REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

DESENVOLVIMENTO DE UMA PLATAFORMA PARA COLETA E SISTEMATIZAÇÃO AUTOMATIZADA DE NOTÍCIAS JORNALÍSTICAS SOBRE NARCOTRÁFICO E APREENSÕES DE DROGAS ILÍCITAS: O CASO DA CARTOGRAFIAS DAS VIOLÊNCIAS NA AMAZÔNIA

DEVELOPMENT OF A PLATFORM FOR THE AUTOMATED COLLECTION AND SYSTEMATIZATION OF JOURNALISTIC NEWS ABOUT DRUG TRAFFICKING AND SEIZURES OF ILLICIT DRUGS: THE CASE OF THE CARTOGRAPHIES OF VIOLENCES IN THE AMAZON

DESARROLLO DE UNA PLATAFORMA PARA RECOPILACIÓN AUTOMATIZADA Y LA SISTEMATIZACIÓN DE NOTICIAS PERIODÍSTICAS SOBRE EL NARCOTRÁFICO Y LAS INCAUTACIONES DE DROGAS ILÍCITAS: EL CASO DE LA CARTOGRAFÍAS DE LAS VIOLENCIAS EN LA AMAZONÍA

Welington Morais Ferreira Júnior
Acadêmico de Bacharelado em Engenharia de Software, Universidade do Estado do Pará, Brasil
E-mail: welmfj@gmail.com

Pedro Leonan Gonzaga de Freitas
Acadêmico de Bacharelado em Engenharia de Software, Universidade do Estado do Pará, Brasil
E-mail: pedroleonan85@gmail.com

Welington Morais Ferreira
Doutorando em Geografia, Universidade Federal do Pará, Brasil
E-mail: welingtonmorais@hotmail.com

Marta de Oliveira Barreiros
Doutora em Engenharia Elétrica, Universidade do Estado do Pará, Brasil
E-mail: marta.do.barreiros@uepa.br

Resumo

A região Amazônica consolidou-se como um eixo central do narcotráfico transnacional, enfrentando graves problemas de subnotificação e fragmentação nos dados criminais oficiais. Para mitigar essa lacuna, a Inteligência de Fontes Abertas (OSINT), baseada em mídias jornalísticas, surge como uma alternativa essencial. Contudo, o monitoramento manual dessas fontes gera um severo gargalo operacional para equipes de pesquisa. Diante disso, este trabalho apresenta o desenvolvimento da plataforma Ártemis, uma solução fundamentada em Engenharia de Software voltada à coleta e à estruturação automatizada de informações jornalísticas sobre o narcotráfico na Amazônia Legal. A pesquisa caracteriza-se como aplicada, exploratória e descritiva, adotando a estratégia de estudo de caso no Instituto Mãe Crioula (IMC). O sistema foi construído a partir do modelo de Desenvolvimento Incremental e da arquitetura em camadas, utilizando Python, FastAPI, feedparser e SQLite. A

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

validação do artefato envolveu testes funcionais, a caracterização descritiva do corpus coletado e uma avaliação qualitativa de usabilidade baseada nas Heurísticas de Nielsen, aplicada a cinco pesquisadores seniores da instituição. A plataforma consolidou um corpus de 1.000 notícias estruturadas, abrangendo os nove estados da Amazônia Legal e 280 veículos distintos, com incremento de 150% no volume em relação ao histórico de coleta manual. A avaliação de usabilidade indicou elevada maturidade e consistência da interface. Conclui-se que o Ártemis atua como infraestrutura de coleta e pré-estruturação que subsidia a posterior análise qualitativa humana, integrando-se ao fluxo de produção do relatório "Cartografias das Violências na Amazônia".

Palavras-chave: Engenharia de Software; Coleta Automatizada de Dados; Inteligência de Fontes Abertas; Narcotráfico; Amazônia.

Abstract

The Amazon region has consolidated itself as a central hub for transnational drug trafficking, facing severe underreporting and fragmentation in official criminal data. To mitigate this gap, Open Source Intelligence (OSINT) based on journalistic media emerges as an essential alternative. However, the manual monitoring of these sources creates a severe operational bottleneck for research teams. In this context, this work presents the development of the Ártemis platform, a Software Engineering solution aimed at the automated collection and structuring of journalistic information on drug trafficking in the Legal Amazon. The research is applied, exploratory, and descriptive, adopting a case study strategy at the Instituto Mãe Crioula (IMC). The system was built using the Incremental Development model and a layered architecture, with Python, FastAPI, feedparser, and SQLite. Validation involved functional testing, a descriptive characterization of the collected corpus, and a qualitative usability assessment based on Nielsen's Heuristics applied to five senior researchers. The platform consolidated a corpus of 1,000 structured news items, covering the nine states of the Legal Amazon and 280 distinct outlets, a 150% increase in volume over the manual collection baseline. The usability assessment indicated high maturity and visual consistency of the interface. It is concluded that Ártemis acts as a collection and pre-structuring infrastructure that supports subsequent human qualitative analysis, integrated into the production of the report "Cartografias das Violências na Amazônia".

Keywords: Software Engineering; Automated Data Collection; Open Source Intelligence; Drug Trafficking; Amazon.

Resumen

La región amazónica se ha consolidado como un eje central del narcotráfico transnacional, enfrentando graves problemas de subnotificación y fragmentación en los datos criminales oficiales. Para mitigar esta brecha, la Inteligencia de Fuentes Abiertas (OSINT), basada en medios periodísticos, surge como una alternativa esencial. Sin embargo, el monitoreo manual de estas fuentes genera un severo cuello de botella operativo para los equipos de investigación. Ante esto, este trabajo presenta el desarrollo de la plataforma Ártemis, una solución fundamentada en Ingeniería de Software orientada a la recopilación y estructuración automatizada de información periodística sobre el narcotráfico en la Amazonía Legal. La investigación es aplicada, exploratoria y descriptiva, adoptando la estrategia de estudio de caso en el Instituto Mãe Crioula (IMC). El sistema se construyó a partir del modelo de Desarrollo Incremental y de la arquitectura en capas, utilizando Python, FastAPI, feedparser y SQLite. La validación involucró pruebas funcionales, la caracterización descriptiva del corpus recopilado y una evaluación cualitativa de usabilidad basada en las Heurísticas de Nielsen, aplicada a cinco investigadores sénior. La plataforma consolidó un corpus de 1.000 noticias estructuradas, abarcando los nueve estados de la Amazonía Legal y 280 medios distintos, con un incremento del 150% en el volumen respecto al histórico de recolección manual. La evaluación de usabilidad indicó elevada madurez y consistencia de la interfaz. Se concluye que Ártemis actúa como

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

infraestructura de recopilación y preestructuración que sustenta el posterior análisis cualitativo humano, integrándose al flujo de producción del informe "Cartografias das Violências na Amazônia".

Palabras clave: Ingeniería de Software; Recopilación Automatizada de Datos; Inteligencia de Fuentes Abiertas; Narcotráfico; Amazonas

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

1. Introdução

A região Amazônica, em virtude de sua vasta extensão territorial e complexidade logística, configura-se como um cenário estratégico e desafiador para a segurança pública. Segundo o Fórum Brasileiro de Segurança Pública e o Instituto Mãe Crioula (2025), a região consolidou-se como eixo central da logística criminal na América Latina, com papel fundamental nas dinâmicas transnacionais do narcotráfico. Nesse contexto, Couto (2023) argumenta que o narcotráfico produz sua própria geografia, desarticulando a vigilância do Estado e explorando a vulnerabilidade social para estabelecer relações de poder cujos líderes permanecem, em geral, invisíveis aos órgãos de fiscalização.

A obtenção de dados estatísticos fidedignos sobre essa criminalidade, contudo, enfrenta obstáculos metodológicos. O 19º Anuário Brasileiro de Segurança Pública (FBSP, 2025) aponta uma desarticulação na sistematização das métricas de apreensão de drogas entre as instituições policiais, que resulta em sobreposições e duplicidade de registros. Somado a isso, o Escritório das Nações Unidas sobre Drogas e Crime (UNODC, 2023) destaca que a escassez de dados desagregados e o baixo detalhamento geográfico das apreensões comprometem a mensuração precisa do fenômeno, caracterizando possivelmente uma subnotificação estrutural. Diante dessa lacuna, organismos como o próprio UNODC (2025) passaram a incorporar fontes abertas e relatórios de apreensões individuais para mitigar as falhas dos sistemas de monitoramento oficiais.

A relevância das fontes abertas é corroborada por trabalhos recentes que demonstram a viabilidade de recolher e processar dados digitais para identificar dinâmicas criminais complexas (HU et al., 2021; VALDIVIA-GRANDA, 2025). Por meio da automação, esses estudos evidenciam que a estruturação de informações públicas dispersas na internet permite mapear padrões operacionais e geográficos noticiados sobre o narcotráfico que, de outra forma, permaneceriam fragmentados ou ocultos diante do grande volume e da diversidade dos dados digitais.

Nesse cenário, o Instituto Mãe Crioula (IMC), organização independente e sem fins lucrativos voltada à pesquisa científica e a ações comunitárias em segurança pública, justiça climática, crimes ambientais e direitos humanos na Amazônia (INSTITUTO MÃE CRIOULA, 2023), desenvolve, em parceria com o

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

FBSP, o relatório "Cartografias das Violências na Amazônia", dedicado ao mapeamento de crimes ambientais e dinâmicas de facções. Sua 4ª edição (2025) adota prática metodológica alinhada à do UNODC, complementando dados institucionais com fontes abertas. Entre estas, os dados jornalísticos são essenciais ao relatório, embora sua natureza não estruturada e sua diversidade imponham desafios de coleta e padronização que excedem a capacidade de processamento manual. Diante desse gargalo, a Engenharia de Software oferece um caminho de intervenção: conforme Silva e Lima (2024), a automação substitui tarefas manuais repetitivas por sistemas automatizados, proporcionando execução mais rápida e precisa e evitando inconsistências inerentes à ação humana.

Assim, formula-se a questão de pesquisa: de que maneira o desenvolvimento de um software voltado à automação da coleta de dados abertos pode otimizar o fluxo de triagem qualitativa de informações sobre o narcotráfico na Amazônia pelo Instituto Mãe Crioula? A pesquisa posiciona-se na interseção entre a Engenharia de Software e as ciências sociais aplicadas, oferecendo resposta tecnológica a um problema de dispersão de dados na Amazônia Legal. Sob os aspectos científicos e técnicos, justifica-se ao propor uma solução replicável de automação e tratamento de dados não estruturados, demonstrando como ferramentas de extração de fontes abertas viabilizam padronização, velocidade e integridade no consumo de fluxos informacionais massivos. Do ponto de vista prático e institucional, otimiza o capital humano do IMC, eliminando o desgaste da triagem manual e permitindo que os pesquisadores redirecionem esforços para a análise qualitativa, o que eleva o rigor e o impacto dos relatórios.

Dessa forma, este trabalho descreve o desenvolvimento da plataforma Ártemis, solução fundamentada em Engenharia de Software voltada à coleta e estruturação automatizada de informações jornalísticas, já integrada ao fluxo de pesquisa do IMC como subsídio à produção da 5ª edição (2026) do relatório "Cartografias das Violências na Amazônia". Mais do que um suporte, o Ártemis estabelece uma infraestrutura escalável para o monitoramento informacional contínuo, propondo uma metodologia de engenharia replicável para contextos de dados massivos e dispersos. Sua validação foi realizada junto aos pesquisadores do IMC, com vistas à viabilidade operacional do fluxo automatizado.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

O trabalho estrutura-se em cinco seções. A seção 2 apresenta a revisão de literatura; a seção 3 detalha os procedimentos metodológicos e o aparato técnico; a seção 4 expõe e discute os resultados; e a seção 5 sintetiza as considerações finais e perspectivas futuras.

2. Revisão da Literatura

2.1 A Vulnerabilidade Territorial e a Invisibilidade de Dados Oficiais sobre o Crime Organizado na Amazônia.

A configuração contemporânea do narcotráfico na Amazônia Legal indica uma transição estrutural em que a região deixa de ser apenas rota logística para consolidar-se como sistema de controle territorial complexo. Essa transformação evidencia-se na capilarização das facções, com destaque para o Comando Vermelho (CV), que expandiu sua presença para 286 municípios, crescimento de 123% em dois anos, consolidando hegemonia sobre eixos estratégicos como a "Rota do Solimões", onde a porosidade das fronteiras favorece o fluxo de ilícitos (FBSP; IMC, 2025). O UNODC (2023) confirma essa dinâmica ao caracterizar a bacia amazônica como um "ecossistema de crimes convergentes", no qual organizações como o PCC, o CV e dissidências das FARC integram o narcotráfico a economias ambientais ilícitas por meio do controle territorial.

Os sistemas oficiais, contudo, falham no registro de crimes em áreas remotas e ribeirinhas, onde atividades como a pirataria fluvial e o tráfico de insumos para o garimpo nem sempre são classificadas (FBSP; IMC, 2025). O UNODC (2023) aprofunda essa lacuna ao apontar que a escassez de dados desagregados e o baixo detalhamento geográfico das apreensões impedem a mensuração precisa do fenômeno, sugerindo uma subnotificação estrutural que mascara a real magnitude da influência criminal (FBSP, 2025). Essa fragilidade é amplificada pela natureza adaptativa das redes, Couto (2023) argumenta que o narcotráfico organiza-se a partir do controle de fluxos e rotas, configurando territórios-rede que submetem as populações à governança criminal. Cruzada com os mais de 1.100 rios de difícil monitoramento descritos pelo UNODC (2023), evidencia-se que a velocidade da dinâmica criminal supera o tempo de resposta dos registros administrativos.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

2.2 Inteligência de Fontes Abertas (OSINT) e o Ambiente de Dados Não Estruturados.

A Inteligência de Fontes Abertas (OSINT) compreende o uso sistemático e orientado de informações obtidas em mídias digitais, registros oficiais, plataformas públicas e conteúdo jornalístico para a produção de conhecimento (MOREIRA; OLIVEIRA; FERREIRA, 2026). Na segurança pública, a necessidade de ampliar o escopo de análise é impulsionada pela "explosão informacional" da era digital (LEITE, 2014), tornando-se ainda mais relevante diante da natureza das estatísticas criminais oficiais que, conforme Miranda e Muniz (2022), não constituem retratos neutros da realidade, mas construções sociais resultantes de práticas seletivas e burocráticas ajustadas às prioridades da cultura policial. Nesse contexto, as mídias de notícias desempenham papel essencial na obtenção de dados atualizados e na revelação de fatos ausentes dos registros oficiais (MOREIRA; OLIVEIRA; FERREIRA, 2026).

A viabilidade de monitorar esses dados, porém, colide com seu volume e volatilidade. Leite (2014) adverte que o fluxo massivo de informação bruta desafia a produção de conhecimento em tempo eficiente, com risco de perda de dados cruciais sob triagem manual. Moreira, Oliveira e Ferreira (2026) complementam que, embora acessíveis, esses dados só adquirem valor analítico mediante processos rigorosos de coleta, organização, filtragem, contextualização e interpretação. Para amenizar esse gargalo, Furtado (2022) destaca a relevância de soluções tecnológicas para o tratamento de dados não estruturados, observando que a exploração de evidências textuais em meios digitais requer métodos de inteligência artificial para transformar textos livres em conhecimento estruturado.

Consolida-se, assim, a OSINT não como mera coleta, mas como sistema de produção de conhecimento que integra tecnologia e metodologia para converter informações dispersas em insumos estratégicos (MOREIRA; OLIVEIRA; FERREIRA, 2026). Sem substituir os métodos tradicionais de investigação, a manipulação de dados abertos os complementa, elevando a capacidade preditiva e analítica ao cruzar relatórios, notícias e bancos oficiais (FURTADO, 2022) e fornecendo superioridade informacional ao tomador de decisão em tempo ágil (LEITE, 2014; MOREIRA; OLIVEIRA; FERREIRA, 2026). Cabe ressaltar, contudo, que esse ciclo

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

abrange desde a coleta até a análise interpretativa, o artefato proposto neste trabalho situa-se na etapa inicial, atuando como infraestrutura de coleta e pré-estruturação de fontes abertas, enquanto a interpretação, a validação substantiva e a análise de padrões permanecem a cargo da curadoria humana dos pesquisadores.

2.3 Extração de dados não estruturados na web

A internet consolidou-se como o maior repositório público de informações do mundo, em crescimento contínuo, massivo e exponencial (GRACIANO, 2022). Como a maior parte desse ecossistema é composta por dados não estruturados dispersos em páginas eletrônicas, impõem-se desafios complexos à recuperação, triagem e análise sistemática da informação (GRACIANO, 2022). Diante disso, o web scraping surge como abordagem técnica indispensável, conceituado como a prática de coletar e extrair dados de sites por qualquer meio automatizado que não seja um programa interagindo com uma API ou um operador humano utilizando manualmente um navegador (MITCHELL, 2018).

Sob a perspectiva computacional, o fluxo inicia-se com requisições a um servidor web para resgatar dados, rotineiramente estruturados em HTML e arquivos adjacentes; em seguida, o script executa a análise sintática (parsing), examinando a árvore do código para isolar e capturar os elementos textuais demandados (MITCHELL, 2018). A finalidade é transformar dados heterogêneos e não estruturados em estruturas organizadas e legíveis, persistidas em planilhas locais ou bancos de dados para manipulação analítica (MITCHELL, 2018; GRACIANO, 2022).

A substituição dos mecanismos tradicionais de coleta por fluxos automatizados é fundamental para a produtividade e a integridade no tratamento das informações (GRACIANO, 2022). Métodos manuais exigiriam milhares de horas para agrupar, copiar e padronizar registros textuais fragmentados (UPADHYAY et al., 2017 apud GRACIANO, 2022). A extração por scrapers elimina esse esforço repetitivo, reduzindo custos de tempo e mão de obra ao fornecer fluxos massivos de dados de forma quase instantânea, escala que demandaria horas a dias por mecanismos convencionais (UPADHYAY et al., 2017 apud GRACIANO, 2022; MITCHELL, 2018).

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

2.4 Trabalhos relacionados

No monitoramento do narcotráfico em ambientes digitais, Hu et al. (2021) desenvolveram um sistema automatizado de coleta no Instagram que emprega um modelo profundo multimodal (texto e imagem) para classificar simultaneamente nove categorias de entorpecentes. Em direção próxima, Valdivia-Granda (2025) propôs um pipeline em Python que extrai notícias por APIs e feed RSS sobre o tráfico internacional de armas, aplicando modelos de linguagem (Transformers) para geocodificar ocorrências e categorizá-las segundo a taxonomia estatística da ONU (ICCS).

Embora convirjam no uso da automação, o Ártemis diferencia-se de ambos, conforme sintetiza a Tabela 1. Frente a Hu et al. (2021), distingue-se pela fonte e pelo formato dos dados, ao especializar-se em portais de imprensa puramente textuais, em vez de redes sociais com análise multimodal. Frente a Valdivia-Granda (2025), difere pelo escopo temático e pela escala, substituindo o monitoramento global de armamentos por um foco regional aprofundado no narcotráfico na Amazônia Legal. Sobretudo, ao contrário das duas abordagens, que incorporam classificação e análise automatizadas, o Ártemis posiciona-se como infraestrutura leve de coleta e pré-estruturação, projetada para o fluxo de trabalho local de uma instituição de pesquisa real, deixando a interpretação analítica a cargo da curadoria humana.

Tabela 1 - Comparativo entre os trabalhos relacionados

Critério	**Hu et al. (2021)**	**Valdivia-Granda (2025)**	**Ártemis(Proposta)**
Fonte de dados	Instagram (posts e comentários; texto e imagem)	Notícias multilíngues, via APIs públicas/assinatura e feed RSS	Feed RSS do Google News (notícias jornalísticas), com foco na Amazônia Legal
Escala	4.648 eventos rotulados	8.171 relatos (2014–2024), em mais de	1.000 notícias (janeiro–maio 2026),

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

		80 países	nos 9 estados, de 280 fontes
Método de extração	Coleta automatizada por hashtags, guiada por modelo de imagem (VGG-16) e expansão iterativa	Pipeline de ingestão (APIs e RSS) com tradução automática, normalização e remoção de duplicatas	Ingestão automatizada de feed RSS (feedparser).
Classificação	Aprendizado multimodal multirrótulo (BERT + ResNet50);10 rótulos de droga	Classificação por transformadores (TopicBERT e ICCS-RoBERTa) na taxonomia ICCS/UNODC	Não realiza classificação automática; filtragem por palavra-chave e curadoria humana
Geocodificação	Não realiza	Sim, NER (SpaCy/Flair) e geocodificação (GeoNames)	Não realiza; atribuição territorial por curadoria humana
Privacidade	Não discutida; coleta perfis e páginas de usuários	Minimização explícita (apenas elementos em JSON; conteúdo integral não persistido)	Persiste apenas metadados (título, fonte, data, URL); não reproduz o conteúdo integral
Replicabilidade	Dataset disponibilizado; hiperparâmetros e ambiente informados	Alta; hiperparâmetros detalhados e saídas estruturadas reprodutíveis	Ferramentas abertas e acessíveis; reprodutibilidade limitada pela dependência do feed do Google News

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Validação	Quantitativa	Quantitativa	Qualitativa
Finalidade analítica	Detecção automática de eventos e tipos de droga	Mapeamento de estrutura e dinâmica (entidades, rotas, correlações; alerta antecipado)	Infraestrutura de coleta e pré-estruturação OSINT; a análise interpretativa permanece humana

Fonte: Elaborado pelos autores

3. Metodologia

3.1 Classificação da Pesquisa

Quanto à natureza, esta pesquisa caracteriza-se como aplicada, entendida como o conjunto de atividades em que conhecimentos previamente adquiridos são utilizados para coletar, selecionar e processar dados com vistas à geração de impacto (FLEURY; WERLANG, 2017). Tal aplicação materializa-se no desenvolvimento da plataforma Ártemis, que mobiliza conceitos da Engenharia de Software para coletar informações jornalísticas e ampliar a capacidade de monitoramento do narcotráfico na Amazônia pelo Instituto Mãe Crioula (IMC). Quanto aos objetivos, é exploratória e descritiva: exploratória ao investigar como a automação pode apoiar a triagem de dados jornalísticos sobre o narcotráfico, e descritiva ao detalhar a arquitetura e os fluxos informacionais estruturados pelo sistema (GUERRA, 2024).

Adota-se ainda a estratégia de estudo de caso, útil para investigar fenômenos complexos em seu contexto real (SÁTYRO; D'ALBUQUERQUE, 2020), delimitada ao fluxo operacional do IMC, cenário em que o artefato foi desenvolvido, implantado e validado de forma customizada para a produção da nova edição do relatório "Cartografias das Violências na Amazônia". Por fim, quanto à abordagem do problema, o estudo é qualitativo (GUERRA, 2024), uma vez que a plataforma atua na captação, centralização e estruturação de material estritamente textual obtido via feed RSS, organizando-o para uma triagem contextualizada que subsidia diretamente a posterior interpretação analítica humana.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

3.2 Compreensão do Processo de Coleta de Informações e Elicitação de Requisitos

Para compreender as etapas de elaboração da nova edição do relatório e mapear as necessidades computacionais da instituição, realizou-se um ciclo de quatro reuniões de levantamento com cinco pesquisadores seniores do IMC, em março de 2026. Os encontros, semidirigidos e apoiados em roteiro estruturado, abordaram o fluxo de trabalho macro da instituição, as metodologias de coleta em fontes abertas e as limitações operacionais da equipe, afunilando progressivamente as discussões para os pontos críticos da triagem informacional.

Na fase seguinte, as necessidades técnicas do sistema foram elicitadas e categorizadas segundo a taxonomia de Sommerville (2018): os Requisitos Funcionais (RF), que descrevem os serviços que o sistema deve fornecer e como deve reagir a entradas específicas, e os Requisitos Não Funcionais (RNF), que estabelecem restrições e critérios de qualidade, como o tempo, padrões e usabilidade, sendo aplicáveis ao sistema como um todo. Essa abordagem permitiu traduzir as demandas operacionais do instituto em especificações técnicas, detalhadas na seção de resultados.

3.3 Processo de Desenvolvimento e Arquitetura

Para a construção do sistema, foi seguido o processo de Desenvolvimento Incremental. Este modelo se baseia na ideia de desenvolver uma implementação inicial, obter feedback dos usuários e evoluir o software através desses retornos, sendo a abordagem mais comum para o desenvolvimento de aplicações e produtos de software (SOMMERVILLE, 2018, pág 35). No contexto do Ártemis, este modelo foi seguido pois, em um momento inicial, havia apenas a noção de uma funcionalidade principal, que é realizar a busca e mitigar o gargalo.

Em relação a arquitetura do sistema, utilizou-se a Arquitetura em Camadas, escolhida devido a sua capacidade de modularizar programas ricos em informações através da separação desses módulos em partes independentes, dividindo o código em três camadas principais: Camada de Interface (UI), Camada de Lógica de Negócios (BLL) e uma Camada de Acesso aos Dados (DAL) (FOWLER, 2015), o diagrama de arquitetura se encontra na seção de resultados.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

3.4 Modelagem do sistema

A representação dinâmica dos comportamentos do sistema foi definida por meio de diagramas baseados na Linguagem de Modelagem Unificada (UML), entendida como uma linguagem visual de propósito geral, aplicável a diferentes domínios e independente de um processo de desenvolvimento específico (GUEDES, 2011). Optou-se por um diagrama de casos de uso e três diagramas de atividades. Conforme Guedes (2011), o diagrama de casos de uso, mais geral e de linguagem acessível, é empregado no levantamento e análise de requisitos para representar o comportamento global do software, enquanto o diagrama de atividades descreve os passos necessários à conclusão de uma atividade específica, concentrando-se no fluxo de controle.

3.5 Stacks tecnológicas e lógica de construção do sistema

O desenvolvimento técnico priorizou portabilidade, baixo consumo de processamento local e agilidade. A lógica de negócios foi construída em Python com o framework FastAPI, escolhido por seu desempenho e pelo suporte nativo a operações assíncronas, que permitem executar varreduras em segundo plano sem comprometer a usabilidade. Para a captura das informações jornalísticas, utilizou-se a biblioteca feedparser, voltada ao consumo do feed RSS do Google News, trabalhando exclusivamente com arquivos de texto estruturados. Embora o web scraping clássico opere sobre páginas HTML, a opção pela ingestão de feed RSS torna o processo mais leve e estável.

A API organiza o fluxo em cinco endpoints principais. O processo inicia-se na verificação de disponibilidade do serviço (GET /). A coleta ocorre no endpoint de investigação (GET /investigar), que recebe os parâmetros de busca, idioma e filtros temporais (after e before) e aciona o motor de busca estruturada para capturar e formatar as notícias do agregador. A manutenção dos dados conta com os endpoints de listagem (GET /acervo) e de remoção (DELETE /deletar/{noticia_id}), este responsável por excluir um registro a partir de seu identificador único.

A persistência adota o banco SQLite, pela portabilidade e simplicidade de implantação. O arquivamento de novas notícias ocorre no endpoint POST /salvar, onde se executa a rotina de deduplicação por URL: ao receber a notícia em JSON,

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

a API verifica se a URL já consta no acervo; se inédita, o registro é salvo; se idêntica a uma já existente, a inserção é rejeitada com um status de erro controlado, evitando redundância. O esquema do banco é apresentado na seção de resultados. A interface foi desenvolvida em HTML5 e JavaScript, com o framework Tailwind CSS para a estilização visual.

3.6 Procedimentos de Validação e Avaliação

Os procedimentos de validação e avaliação estruturaram-se em três frentes: a verificação de conformidade técnica, a caracterização descritiva do corpus e a avaliação de usabilidade. A conformidade técnica baseou-se em testes funcionais sobre os Requisitos Funcionais, mediante inserção controlada de entradas válidas e inválidas para validar o motor de varredura, a integridade do parsing do XML, as rotinas de bloqueio de duplicidade e a persistência no banco SQLite. O corpus consolidado foi caracterizado de forma descritiva quanto à distribuição por estado, veículo de imprensa, tipo de droga e incidência de duplicidade, em coerência com a vertente descritiva da pesquisa. A avaliação de usabilidade, conduzida junto a cinco pesquisadores seniores do IMC, teve como referencial as dez Heurísticas de Nielsen (NIELSEN, 1994), com percepções coletadas por questionário estruturado em escala Likert de cinco pontos (1 = Discordo Totalmente a 5 = Concordo Totalmente). Complementarmente, a operação do motor de busca foi validada em duas fases: a inicial (janeiro a março de 2026), centrada no cruzamento de siglas e nomes dos estados com quatro palavras-chave essenciais (maconha, cocaína, skunk e crack), e a de produção expandida (abril a maio de 2026), voltada às demandas de longo prazo do instituto, cujo refinamento de strings e operadores está consolidado na Tabela 2.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Tabela 2 – Palavras-chave utilizadas

Sigla dos estados	**Estados**	**Tráfico de Drogas**	**Tipos de Drogas**	**Organização criminosa**
PA	Pará	Microtráfico	Drogas sintéticas	Facção
MA	Maranhão	Narcotráfico	Key	Organização Criminosa
AP	Amapá	Apreensão de Drogas	Cocaína	Faccionado
TO	Tocantins	Apreende Drogas	Maconha	Tribunal do Crime
MT	Mato Grosso		Skunk	Grupo Criminoso
AM	Amazonas		Crack	
RO	Rondônia		Oxi	
RR	Roraima		Ecstasy	
AC	Acre		MD	
			Loló	

Fonte: Elaborado pelos autores

3.7 Aspectos éticos e legais

O desenvolvimento de ferramentas de coleta automatizada sobre criminalidade exige a observância de aspectos éticos e legais. Quanto aos direitos autorais, o Ártemis persiste exclusivamente metadados jornalísticos (título, fonte, data e URL), sem reproduzir o conteúdo integral das matérias, preservando os direitos das fontes e remetendo o usuário ao veículo original pelo endereço armazenado.

No quesito da proteção de dados, embora a Lei Geral de Proteção de Dados (BRASIL, 2018) incida sobre o tratamento de dados pessoais, o Ártemis não coleta dados de indivíduos, restringindo-se a metadados de publicações de acesso público. Esse desenho alinha-se ao princípio da necessidade (art. 6º, III), que orienta a coleta apenas do estritamente indispensável à finalidade; ademais, a própria Lei estabelece que o tratamento de dados de acesso público deve considerar a

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

finalidade que justificou sua disponibilização (art. 7º, §3º), o que reforça a opção por uma coleta minimalista e orientada à pesquisa. Quanto aos termos de uso, a captação ocorre pelo consumo de um feed RSS público, disponibilizado pelo próprio agregador para redistribuição, sem contornar barreiras técnicas ou raspar páginas protegidas.

Por fim, a interpretação substantiva permanece sob curadoria humana dos pesquisadores. Por não automatizar e nem classificar atores ou territórios, a ferramenta mitiga o risco de estigmatização individual ou regional, reforçando seu posicionamento como infraestrutura de coleta, e não como sistema autônomo de inteligência.

4. Resultados

4.1 Diagnóstico do Gargalo Operacional

Identificou-se que a metodologia manual dividia os nove estados da Amazônia Legal entre a equipe, ficando cada pesquisador responsável por dois ou três estados, com coleta executada por buscas manuais não sistematizadas na web para preencher bases destinadas a análises quantitativas e qualitativas. Os pesquisadores relataram elevado desgaste operacional decorrente da densidade informacional da web e do fator temporal, pois, dada a alta frequência de publicações, a interrupção da coleta por apenas 24 horas gerava risco crítico de perda de eventos relevantes. Como consequência, o tempo dedicado à extração superava o disponível para a análise, comprometendo a participação dos pesquisadores nas demais etapas e exigindo constantes realocações de responsabilidades. Disso resultavam lacunas informacionais e perda de precisão no monitoramento das dinâmicas criminais, afetando a agilidade e o rigor do relatório final.

4.2 Concepção da Plataforma e Especificação de Requisitos

Para mitigar o desgaste da equipe e fornecer o suporte computacional demandado pelo IMC, projetou-se o sistema "Ártemis". A alcunha, em referência à deusa grega da caça, simboliza a proposta da ferramenta: um mecanismo de busca preciso, de largo alcance e focado na captura sistemática de informações dispersas.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

A partir da elicitação descrita na metodologia, as demandas dos pesquisadores foram convertidas em especificações técnicas: a Tabela 3 consolida os Requisitos Funcionais (RF) obrigatórios para a automação do fluxo de trabalho, e a Tabela 4 descreve os Requisitos Não Funcionais (RNF) que delimitam as restrições de tecnologia, desempenho e interface.

Tabela 3 - Requisitos Funcionais

ID	Requisito	Descrição	Status
RF01	Função principal	O sistema deve realizar a coleta de notícias na web	Implementado e validado
RF02	Busca por parâmetros	O sistema deve permitir a inserção de uma ou mais palavras-chave	Implementado e validado
RF03	Tradução de parâmetros	O sistema deve conter uma lógica de tradução de palavras-chave para o caso de haver inserção de termos em outro idioma	Planejado (Não foi validado)
RF04	Extração de metadados	O sistema deve capturar os metadados das notícias (Título, fonte, data, URL)	Implementado e validado
RF05	Filtros de busca	O sistema deve possuir filtros de data, quantidade de notícias e idioma	Implementado e validado
RF06	Persistência de dados	O sistema deve permitir que o usuário armazene notícias	Implementado e validado
RF07	Visualização dos dados	O sistema deve permitir que o usuário visualize as notícias que forem armazenadas	Implementado e validado
RF08	Controle do acervo	O sistema deve permitir que o usuário tenha controle sobre as notícias armazenadas	Implementado e validado
RF09	Seleção em massa	O sistema deve permitir a seleção total e individual das notícias salvas	Implementado e validado
RF10	Exportação CSV	O sistema deve possuir uma funcionalidade de exportação das notícias armazenadas para um arquivo estruturado em CSV	Implementado e validado

Fonte: Elaborado pelos autores

Tabela 4 - Requisitos Não funcionais

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

ID	Requisito	Descrição
RNF01	Usabilidade	A interface deve adotar paleta de baixa saturação e contraste adequado à leitura.
RNF02	Desempenho de Resposta	As requisições de busca devem ser processadas de forma assíncrona, sem bloquear a interface, com tempo de resposta ≤ 1,5 segundos por consulta.
RNF03	Desempenho de Exportação	A exportação do acervo em CSV deve concluir em ≤ 3 segundos para até 1.000 registros.
RNF04	Escalabilidade	O acervo deve suportar ao menos 10.000 registros persistidos sem degradação perceptível de desempenho.
RNF05	Confiabilidade	Em caso de indisponibilidade do feed RSS, o sistema deve emitir alerta e preservar integralmente o acervo já persistido, sem perda de dados.
RNF06	Integridade	O sistema deve garantir taxa de duplicidade nula no acervo, por deduplicação baseada na URL como chave única (0% de URLs duplicadas).
RNF07	Portabilidade	O sistema deve operar em diversos tipos de navegadores caso seja necessário
RNF08	Segurança de dados	O sistema deve persistir exclusivamente metadados jornalísticos (título, fonte, data e URL), sem dados pessoais, em conformidade com o princípio da minimização (LGPD).
RNF09	Recuperação	O banco de dados (SQLite) deve permitir cópia de segurança recuperável do acervo.
RNF10	Feedback visual	A interface deve exibir animações de transição e estados de carregamento que informem o usuário sobre os processos internos em execução.
RNF11	Separação de dependências	Os módulos de processamento, armazenamento e interface devem permanecer desacoplados, conforme a arquitetura em camadas (UI/BLL/DAL).

Fonte: Elaborado pelos autores

4.3 Artefatos do sistema

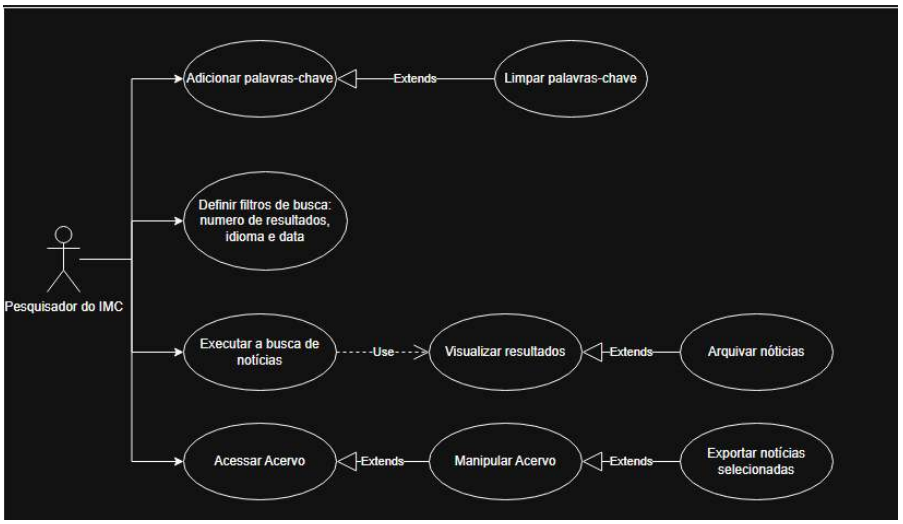
O Diagrama de Casos de Uso (Figura 1) demonstra as interações do usuário organizadas em quatro macro funcionalidades do sistema: inserção de parâmetros, delimitação de escopo, motor de varredura e gestão de dados. As relações de

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

<include> e <extend> foram empregadas para indicar, respectivamente, a obrigatoriedade do retorno de uma resposta após a busca e os fluxos condicionais da interface, como a limpeza de campos e a exportação seletiva.

Figura 1 – Diagrama de Casos de Uso



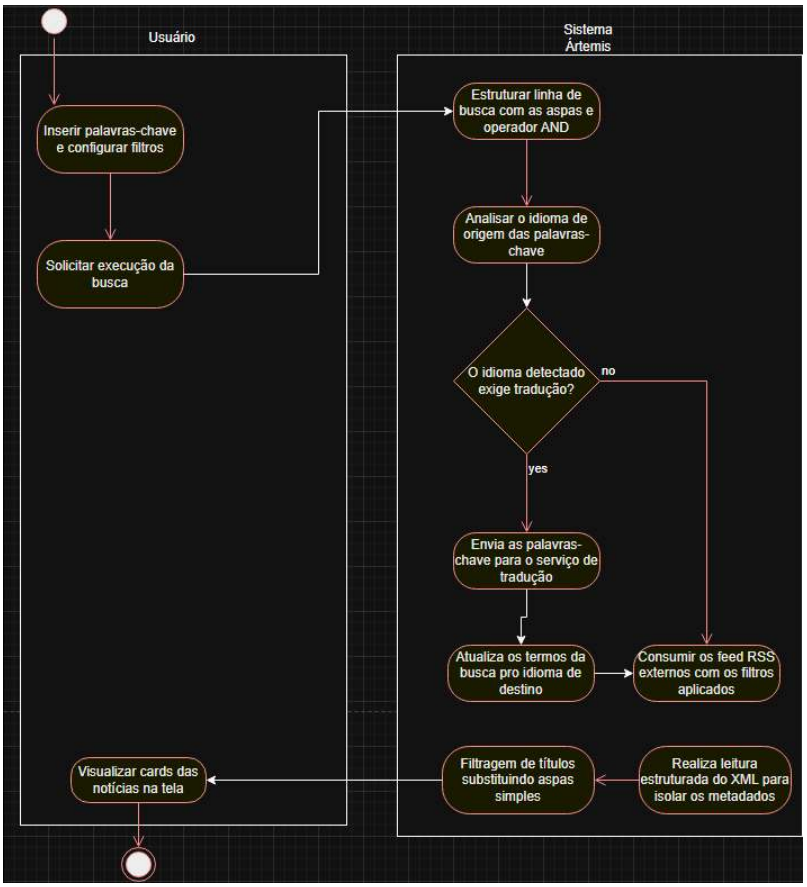
Fonte: Elaborado pelos autores

O primeiro Diagrama de Atividades (Figura 2) ilustra o fluxo lógico da coleta de informações. O modelo destaca o tratamento computacional inicial dos termos (encapsulamento em aspas e injeção do operador AND), a validação para internacionalização das palavras-chave e as etapas de parsing do feed RSS externo, culminando na renderização dinâmica dos resultados para o usuário.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Figura 2 - Diagrama de Atividades 1



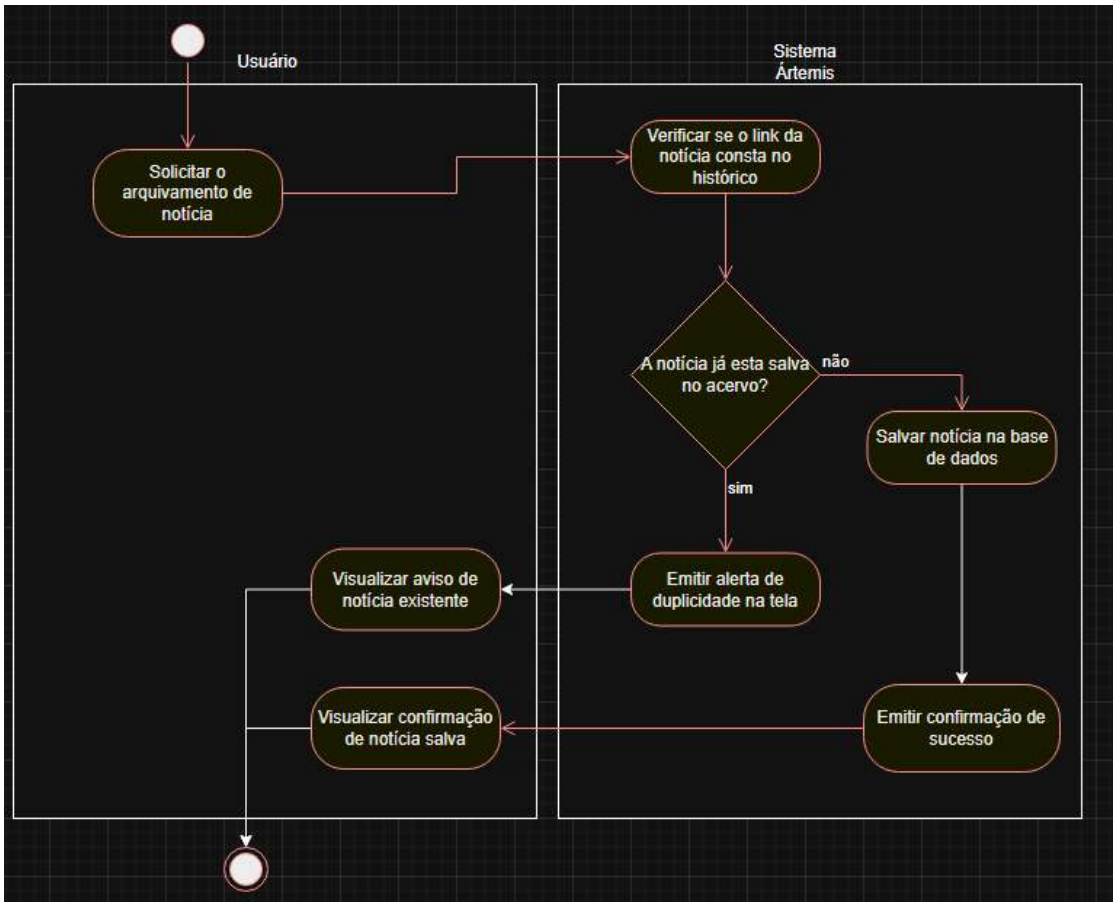
Fonte: Elaborado pelos autores

O segundo Diagrama de Atividades (Figura 3) detalha a rotina de persistência e integridade de dados do acervo. O fluxo evidencia o mecanismo automatizado de checagem contra duplicidade baseado na URL da notícia, dividindo o comportamento do software entre a rejeição do registro (com emissão de alerta visual) ou o salvamento definitivo do conteúdo inédito no banco de dados.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Figura 3 - Diagrama de Atividades 2



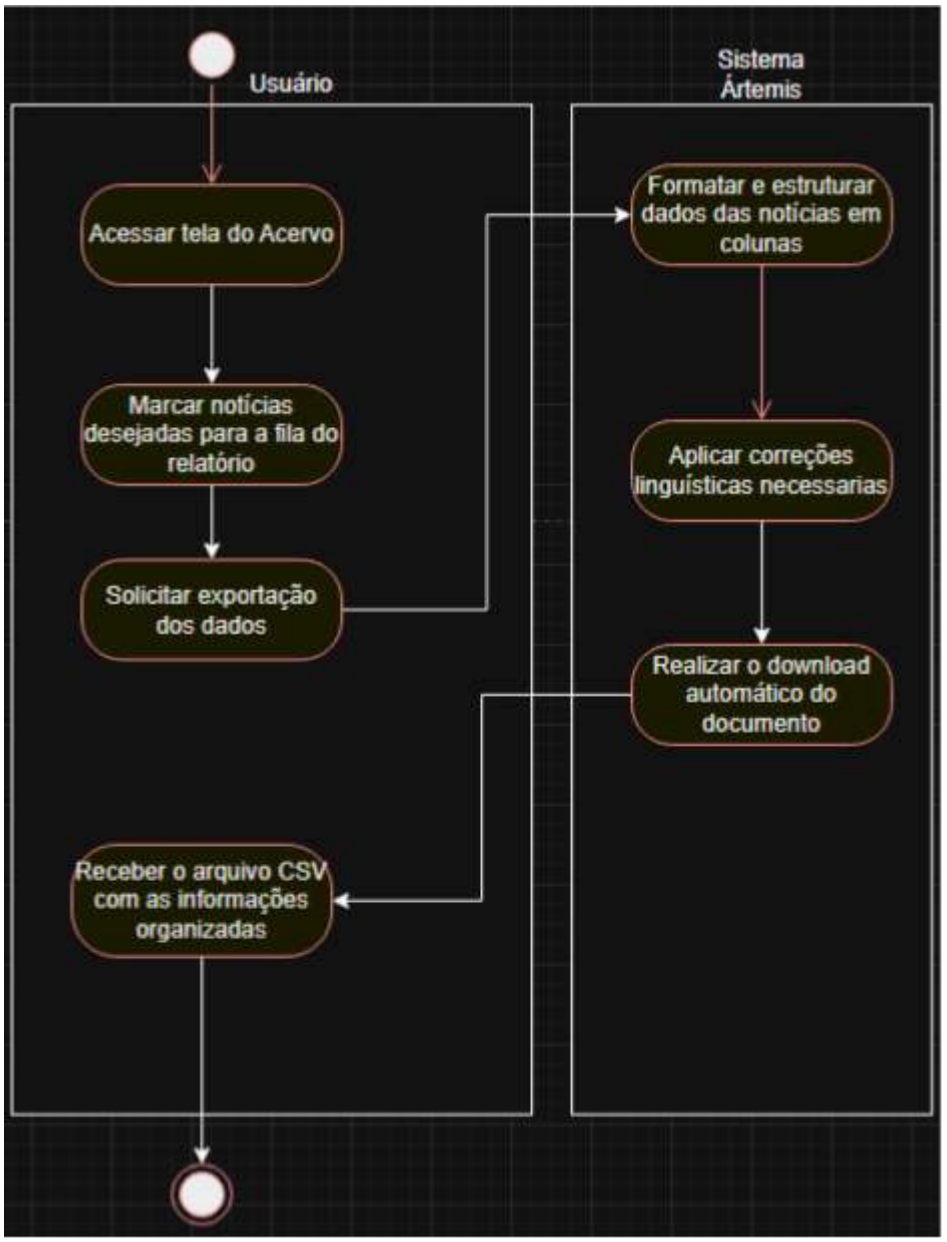
Fonte: Elaborado pelos autores

O terceiro Diagrama de Atividades (Figura 4) representa o procedimento de extração de dados para relatórios. O processo descreve sequencialmente a triagem manual dos registros pelo pesquisador, o tratamento linguístico de codificação para mitigar erros de acentuação em planilhas eletrônicas e a entrega final do arquivo estruturado em formato CSV.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Figura 4 - Diagrama de Atividades 3



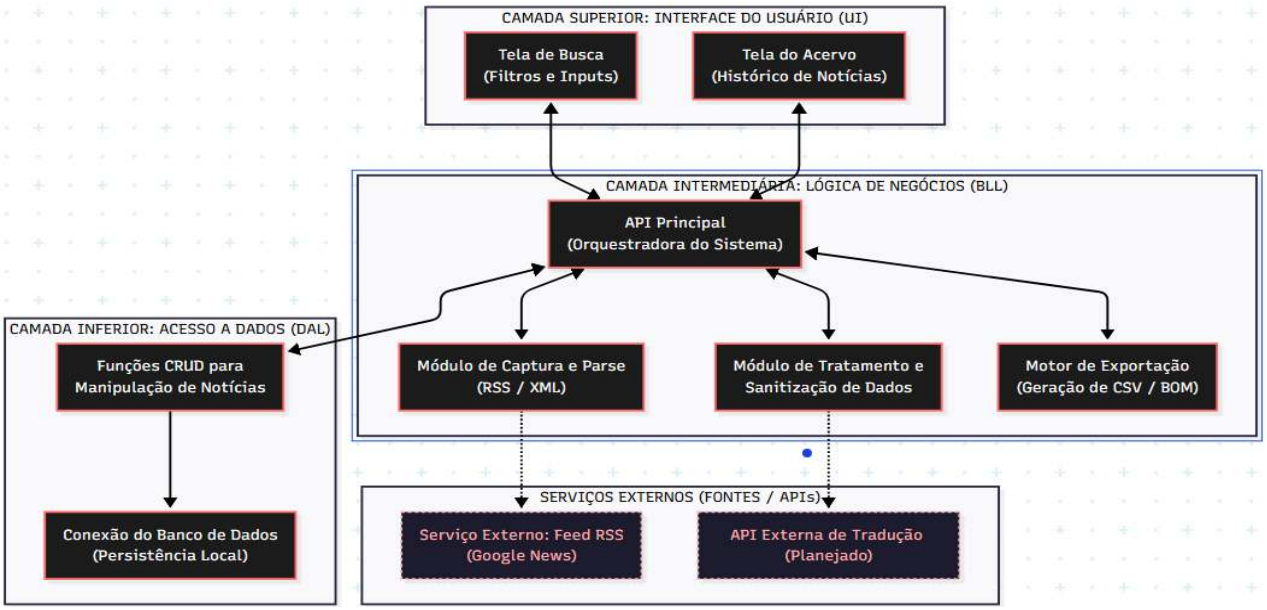
Fonte: Elaborado pelos autores

A Figura 5 apresenta a arquitetura da plataforma, evidenciando a distribuição dos componentes pelas três camadas. Na UI situam-se as telas de Busca e de Acervo; na BLL, a API Principal orquestra os módulos de Captura e Parse (RSS/XML), Tratamento e Exportação; e na DAL concentram-se as funções de manipulação e a conexão com a base local. O serviço externo de tradução representado corresponde a uma funcionalidade planejada (RF03), não validada nesta versão.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Figura 5 - Diagrama de Arquitetura



Fonte: Elaborado pelos autores

A Figura 6 detalha o esquema do banco de dados, composto pelos campos id, titulo, fonte, data_publicacao e url. A restrição UNIQUE aplicada à coluna url constitui o mecanismo que sustenta a deduplicação, impedindo, no nível do banco, a inserção de registros com endereço já indexado.

Figura 6 - Esquema do banco de dados

Esquema do banco de dados

Chave	Nome do Campo	Tipo de dado
PK	id	INTEGER PRIMARY KEY AUTOINCREMENT
	titulo	TEXT
	fonte	TEXT
	data_publicacao	DATETIME
UNIQUE	url	TEXT

Fonte: Elaborado pelos autores

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

4.4 Telas do Sistema

O alinhamento dos requisitos funcionais e não-funcionais da plataforma Ártemis consolidou-se em uma interface gráfica baseada na arquitetura de página única (SPA). O sistema foi estruturado em duas telas principais interconectadas, projetadas para simplificar o fluxo de trabalho dos pesquisadores. A primeira, denominada Tela de Busca (Figura 7), concentra os mecanismos de entrada de dados, permitindo a inserção de termos-chave e a aplicação dos filtros. Essa interface operacionaliza o motor de coleta assíncrono, retornando os resultados em formato de cartões interativos que exibem o título, o veículo de imprensa, a data de publicação e o link de acesso.

Figura 7 – Tela principal do sistema



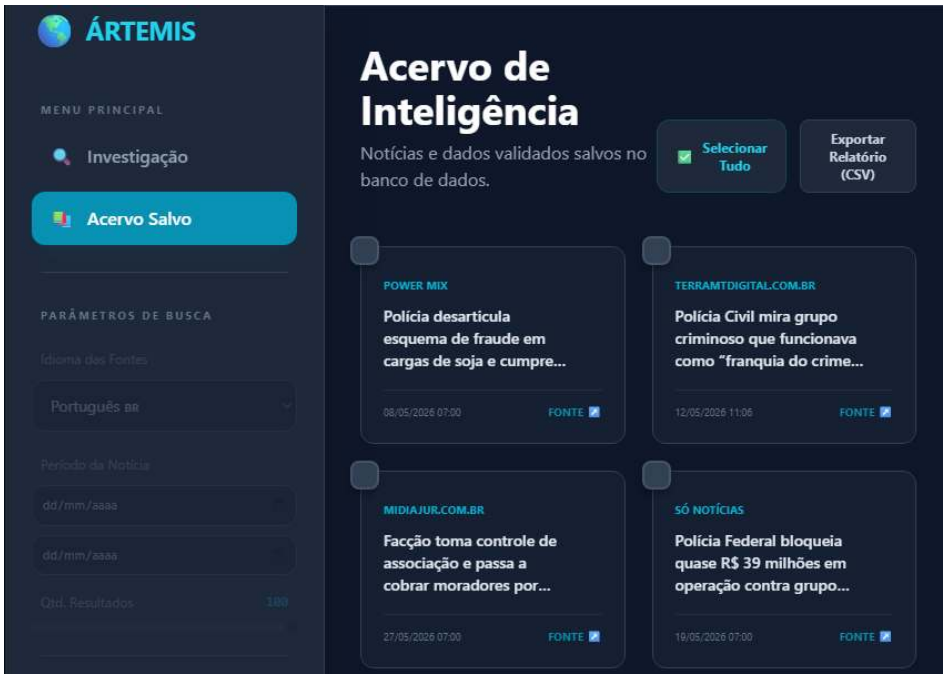
Fonte: Elaborado pelos autores

A segunda interface, denominada Tela do Acervo (Figura 8), funciona como o ambiente de armazenamento da plataforma. Nessa tela, são centralizados todos os registros que o pesquisador selecionou como relevantes para o monitoramento do narcotráfico. A interface disponibiliza ferramentas para a manipulação seletiva dos dados, permitindo ao usuário revisar o conteúdo consolidado e disparar o comando de exportação para gerar o arquivo CSV.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Figura 8 – Tela do Acervo



Fonte: Elaborado pelos autores

4.5 Validação Funcional

A verificação de conformidade técnica do sistema foi conduzida por meio de testes funcionais, que avaliam se o software se comporta conforme as especificações a partir de suas entradas e saídas (SOMMERVILLE, 2018). Para cada requisito funcional, submeteram-se entradas válidas e inválidas, comparando-se a saída observada à esperada, conforme sintetiza a Tabela 5. Os resultados confirmaram a aderência do artefato aos requisitos especificados, à exceção do RF03, tratado como funcionalidade planejada e não validada nesta versão.

Tabela 5 - Matriz de testes funcionais

RF	Cenário (entrada válida)	Entrada inválida	Saída esperada	Resultado
RF01 – Coleta	Termos "cocaína" + “PA”	Campo de busca vazio	Retorno das notícias correspondentes / alerta de campo obrigatório	Validado

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

RF02 – Busca por parâmetros	Uma ou mais palavras-chave	Apenas caracteres especiais	Composição da query (aspas + AND) / validação da entrada	Validado
RF03 - Tradução de parâmetros	-	-	-	Não validado (planejado)
RF04 - Extração de metadadados	Notícia com título, fonte, data e URL	Notícia sem data	Persistência dos quatro metadados / tratamento da ausência	Validado
RF05 - Filtros de busca	Filtro de data e de quantidade	Intervalo de data invertido	Aplicação dos filtros / validação (filtro de idioma planejado)	Validação parcial
RF06 - Persistência	Salvar notícia inédita	Salvar URL ja existente	Gravação / bloqueio da duplicata pela constraint UNIQUE	Validado
RF07 - Visualização	Abrir acervo com N registros	Acervo vazio	Listagem dos registros / exibição de estado vazio	Validado
RF08 - Controle do Acervo	Excluir registro selecionado	Acionar exclusão sem seleção	Execução da ação / bloqueio sem seleção	Validado
RF09 - Seleção em massa	Selecionar todos os registros	Selecionar em lista vazia	Marcação total / Nenhuma ação	Validado
RF10 - Exportação CSV	Exportar acervo preenchido	Exportar acervo vazio	Geração do CSV íntegro / aviso de acervo vazio	Validado

Fonte: Elaborado pelos autores

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

4.6 Caracterização Descritiva do Corpus

Para qualificar o material consolidado, procedeu-se a uma análise descritiva do corpus reunido nas duas fases operacionais (janeiro a maio de 2026), em coerência com a vertente descritiva da pesquisa. O conjunto totalizou 1.000 notícias estruturadas, incremento de 150% em relação aos 400 registros obtidos manualmente no ciclo anterior. Ressalta-se que esse aumento expressa a ampliação da capacidade de coleta, e não, por si só, ganho de qualidade analítica, razão pela qual a composição do acervo é detalhada a seguir.

Quanto à integridade, a deduplicação por URL mostrou-se eficaz: não houve registro com link duplicado e apenas dois títulos idênticos foram observados (0,2%), residuais de republicações sob o mesmo endereço. Reconhece-se, porém, que o controle por URL não captura a duplicidade no nível do evento, quando uma mesma ocorrência é noticiada por veículos distintos sob endereços diferentes, limitação discutida na seção seguinte.

Quanto à diversidade, o corpus reuniu notícias de 280 veículos distintos, com baixa dependência de um único emissor. O G1 concentrou a maior parcela com 213 registros (21,3%) e os nove veículos mais recorrentes (Tabela 6) responderam por 40,9% do total, enquanto 138 veículos contribuíram com uma única notícia cada, configurando uma cauda longa de fontes locais e institucionais, entre elas portais governamentais e corporações de segurança pública estaduais. Esse padrão atenua a preocupação de que a coleta via feed RSS privilegiaria apenas grandes portais.

Tabela 6 - Fontes jornalísticas mais recorrentes

Fonte	**Notícias**	**%**
G1	213	21,3%
GOV.BR	66	6,6%
Polícia Militar do Estado de Rondônia	30	3%
O Liberal	26	2,6%
ac24horas	23	2,3%
CenárioMT	16	1,6%
AM POST	12	1,2%

REMUNOM
ISSN 2178-6925

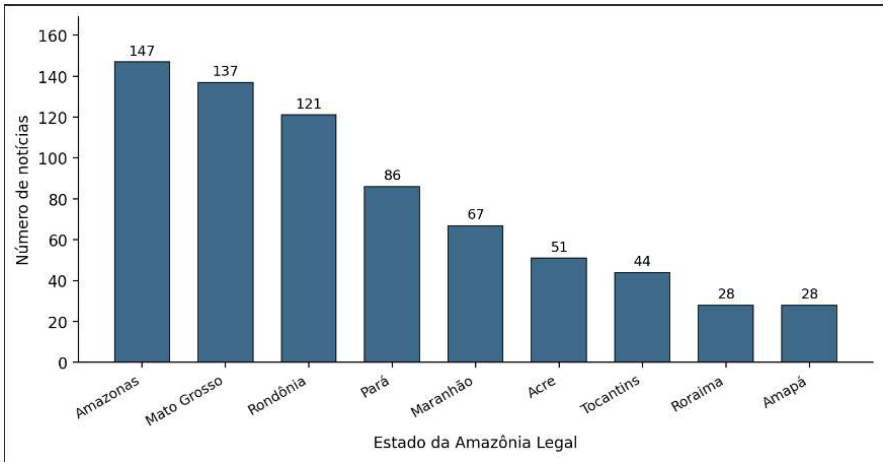
Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

News Rondônia	12	1,2%
LOCAL MT	11	1,1%

Fonte: Elaborado pelos autores

A distribuição territorial contemplou os nove estados da Amazônia Legal (Figura 9). Dos 1.000 registros, 709 foram atribuídos de forma única a um estado por inferência textual a partir do título e da fonte, 81 mencionavam mais de uma unidade federativa e 210 não explicitavam a localidade no título, caso em que a referência espacial decorria do termo de busca. Amazonas (147), Mato Grosso (137) e Rondônia (121) lideraram o volume, ao passo que Roraima (28) e Amapá (28) registraram a menor incidência, refletindo irregularidades na cobertura midiática regional.

Figura 9 - Gráfico de quantidade de notícias por estado



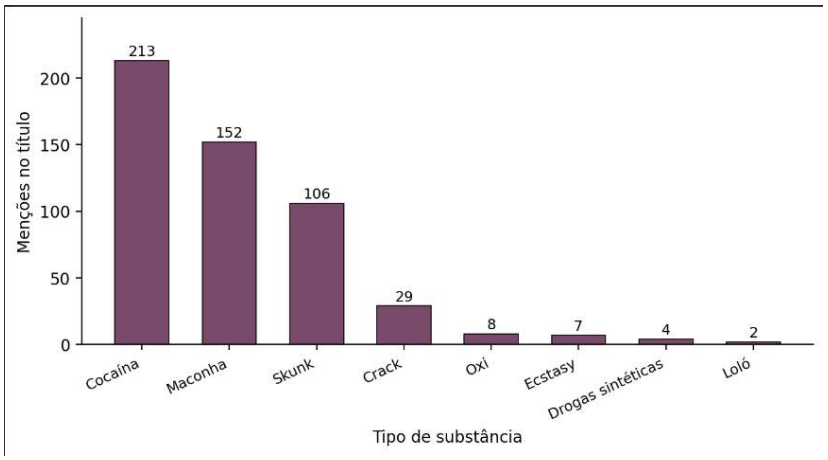
Fonte: Elaborado pelos autores

Em relação às substâncias, 479 títulos (47,9%) não nomeavam o entorpecente, por tratarem de operações ou apreensões genéricas. Entre os que especificaram, predominaram cocaína (213), maconha (152) e skunk (106), seguidas pelo crack (29) e por quantidades residuais de pasta base, oxi e drogas sintéticas (Figura 10). Ressalta-se que tais menções não são mutuamente exclusivas, podendo um mesmo título referir mais de uma substância.

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

Figura 10 - Menções de substância nos títulos



Fonte: Elaborado pelos autores

Em síntese, a plataforma ampliou o volume de coleta preservando diversidade de fontes e abrangência sobre toda a Amazônia Legal, fornecendo um acervo estruturado e heterogêneo apto a subsidiar a posterior triagem e interpretação qualitativa por parte dos pesquisadores.

4.7 Avaliação de Usabilidade

Os resultados indicaram elevada maturidade da interface: oito das dez heurísticas obtiveram avaliação máxima unânime, com destaque para o controle e liberdade do usuário, a consistência visual e o reconhecimento em vez de recordação, atributos que confirmam o caráter intuitivo do fluxo operacional. As duas heurísticas de suporte, recuperação de erros e ajuda e documentação, concentraram as oportunidades de melhoria, sobretudo a inclusão de dicas contextuais e de um manual integrado, recomendações incorporadas ao planejamento de evolução do software, a avaliação de usabilidade esta melhor descrita na tabela 7. A homogeneidade institucional do painel é reconhecida como limitação e discutida na Seção 4.8.

Tabela 7 - Avaliação de usabilidade baseada nas Heurísticas de Nielsen

Heurística de Nielsen	**Pergunta aplicada**	**Distribuição das Respostas**	**Principais Comentários e Implicações para**

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

		(Escala Likert 1 a 5)	Melhoria
H1 - Visibilidade do status do sistema	O sistema me mantém informado de forma clara e rápida sobre o andamento das ações (como carregamentos, buscas e confirmações de salvamento)?	Nota 5: 5 pesquisadores Notas 4 a 1: 0 pesquisadores	Consenso positivo absoluto. O uso de feedbacks visuais de carregamento assíncrono no FastAPI garante clareza. Manter padrão atual.
H2 - Correspondência com o mundo real	Os termos, ícones e palavras utilizados nas telas são familiares e fáceis de compreender?	Nota 5: 5 pesquisadores Notas 4 a 1: 0 pesquisadores	Interface adequada ao jargão da pesquisa em segurança pública e inteligência. Sem necessidade de alterações.
H3 - Controle e liberdade do usuário	É fácil desfazer uma ação, voltar para a tela anterior ou corrigir um comando caso eu clique em algo por engano?	Nota 5: 5 pesquisadores Notas 4 a 1: 0 pesquisadores	Unanimidade. O fluxo de navegação e as rotas permitem retrocessos fáceis sem aprisionar o usuário.
H4 - Consistência e padrões	O sistema mantém uma padronização visual e funcional (os botões e	Nota 5: 5 pesquisadores Notas 4 a 1: 0 pesquisadores	Identidade visual íntegra e reaproveitamento de componentes em todas as camadas da

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

	componentes significam a mesma coisa em todas as telas)?		UI.
H5 - Prevenção de erros	A interface do sistema é intuitiva e projetada de forma a evitar que eu cometa erros por distração durante o uso?	Nota 5: 5 pesquisadores Notas 4 a 1: 0 pesquisadores	Bloqueios prévios na interface (como restrição de campos e botões desativados antes do preenchimento) evitam falhas operacionais.
H6 - Reconhecimento em vez de recordação	As opções de menus, filtros e ações estão visíveis e fáceis de encontrar, sem que eu precise memorizar o caminho?	Nota 5: 5 pesquisadores Notas 4 a 1: 0 pesquisadores	Redução da carga cognitiva. Menus fixos e barras laterais de filtragem eliminam a necessidade de decorar caminhos de busca.
H7 - Flexibilidade e eficiência de uso	O sistema responde rápido aos comandos e permite realizar as tarefas (como buscas e exportações) de forma ágil e eficiente?	Nota 5: 5 pesquisadores Notas 4 a 1: 0 pesquisadores	Os avaliadores destacaram a expressiva redução de tempo na coleta e qualificação de dados.
H8 - Estética e design minimalista	O design das telas é limpo, agradável e apresenta	Nota 5: 5 pesquisadores Notas 4 a 1: 0	Estética minimalista validada. Foco exclusivo nos

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

	apenas as informações e botões necessários, sem poluição visual?	pesquisadores	metadados necessários para a pesquisa, reduzindo ruídos visuais.
H9 - Reconhecer, diagnosticar e recuperar erros	Quando ocorre algum problema no sistema, as mensagens emitidas são claras e me ajudam a entender como resolver a situação?	Nota 5: 3 pesquisadores Nota 4: 2 pesquisadores Notas 3 a 1: 0 pesquisadores	Retorno predominantemente positivo. Melhoria considerada: Tratar exceções de rede no back-end para exibir mensagens ainda mais amigáveis ao usuário final quando ocorre uma falha.
H10 - Ajuda e documentação	Se necessário, o sistema oferece orientações, dicas contextuais ou instruções claras sobre como utilizá-lo?	Nota 5: 3 pesquisadores Nota 3: 2 pesquisadores Notas 2 a 1: 0 pesquisadores	Ponto de atenção moderado. Implicação de melhoria: Inclusão de tooltips (dicas flutuantes ao passar o mouse) e um manual rápido integrado em PDF para sanar dúvidas contextuais.

Fonte: Elaborado pelos autores

4.8 Limitações

Apesar dos resultados alcançados, o estudo apresenta limitações que delimitam suas conclusões e indicam caminhos futuros. A principal reside na dependência estrutural do feed RSS do Google News: por consumir um agregador, a plataforma sujeita-se aos critérios de indexação, ranqueamento e disponibilidade de terceiros, o que pode introduzir vieses de cobertura, sub-representar veículos

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

locais pouco indexados e limitar a reprodutibilidade caso o serviço seja alterado. Soma-se a isso a natureza não neutra das fontes jornalísticas, que dependem de agenda midiática, visibilidade regional e releases policiais (MIRANDA; MUNIZ, 2022), assim, a distribuição territorial observada deve ser lida como expressão da cobertura midiática, e não como medida direta da incidência do fenômeno.

No plano técnico, a deduplicação por URL, embora eficaz contra registros idênticos, não captura a duplicidade no nível do evento, ou seja, quando uma mesma ocorrência é noticiada por veículos distintos sob endereços diferentes. Sua consolidação permanece dependente de triagem humana, pois o sistema não realiza análise semântica automatizada: a extração de entidades, a geocodificação e a interpretação de padrões continuam a cargo dos pesquisadores, em coerência com o escopo de infraestrutura de coleta assumido pela ferramenta. Soma-se ainda o suporte à coleta multilíngue (RF03), previsto nos requisitos e diagramas, mas tratado como funcionalidade planejada e não validada, dado que os testes nessa frente não produziram resultados maduros para incorporação ao escopo atual.

Por fim, a avaliação de usabilidade contou com cinco pesquisadores seniores de uma mesma instituição. Tal recorte é adequado à validação contextual do estudo de caso, mas limita a generalização a outros perfis e organizações, podendo trabalhos futuros ampliar a amostra para usuários juniores, analistas de dados e avaliadores externos.

5. Conclusão

O presente trabalho teve como objetivo desenvolver e validar uma plataforma voltada à automação da coleta de dados jornalísticos sobre o narcotráfico e a apreensão de drogas ilícitas na Amazônia, integrada ao fluxo de pesquisa do Instituto Mãe Crioula, em resposta à questão de como uma solução de software poderia otimizar a triagem qualitativa dessas informações.

Os resultados demonstraram que a plataforma cumpriu seu propósito. A validação funcional confirmou a aderência aos requisitos especificados, e a caracterização descritiva do corpus evidenciou a capacidade de consolidar, em volume e de forma estruturada, o material necessário ao fluxo de pesquisa, viabilizando a extração de 1.000 notícias, equivalente ao incremento de 150%

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

sobre os 400 registros do ciclo anterior. A avaliação de usabilidade indicou maturidade, confiabilidade técnica e utilidade prática sob a perspectiva dos usuários.

Como contribuição científica e tecnológica, o Ártemis oferece uma arquitetura replicável em camadas (Python, FastAPI, feedparser e SQLite), demonstrando a viabilidade do uso de fontes abertas (OSINT) estruturadas via RSS para o monitoramento contínuo de dinâmicas criminais em contextos de dados massivos e dispersos, já integrada à produção da 5ª edição do relatório "Cartografias das Violências na Amazônia" (2026).

Reconhecem-se limitações: a dependência do feed RSS do Google News restringe a coleta aos conteúdos indexados, podendo excluir fontes locais de menor visibilidade algorítmica; e a ausência de Processamento de Linguagem Natural (PLN) para extração de entidades ainda exige intervenção manual na análise. Para trabalhos futuros, sugerem-se: (i) módulos de PLN para classificação automática e extração de metadados semânticos; (ii) a expansão da matriz de busca para diários oficiais e boletins de segurança pública; (iii) um dashboard analítico para visualização espaço-temporal dos eventos; e (iv) a adaptação da arquitetura a outros temas, como crimes ambientais e violência contra populações vulneráveis.

Em suma, o Ártemis demonstra que a Engenharia de Software pode oferecer respostas concretas e escaláveis a problemas metodológicos das ciências sociais aplicadas, ampliando a capacidade de monitoramento informacional e fortalecendo a produção de conhecimento sobre a segurança pública na Amazônia.

Referências

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 20 jun. 2026

COUTO, Aiala Colares. Geografia das redes do narcotráfico na Amazônia. **Revista GeoAmazônia**, Belém, v. 11, n. 22, p. 46-67, 2023. Disponível em: https://periodicos.ufpa.br/index.php/geoamazonia/article/view/13828. Acesso em:

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

28 abr. 2026

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA (FBSP). **Anuário Brasileiro de Segurança Pública 2025**. São Paulo, 2025. Disponível em: https://publicacoes.forumseguranca.org.br/items/c3605778-37b3-4ad6-8239-94e4cb236444. Acesso em: 27 abr. 2026.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA (FBSP); INSTITUTO MÃE CRIOULA (IMC). **Cartografias das Violências na Amazônia**. 4. ed. São Paulo: FBSP; Belém: IMC, 2025. Disponível em: https://forumseguranca.org.br/wp-content/uploads/2025/11/cartografias-violencia-amazonia-2025.pdf. Acesso em: 27 abr. 2026.

FOWLER, Martin. **Presentation Domain Data Layering.** [S.l.], 2015. Disponível em: https://martinfowler.com/bliki/PresentationDomainDataLayering.html. Acesso em: 15 mai. 2026.

FLEURY, Maria Tereza Leme; WERLANG, Sérgio R. C. **Pesquisa aplicada**: conceitos e abordagens. Anuário de Pesquisa GVpesquisa. 2017. Disponível em: https://periodicos.fgv.br/apgvpesquisa/article/view/72796/69984. Acesso em: 15 mai 2026.

FURTADO, Vasco. Inteligência artificial na segurança pública: conceitos, perspectivas e desafios. In: LIMA, Renato Sérgio de (org.); BARROS, Betina Warmling (org assist.). **Estatísticas de segurança pública**: produção e uso de dados criminais no Brasil. São Paulo: Fórum Brasileiro de Segurança Pública, 2022. cap. 10, p. 178-189.

GUEDES, Gilleanes T. A. **UML 2**: uma abordagem prática. 2. ed. São Paulo: Novatec, 2011

GUERRA, A. de L. e R. M**etodologias e classificação das pesquisas científicas**. Recima21, [S. l.], 2024. Disponível em: https://recima21.com.br/index.php/recima21/article/view/5584. Acesso em: 15 mai. 2026

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

GRACIANO, Helton Luiz dos Santos. **ScraperCI**: um protótipo de web scraper para coleta de dados. 2022. 79 f. Dissertação (Pós Graduação em Ciência da Informação). Universidade Federal de São Carlos. São Carlos, 2022. Disponível em: https://repositorio.ufscar.br/server/api/core/bitstreams/796636e7-3349-4ecb-8f8b-b19621d7da3c/content. Acesso em: 15 mai. 2026.

HU, Chuanbo; YIN, Minglei; LIU, Bin; LI, Xin; YE, Yanfang. **Detection of Illicit Drug Trafficking Events on Instagram: A Deep Multimodal Multilabel Learning Approach.** In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 30., 2021, New York, 2021. 9 p. Disponível em: https://dl.acm.org/doi/10.1145/3459637.3481908. Acesso em: 9 jun. 2026.

INSTITUTO MÃE CRIOULA (IMC). **Sobre o instituto**. Belém, 2023. Disponível em: https://institutomaecrioula.org/sobre-o-instituto/. Acesso em: 27 abr. 2026.

LEITE, Sara Souza. **O Emprego das Fontes abertas no Âmbito da Atividade de Inteligência Policial**. Revista Brasileira de Ciências Policiais. Brasília, v. 5, n. 1, p. 11-45, 2014. Disponível em: https://www.sindipoldf.org.br/arquivos/93bea178e0.pdf. Acesso em: 29 abr. 2026

MITCHELL, Ryan. **Web Scraping com Python:** Coletando mais dados na web moderna. 2. ed. Sebastopol: O'Reilly Media, 2018.

MIRANDA, Ana Paula Mendes de; MUNIZ, Jacqueline de Oliveira. Mudar para ficar igual? A cartorialização da informação e seus efeitos na gestão da segurança pública. In: LIMA, Renato Sérgio de (org.); BARROS, Betina Warmling (org. assist.). **Estatísticas de segurança pública**: produção e uso de dados criminais no Brasil. São Paulo: Fórum Brasileiro de Segurança Pública, 2022. cap. 5, p. 102-121.

NIELSEN, Jakob. Enhancing the explanatory power of usability heuristics. In: **Conference on human factors in computing systems**, 1994, Boston. New York: Association for Computing Machinery, 1994. p. 152-158. Disponível em: https://dl.acm.org/doi/10.1145/191666.191729. Acesso em: 20 jun. 2026

SÁTYRO, Natália Guimarães Duarte; D'ALBUQUERQUE, Raquel Wanderley. O

REMUNOM
ISSN 2178-6925

Received: 24/05/2026 - **Accepted:** 30/06/2026
Vol: 13.14
DOI: 10.66104/3ntg3y15
Pages: 1-34

que é um estudo de caso e quais as suas potencialidades. **Sociedade e Cultura**, Goiânia, v. 23, 2020. Disponível em: https://revistas.ufg.br/fcs/article/view/55631. Acesso em: 15 mai. 2026.

SOMMERVILLE, Ian. **Engenharia de Software**. 10. ed. São Paulo: Pearson, 2018

SILVA, Everson Esteves da; LIMA Danielli Araújo. **Automação de coleta de dados:** benefícios e aplicação prática. 11º Encontro de Pesquisa e Extensão (EnPE). Minas Gerais, 2024. Disponível em: https://www.researchgate.net/publication/390797651_AUTOMACAO_DE_COLETA_DE_DADOS_BENEFICIOS_E_APLICACAO_PRATICA. Acesso em: 30 abr. 2026

UNITED NATIONS OFFICE ON DRUGS AND CRIME (UNODC). **The Drugs-Crime Nexus in the Amazon Basin**: How a complex crime ecosystem is endangering the world's largest rainforest and imperilling efforts to combat climate change. Vienna, 2023. Disponível em: https://igarape.org.br/wp-content/uploads/2024/01/sc_en_the-drugs-crime-nexus-in-the-amazon-basin.pdf. Acesso em: 28 abr. 2026.

UNITED NATIONS OFFICE ON DRUGS AND CRIME (UNODC). **World Drug Report 2025 Methodological Annex**. Vienna, 2025. Disponível em: https://www.unodc.org/documents/data-and-analysis/WDR_2025/Annex/WDR2025_Methodological_Annex.pdf. Acesso em: 27 abr. 2026

VALDIVIA-GRANDA, W. A. Structure and dynamics mapping of illicit firearms trafficking using artificial intelligence models. **Frontiers in Big Data**, [S. l.], v. 8, e1648730, 2025. Disponível em: https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2025.1648730/full. Acesso em: 9 jun. 2026.