

# Aplicação de técnicas de raspagem de dados para criação de um banco de dados nacional sobre pessoas desaparecidas

Luiz Fernando da Silva Junior<sup>1</sup>, João Benedito dos Santos Junior<sup>2</sup>

<sup>1</sup> Aluno de Graduação do Curso de Ciência da Computação da PUC Minas Campus de Poços de Caldas – MG, Brasil

<sup>2</sup> Orientador e Docente do Departamento de Ciência da Computação da PUC Minas Campus de Poços de Caldas – MG, Brasil

luizfsjunior.2002@gmail.com, joao@pucpcaldas.br

**Abstract.** *This article describes the development of a tool that, through data scraping techniques, can collect information from police reports, news websites and social networks. This information can be used to create a national database on missing persons with the aim of assisting in the investigation of disappearance cases throughout Brazil.*

**Resumo.** *Este artigo descreve o desenvolvimento de uma ferramenta que através de técnicas de raspagem de dados consegue coletar informações de boletins de ocorrência, sites de notícias e redes sociais. Essas informações podem ser utilizadas para criar um banco de dados nacional sobre pessoas desaparecidas com o objetivo de auxiliar na investigação em casos de desaparecimento por todo território brasileiro*

## **1. Introdução**

O desaparecimento de pessoas é um problema grave que afeta milhões de pessoas em todo o mundo. No Brasil, estima-se que cerca de 60 mil pessoas desaparecem todos os anos. A investigação de casos de desaparecimento pode ser um processo complexo e desafiador. Isso ocorre porque, muitas vezes, as informações sobre o desaparecimento são limitadas. No Brasil, em 2022, foram registrados 74.061 casos de desaparecimento. Todavia, não existe uma tecnologia ou ferramenta que auxilie a investigação em nível nacional, muito menos uma padronização dos dados.

Esse projeto tem como objetivo, criar uma ferramenta que permita a padronização de dados a nível nacional e uma junção desses mesmos dados para um ajudar a inteligência policial utilizando de raspagem de dados, que pode ser utilizada para coletar informações de forma automática.

## **2. Empecilho na busca de pessoas desaparecidas**

A busca por pessoas desaparecidas não é de longe uma tarefa fácil, pois além de não existir um padrão de dados a nível nacional, a investigação é um processo complexo que demanda rapidez.

### **2.1. Dificuldade na investigação**

A investigação de pessoas desaparecidas é um processo complexo e desafiador, pois muitas vezes as informações sobre o desaparecimento são limitadas ou contraditórias. Além disso, os casos de desaparecimento podem ser motivados por diversos fatores, o que torna ainda mais difícil a identificação da pessoa desaparecida e a resolução do caso.

Alguns dos desafios mais comuns enfrentados na investigação de pessoas desaparecidas incluem:

**Falta de informações:** Muitas vezes, as informações sobre o desaparecimento são limitadas ou contraditórias. Isso pode dificultar a identificação da pessoa desaparecida e a investigação do caso.

**Motivação do desaparecimento:** Os casos de desaparecimento podem ser motivados por diversos fatores, como crime, suicídio ou fuga. Isso pode dificultar a investigação do caso e a identificação da pessoa desaparecida.

**Tempo:** O tempo é um fator importante na investigação de pessoas desaparecidas. Quanto mais tempo passa, mais difícil se torna encontrar a pessoa desaparecida.

Para superar esses desafios, é importante que as autoridades policiais tenham acesso a informações confiáveis e atualizadas. Além disso, é importante que as investigações sejam conduzidas de forma profissional e ética, respeitando os direitos das famílias das vítimas.

## **2.2. Falta de padronização dos dados**

A falta de padronização dos dados é um dos principais desafios para a investigação de pessoas desaparecidas. No Brasil, os dados sobre pessoas desaparecidas são coletados de diversas fontes, incluindo boletins de ocorrência, sites de notícias e redes sociais. No entanto, essas fontes utilizam diferentes formatos e nomenclaturas para coletar as informações, o que dificulta a integração dos dados e a identificação de padrões. Além disso, a exposição de dados sobre pessoas desaparecidas varia de estado para estado. Alguns estados têm sistemas de dados mais avançados, com informações mais detalhadas e atualizadas. Outros estados, no entanto, têm sistemas de dados mais rudimentares, com informações limitadas e desatualizadas.

Essa falta de padronização e exposição de dados dificulta a identificação de pessoas desaparecidas e a investigação de seus casos. Para superar esse desafio, é importante que as autoridades policiais trabalhem para padronizar os dados sobre pessoas desaparecidas e criar sistemas de dados mais avançados e integrados.

## **3. Possíveis soluções para obstáculos em investigações**

### **3.1. Padronização dos dados**

As autoridades policiais devem trabalhar para padronizar os dados sobre pessoas desaparecidas. Isso pode ser feito por meio da criação de um modelo de dados comum que seja utilizado por todas as fontes.

### **3.2. Integração dos dados**

Os dados sobre pessoas desaparecidas devem ser integrados em um único sistema. Isso facilitará a identificação de padrões e a investigação dos casos, podendo ser feito através da coleta de dados de maneira automática pela ferramenta de raspagem de dados.

### **3.3. Criação de um banco de dados nacional**

A criação de um banco de dados nacional sobre pessoas desaparecidas pode ajudar a superar os empecilhos de falta de padronização e exposição de dados. O banco de dados pode centralizar informações de diferentes fontes e padronizar os dados, o que facilitará a investigação dos casos, e com isso pode-se transformar os dados em uma forma mais gráfica para facilitar a interpretação, ajudando a achar padrões de desaparecimentos, por exemplo.

#### **4. Metodologia**

Os dados foram coletados de diversas fontes, incluindo boletins de ocorrência e redes sociais. Para coletar os dados de boletins de ocorrência, foi utilizado um script em *Python* para acessar alguns sites da Polícia Civil de alguns estados. Para coletar os dados de redes sociais, foi utilizado um script em *Python* com a biblioteca *Instaloader* para acessar o instagram. O script extraiu os *posts* com os cartazes criados pela polícia com alguns dados com a situação atual da pessoa, nome, idade.

Alguns dados coletados, como cidade do desaparecimento poderão ser utilizados para mapear as regiões com mais desaparecimentos, auxiliando ainda mais na descoberta de padrões e com isso podendo trazer melhores resultados nas investigações.

#### **5. Considerações finais**

O desaparecimento de pessoas é um problema grave no Brasil, com cerca de 60 mil casos registrados anualmente. A investigação desses casos é complexa devido à falta de informações, motivos diversos para os desaparecimentos e à falta de padronização dos dados. No entanto, este projeto propõe soluções, como a padronização dos dados, a integração em um sistema único e a criação de um banco de dados nacional, que podem melhorar significativamente a investigação. A metodologia de coleta de dados abrange boletins de ocorrência e redes sociais, permitindo a análise abrangente dos casos e o mapeamento de áreas com mais desaparecimentos. Essas soluções têm o potencial de trazer resultados mais eficazes na busca por pessoas desaparecidas e oferecer esperança às famílias e comunidades afetadas.

#### **6. Referências**

BARBOSA, J. E. M. Raspagem de Dados Jurídicos Utilizando Scrapy. [s.d.].

DOMINGUES, R. S. M. V. Os principais estilos de raspagem de palheta de oboé no Brasil: um estudo espectrográfico para caracterização timbrística. Anais do SIMPOM, n. 4, 12 jul. 2016.

FERREIRA, L. C. DE M. De problema de família a problema social: notas etnográficas sobre o desaparecimento de pessoas no Brasil contemporâneo. Anuário Antropológico, n. v.38 n.1, p. 191–216, 1 jun. 2013.

FORUMSEG. Desaparecidos no Brasil: da contagem de registros às responsabilidades do Estado. Disponível em:

<<https://fontesegura.forumseguranca.org.br/desaparecidos-no-brasil-da-contagem-de-registros-as-responsabilidades-do-estado/>>. Acesso em: 6 nov. 2023.

L13812. Disponível em:

<[https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/lei/l13812.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/l13812.htm)>. Acesso em: 6 nov. 2023.

LEAL, E. M. A dúvida mais persistente: as formas de governo do desaparecimento de pessoas no Brasil. 2017.

NETO, M. B. Apache Kafka: Implementação da Técnica de Replicação de Banco de Dados Baseada em Middleware para o Contexto de Raspagem de Dados. 2022.

SOUZA, W. M. A. S. DE; FILHO, W. F. T.; SANTOS, W. B. Ferramenta de Web-Scraping: Impactos da COVID-19 na Indústria de Software. Anais Estendidos do Simpósio Brasileiro de Sistemas de Informação (SBSI). Anais... Em: ANAIS ESTENDIDOS DO XVII SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO. SBC, 7 jun. 2021. Disponível em: <[https://sol.sbc.org.br/index.php/sbsi\\_estendido/article/view/15354](https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/15354)>. Acesso em: 6 nov. 2023