mais sofisticadas. Recomenda-se a continuidade deste trabalho com avaliações mais detalhadas da eficácia e segurança do processo de anonimização.

Palavras-chave: Crimes cibernéticos; Privacidade; Ferramenta computacional; Análise léxica.