

UNIMAR TECH SUMMIT

II MOSTRA DE INICIAÇÃO
CIENTÍFICA E EXTENSÃO

CADERNO DE RESUMOS

2026

Aplicação de Redes Neurais Convolucionais na Detecção Precoce do Câncer de Mama em Mamografias

Application of Convolutional Neural Networks in the Early Detection of Breast Cancer in Mammograms

Guilherme Lopes Bertacini – guilb2005@gmail.com
Henrique Leal Tavares – henriquetavares@unimar.br

INTRODUÇÃO

O câncer de mama é a neoplasia de maior incidência e mortalidade entre as mulheres (WHO, 2024; INCA, 2022). O rastreamento periódico por mamografia digital é o principal método de diagnóstico precoce, elevando a sobrevida a mais de 90% (SBOC, 2023). Contudo, a alta densidade tecidual e contornos lesivos sutis geram considerável variabilidade interobservador na interpretação radiológica (Siqueira; Koch; Paquete, 2018). Nesse cenário, sistemas CAD baseados em Redes Neurais Convolucionais (CNNs) atuam como suporte crucial (Litjens et al., 2017). Este trabalho avalia a hipótese de que a engenharia de dados offline — combinando isolamento anatômico da mama (cropping) e equalização adaptativa de contraste local (CLAHE) — potencializa as cabeças em cascata da Cascade R-CNN (Cai; Vasconcelos, 2018) ao eliminar gradientes nulos de fundo e realçar transições glandulares. O objetivo foi quantificar o impacto desse pré-processamento na detecção de lesões na base INbreast (Moreira et al., 2012) frente a dados não tratados.

DESENVOLVIMENTO

A pesquisa utilizou estação Ubuntu 22.04 LTS com GPU NVIDIA Orin, Python 3.10, PyTorch 2.5 e MMDetection v3.3.0. Adotou-se o conjunto público INbreast (Moreira et al., 2012), com mamografias digitais diretas (FFDM) e anotações validadas por biópsia em formato COCO. Empregou-se a Cascade R-CNN por sua arquitetura em múltiplos estágios com limiares crescentes de IoU, propiciando refinamento iterativo das caixas delimitadoras. A rede operou com backbone ResNet-50 e FPN, estágios iniciais congelados (frozen_stages=2), otimizador AdamW ($\text{lr}=1 \times 10^{-4}$), decay de 0,05 e Cosine Annealing por 100 épocas.

Avaliaram-se dois cenários de treinamento: o Experimento 1 (Base), com imagens brutas (640x640 pixels); e o Experimento 2, sob engenharia de dados offline.

No Experimento 2, aplicou-se cropping automático (varredura vertical a 20% da altura para orientar a mama, corte 30, margem de 200 pixels e ajuste das caixas COCO) e equalização CLAHE (clip limit 2.0, blocos 8x8) na mama isolada. No Experimento 1, a perda convergiu na 100ª época com $\text{mAP}_{50} = 0,0820$ ($\text{mAP}_s = 0,0280$; $\text{mAP}_m = 0,2050$; $\text{mAP}_l = 0,3000$). No Experimento 2, o ápice ocorreu na 79ª época: embora $\text{mAP}_{50} = 0,0610$ pela menor resposta a pequenas estruturas ($\text{mAP}_s = 0,0130$), obteve-se salto nas lesões consolidadas: $\text{mAP}_m = 0,3030$ (+9,8% absolutos) e $\text{mAP}_l = 0,3570$ (+5,7% absolutos). O cropping eliminou os gradientes nulos de fundo, focando as convoluções no tecido glandular; a retração em microcalcificações aponta para a calibragem de âncoras na RPN.

CONCLUSÃO

Os objetivos foram plenamente atingidos, comprovando que a engenharia de

dados offline é decisiva no desempenho da Cascade R-CNN em mamografia. Os expressivos ganhos de até 9,8% em lesões médias e grandes confirmam a relevância clínica do pipeline na triagem assistida e na mitigação da variabilidade interobservador. Como trabalhos futuros, prevê-se o ajuste de escala de âncoras na RPN para microlesões e a extensão multibase via Transfer Learning para superar o deslocamento de domínio radiológico.

REFERÊNCIAS

CAI, Z.; VASCONCELOS, N. Cascade R-CNN: delving into high quality object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2018, Salt Lake City. Proceedings [...]. Salt Lake City: IEEE, 2018. p. 6154-6162.

INSTITUTO NACIONAL DE CÂNCER (INCA). Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro: INCA, 2022.

LITJENS, G. et al. A survey on deep learning in medical image analysis. Medical Image Analysis, v. 42, p. 60-88, 2017.

MOREIRA, I. C. et al. INbreast: toward a full-field digital mammographic database. Academic Radiology, v. 19, n. 2, p. 236-248, 2012.

SIQUEIRA, S. A. C.; KOCH, H. A.; PAQUETE, A. D. Variabilidade interobservador na classificação BI-RADS® de lesões mamárias. Radiologia Brasileira, v. 51, n. 6, p. 383-388, 2018.

SOCIEDADE BRASILEIRA DE ONCOLOGIA CLÍNICA (SBOC). Detecção precoce do câncer de mama pode aumentar em até 90% as chances de cura. [S.l.]: SBOC, 2023.

WORLD HEALTH ORGANIZATION (WHO). Breast cancer. Geneva: WHO, 2024.

CAA Inteligente: processamento de linguagem natural em comunicação alternativa para uso por pessoas neurodivergentes

Smart AAC: natural language processing in augmentative and alternative communication for use by neurodivergent individuals

Amanda Villar Moura – amandavillarmoura@gmail.com
Douglas Rodrigues – douglasrodrigues@unimar.br

INTRODUÇÃO

Crianças e adultos não verbais utilizam sistemas de Comunicação Aumentativa e Alternativa (CAA), que são geralmente baseados em cartões de figuras (Picture Exchange Communication System - PECS) (Soares; Barreto; Cunha, 2024). Atualmente, a maioria dos aplicativos da área é estática e limitada a botões de ordem fixa, o que reduz a dinâmica das interações (Elsahar et al., 2019) e evidencia a necessidade de soluções tecnológicas mais eficientes. Diante disso, este projeto tem como objetivo desenvolver e testar um aplicativo de CAA inteligente para pessoas neurodivergentes utilizando Processamento de Linguagem Natural (PLN). Especificamente, busca-se aplicar um modelo preditivo, adaptado ao contexto brasileiro, capaz de cruzar variáveis de contexto momentâneo (horário, localização e histórico das três últimas palavras) para antecipar intenções de fala. Ao sugerir dinamicamente os pictogramas mais prováveis no topo da tela, a ferramenta visa proporcionar uma comunicação mais rápida, fluida e menos frustrante.

DESENVOLVIMENTO

A pesquisa inicia-se com a análise dos aplicativos de CAA existentes no mercado e o estudo do método DHACA (Desenvolvimento das Habilidades de Comunicação no Autismo) (Montenegro et al., 2024). Esse referencial teórico fundamenta a lógica de sugestão dinâmica de pictogramas, alinhada às abordagens de geração automática de sequências de pictogramas propostas por Pereira et al. (2017). Na etapa prática, o tratamento computacional dos dados de entrada (horário, localização e histórico de uso) será realizado em Python, utilizando a biblioteca Pandas. A partir dessas variáveis, estrutura-se a relação lógica entre cenários cotidianos e o vocabulário pertinente à realidade brasileira. Em seguida, com o uso da biblioteca Scikit-learn, o algoritmo preditivo de PLN será treinado para identificar padrões de comunicação e retornar sempre as cinco imagens mais adequadas ao contexto do usuário. Por fim, serão realizadas análises para validar a eficácia do modelo preditivo e comprovar se ele atinge o objetivo de otimizar a interação comunicativa.

CONCLUSÃO

Ao final deste projeto, espera-se que a aplicação da inteligência artificial preditiva torne a comunicação de pessoas neurodivergentes muito mais rápida e fluida. Através de um algoritmo capaz de prever intenções de comunicação com base no contexto (horário, localização e histórico), a ferramenta deixará de ser estática para se tornar ativamente adaptativa. Dessa forma, o projeto visa proporcionar uma maior autonomia aos usuários, garantindo uma ferramenta de acessibilidade muito mais responsiva e adequada às suas reais necessidades comunicativas no dia a dia, com perspectivas de validar a eficácia do método preditivo aplicado ao contexto brasileiro.

REFERÊNCIAS

ELSAHAR, Y. et al. Augmentative and alternative communication (AAC) advances: a review of configurations for individuals with a speech disability. *Sensors*, Basel, v. 19, n. 8, art. 1911, p. 1-24, 2019.

MONTENEGRO, A. C. A. et al. Método de desenvolvimento das habilidades de comunicação no autismo – DHACA: validação da aparência e do conteúdo. *CoDAS*, v. 36, n. 3, e20230138, 2024.

PEREIRA, R. et al. Geração automática de sentenças em língua natural para sequências de pictogramas como apoio à comunicação alternativa e ampliada. *LinguaMÁTICA*, v. 9, n. 1, p. 31-39, 2017.

SOARES, Q.; BARRETO, L. M.; CUNHA, C. O. Comunicação alternativa: tecnologia assistiva e desenvolvimento da linguagem. *Humanas Sociais & Aplicadas*, [S.l.], v. 14, n. 43, p. 84-99, 2024.

Ciência de Dados Aplicada ao Monitoramento de Meteoros: classificação automática de eventos por câmera estática via aprendizado de máquina

Data Science Applied to Meteor Monitoring: automatic event classification from static camera footage via machine learning

Ana Júlia Pereira Romera – anajuliapromera@gmail.com
Caio Saraiva Coneglian – caioconeglian@unimar.br

INTRODUÇÃO

O estudo de meteoros por meio de câmeras estáticas tem se consolidado como uma metodologia eficaz e acessível na astronomia observacional, fornecendo informações importantes sobre a dinâmica e composição de pequenos corpos do Sistema Solar (Jenniskens, 2006). Softwares de detecção automática, como o UFO Capture (SonotaCo, [s.d.]), viabilizam o monitoramento contínuo por meio da análise de diferenças entre quadros consecutivos de vídeo (Molau; Gural, 2005). Contudo, esses sistemas geram um volume expressivo de falsos positivos, registrando artefatos de compressão, insetos, satélites artificiais e relâmpagos, o que exige extensa revisão manual e cria um gargalo para a produção científica.

O presente projeto tem o objetivo geral de desenvolver uma pipeline automatizada de classificação desses eventos luminosos utilizando técnicas de aprendizado de máquina. O estudo justifica-se pela necessidade de otimizar a curadoria de dados do observatório e viabilizar a escalabilidade de sistemas de monitoramento. A abordagem metodológica, em fase de implementação, abrange a coleta contínua, a extração de características e a estruturação de um classificador inteligente.

DESENVOLVIMENTO

Os materiais empregados na pesquisa compreendem a câmera estática de monitoramento de meteoros do observatório da UNIMAR, o software de detecção UFO Capture (SonotaCo) e um computador dedicado ao processamento, com linguagem Python (versão 3.10 ou superior) e as bibliotecas OpenCV, NumPy, Pandas, Scikit-learn e Matplotlib.

O desenvolvimento técnico da arquitetura de software e a definição metodológica constituem os resultados parciais alcançados até o momento. A primeira etapa consistirá na operação do sistema de captura, projetado para gerar arquivos de vídeo (.avi) e metadados (.xml) a cada evento luminoso detectado. Sua implantação encontra-se em andamento, havendo um obstáculo técnico de compatibilidade ainda em resolução, o que precede o início efetivo da coleta de dados.

Como entrega estrutural, definiu-se a construção de um dataset rotulado por inspeção manual dos eventos, categorizando-os nas classes: meteoro, satélite, inseto, relâmpago e artefato de câmera. O processamento planejado extrairá características cinemáticas, temporais, morfológicas e fotométricas das imagens via OpenCV. Para a etapa analítica, a pipeline definiu o Random Forest (Breiman, 2001), via Scikit-learn (Pedregosa et al., 2011), como classificador primário. O treinamento e a validação aplicarão validação cruzada estratificada e técnicas para mitigar desbalanceamentos sintéticos entre classes (Chawla et al., 2002), com avaliação por acurácia, precisão, recall, F1-score e matriz de confusão.

CONCLUSÃO

A arquitetura delineada demonstra a viabilidade técnica de integrar conceitos de visão computacional para mitigar o problema dos falsos positivos na astronomia observacional. Como perspectivas de continuidade, o projeto avançará, primeiramente, para a resolução do obstáculo técnico que ainda impede a operação estável do sistema de captura e, na sequência, para a rotulagem do dataset, o treinamento do modelo classificador e a avaliação rigorosa do seu desempenho. Espera-se, com a conclusão da pesquisa, consolidar uma ferramenta funcional aplicável e disponibilizar um dataset rotulado inédito à comunidade científica brasileira.

REFERÊNCIAS

BREIMAN, L. Random forests. Machine Learning, v. 45, n. 1, p. 5-32, 2001.

CHAWLA, N. V. et al. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, v. 16, p. 321-357, 2002.

JENNISKENS, P. Meteor showers and their parent bodies. Cambridge: Cambridge University Press, 2006.

MOLAU, S.; GURAL, P. A review of video meteor detection software. WGN, Journal of the International Meteor Organization, v. 33, n. 1, p. 15-20, 2005.

PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research, v. 12, p. 2825-2830, 2011.

SONOTACO. UFO Capture / UFO Analyzer user guide. [S.l.]: SonotaCo, [s.d.].

Construção Otimizada de Grafos na biblioteca NetFog

Optimized Graph Construction in NetFog library

Tiago Fogolin Ragassi – ticofogolin@gmail.com
Rafael Gutierrez Castanha – rafaelcastanha@unimar.br

INTRODUÇÃO

Grafos de larga escala sustentam aplicações que crescem em volume e criticidade, de motores de recomendação a grafos de conhecimento usados por sistemas de IA, passando por análise de redes sociais e bioinformática. Nesses contextos, o tempo de construção do grafo compete com o de consulta e treinamento, tornando a etapa de ingestão um alvo relevante de otimização.

A NetFog é uma biblioteca de grafos para Python, escrita em Rust via PyO3, que permite a escolha da estrutura de dados mais adequada a cada caso de uso. Este trabalho foca na otimização da etapa de construção do grafo, buscando aproximar ou superar o tempo de bibliotecas consolidadas da área, como *igraph* e *Networkx*.

O objetivo geral é avaliar o quanto camadas de abstração, como rótulos textuais e extração genérica via Foreign Function Interface (FFI), custam em desempenho frente a essas bibliotecas. Os objetivos específicos são: (i) escrever o algoritmo da Compressed Sparse Row (CSR) em complexidade $O(V+E)$; (ii) avaliar os trade-offs de otimizar a fronteira de Foreign Function Interface (FFI) entre Python e Rust; e (iii) quantificar o custo de uma API baseada em rótulos textuais (labels) frente a uma de inteiros puros.

DESENVOLVIMENTO

A CSR foi escolhida para este trabalho por oferecer o melhor desempenho de travessia do motor, já que as arestas de cada vértice ficam ordenadas e contíguas na memória (Beamer; Asanović; Patterson, 2015), o que também explica seu custo de construção mais alto, pois exige conhecer a posição final de cada aresta antes de escrevê-la. Por isso, o recorte deste estudo é a construção estática, não incremental. O algoritmo foi escrito usando contagem de grau seguida por soma de prefixos de GVEL (Sahu; Kothapalli, 2024): o grau final de cada vértice soma arestas existentes e novas; a soma de prefixos define a posição final de cada aresta, permitindo alocação única e exata via cursor por vértice.

Na fronteira de FFI, a extração automática do PyO3 usa protocolo genérico com verificação de tipo por elemento. Na geração em lote, isso foi substituído por acessos por referência emprestada, sem cópia, ao custo de menos verificação e risco de comportamento inesperado.

O benchmark usou grafo direcionado de 2.000.000 de vértices e 30.000.000 de arestas, com 5 execuções por variante, isoladas em processos próprios para evitar estado compartilhado do alocador. Foram comparadas as ingestões da NetFog por rótulos e por inteiros puros, cada uma utilizando a PyO3 sem alterações e com a FFI modificada.

A ingestão por inteiros com a NetFog levou 4,06s em média, 3,07x mais rápida que a melhor biblioteca comparada, *igraph* (12,44s em média). A por rótulos levou 16,21s, 1,30x mais lenta que essa mesma referência. A diferença entre as variantes da NetFog quantifica em ~4,00x o custo da camada de rótulos. A FFI modificada superou o PyO3 puro em ambos os casos (1,70x rótulos; 1,35x ids), confirmando o custo

mensurável da fronteira de FFI.

CONCLUSÃO

Os resultados atendem aos objetivos propostos: a CSR foi escrita em $O(V+E)$, e a ingestão por id puro superou a melhor biblioteca comparada, enquanto a construção com rótulos foi mais lenta, como esperado. Em ambos os casos, a FFI modificada superou a extração automática do PyO3. Esse padrão sugere que, quanto mais próxima do dado bruto a interface, melhor o desempenho de construção: cada camada de abstração adicionada (rótulos textuais, verificação automática, extração genérica) tem custo mensurável. Como limitação, a redução de verificações que sustenta parte desse ganho expõe os caminhos otimizados a comportamento inesperado com entradas malformadas, risco a avaliar antes de uma versão final.

REFERÊNCIAS

BEAMER, S.; ASANOVIĆ, K.; PATTERSON, D. The GAP benchmark suite. arXiv, 2015. Disponível em: <https://arxiv.org/abs/1508.03619>. Acesso em: 11 setembro 2026.

SAHU, S.; KOTHAPALLI, K. GVEL: fast graph loading in edgelist and compressed sparse row (CSR) formats. In: EURO-PAR 2024: PARALLEL PROCESSING WORKSHOPS. Cham: Springer, 2024. Disponível em: https://doi.org/10.1007/978-3-031-90200-0_22.

Da Pergunta à Evidência: busca científica híbrida

From Question to Evidence: hybrid scientific search

Rayssa Gomides Marconato – rayssagomidesmarconato@gmail.com
Victor Hugo Barros Baptista Pereira – victorhugobbpereira02@gmail.com

INTRODUÇÃO

A busca por referências é uma das etapas fundamentais de uma pesquisa científica, mas encontrar trabalhos realmente relevantes pode se tornar uma tarefa complexa diante do crescente volume de publicações disponíveis. Parte dessa dificuldade está no fato de que uma mesma ideia pode ser descrita de diferentes formas: o pesquisador formula sua questão utilizando determinados termos, enquanto um artigo potencialmente relevante pode empregar outro vocabulário para abordar o mesmo conceito. Métodos tradicionais de Recuperação de Informação, como o BM25, são eficientes na correspondência entre os termos presentes na busca e nos documentos (Robertson; Zaragoza, 2009), enquanto modelos de recuperação semântica buscam identificar relações de significado mesmo quando não há correspondência direta entre as palavras (Reimers; Gurevych, 2019). Entretanto, estudos demonstram que abordagens semânticas não necessariamente superam métodos lexicais em todos os domínios (Thakur et al., 2021), o que evidencia características potencialmente complementares entre essas estratégias. A partir desse cenário, este trabalho busca investigar se a combinação entre recuperação lexical e semântica pode melhorar a identificação e o ranqueamento de artigos relevantes para uma questão de pesquisa, comparando experimentalmente métodos lexicais, semânticos e híbridos em coleções científicas de referência.

DESENVOLVIMENTO

A avaliação experimental utilizou três coleções científicas do benchmark BEIR (Thakur et al., 2021): NFCorpus, SciFact e TREC-COVID, totalizando mais de 180 mil documentos. Foram comparadas quatro estratégias de recuperação: BM25, baseada em correspondência lexical (Robertson; Zaragoza, 2009); recuperação semântica por representações vetoriais; uma abordagem híbrida, combinando sinais lexicais e semânticos; e o reranqueamento neural dos resultados recuperados (Nogueira; Cho, 2019). O desempenho foi avaliado principalmente pelo $nDCG@10$, complementado por métricas de precisão e cobertura. Os resultados mostraram que a abordagem híbrida obteve o melhor $nDCG@10$ no NFCorpus (0,3351) e no SciFact (0,6698), enquanto no TREC-COVID o híbrido associado ao reranqueamento alcançou 0,7083, frente a 0,5678 do BM25, representando ganho de 24,74%.

CONCLUSÃO

Os resultados demonstram que a combinação entre recuperação lexical e semântica pode melhorar a identificação e o ranqueamento de literatura científica, confirmando o potencial de complementaridade entre essas abordagens. O método híbrido apresentou o melhor desempenho em duas das três coleções avaliadas, enquanto sua associação ao reranqueamento obteve o maior ganho no TREC-COVID. Como continuidade, pretende-se investigar modelos ajustados a domínios científicos

específicos e aplicar a abordagem em uma ferramenta de apoio à busca por referências acadêmicas.

REFERÊNCIAS

NOGUEIRA, R.; CHO, K. Passage re-ranking with BERT. arXiv, 2019. Disponível em: <https://arxiv.org/abs/1901.04085>. Acesso em: 13 setembro 2026.

REIMERS, N.; GUREVYCH, I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING AND THE INTERNATIONAL JOINT CONFERENCE ON NATURAL LANGUAGE PROCESSING (EMNLP-IJCNLP), 2019, Hong Kong. Proceedings [...]. Hong Kong: Association for Computational Linguistics, 2019. p. 3982-3992. Disponível em: <https://doi.org/10.18653/v1/D19-1410>.

ROBERTSON, S.; ZARAGOZA, H. The probabilistic relevance framework: BM25 and beyond. Foundations and Trends in Information Retrieval, v. 3, n. 4, p. 333-389, 2009. Disponível em: <https://doi.org/10.1561/15000000019>.

THAKUR, N. et al. BEIR: a heterogeneous benchmark for zero-shot evaluation of information retrieval models. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS DATASETS AND BENCHMARKS TRACK, 35., 2021, Online. Proceedings [...]. [S.l.: s.n.], 2021. Disponível em: <https://openreview.net/forum?id=wCu6T5RI5Z>. Acesso em: 13 setembro 2026.

Desempenho da Mesclagem de Bases Bibliográficas via NBVIZ

Performance of Bibliographic Database Merging via NBVIZ

Leonardo Neves Bolfarini – leonardobolfarini@gmail.com

Rafael Gutierrez Castanha – rafaelcastanha@unimar.br

INTRODUÇÃO

A análise cienciométrica depende de metadados bibliográficos organizados e comparáveis. Entretanto, a cobertura e a estrutura dos registros variam entre fontes como Scopus, Web of Science e OpenAlex, o que torna necessária a padronização antes da produção de indicadores e redes científicas (Visser; Van Eck; Waltman, 2021; Priem; Piwowar; Orr, 2022). Este trabalho avaliou o desempenho do fluxo já implementado no NBVIZ, ecossistema de análises de dados científicos, para integrar registros dessas três bases. O objetivo foi verificar o comportamento da mesclagem e da deduplicação em diferentes volumes de dados, identificando a permanência de registros e o tempo necessário para gerar os arquivos de saída.

DESENVOLVIMENTO

Foram definidos três cenários de teste a partir da expressão de busca “graph”: 1.000, 10.000 e 20.000 registros de cada base, que somados correspondem a 3.000, 30.000 e 60.000 registros de entrada. Cada cenário foi executado cinco vezes online, no servidor de produção do NBVIZ, pela própria interface web. Assim, cada execução percorreu o fluxo real da aplicação: o front-end enviou os dados à API, o backend realizou leitura e padronização e a biblioteca nbviz_scientometric_tools executou a mesclagem e a deduplicação. Foram medidos o tempo da requisição no endpoint e, separadamente, o processamento isolado da mesclagem e da gravação dos arquivos de saída.

A deduplicação foi estável nas cinco repetições. Nos cenários de 3.000, 30.000 e 60.000 registros, foram preservados, respectivamente, 2.907, 27.645 e 55.720 registros; as remoções corresponderam a 3,10%, 7,85% e 7,13%. Para reportar a variação, foram calculadas média e mediana. O processamento isolado no backend, composto por mesclagem e gravação, apresentou médias e medianas de 0,045/0,043 s, 0,216/0,219 s e 0,363/0,370 s. Já o tempo da requisição apresentou médias/medianas de 11,877/11,604 s, 95,003/94,800 s e 231,105/189,726 s.

A média e a mediana são apresentadas conjuntamente porque respondem a papéis diferentes: a média evidencia a variação de todas as execuções, enquanto a mediana representa melhor a execução típica quando há valores extremos. No cenário de 60.000 registros, uma execução de 401,973 s elevou a média do tempo de requisição, mas não alterou de modo relevante o processamento isolado do backend. A diferença entre cerca de 0,37 s de processamento e 189,726 s na requisição mediana indica que o gargalo está no restante do fluxo da requisição, e não na mesclagem. Esse tempo inclui leitura e padronização, além da consulta paginada à OpenAlex; como as fases ainda não foram temporizadas individualmente, não é correto atribuir toda a latência a uma única fonte.

A continuidade proposta não é reescrever a lógica de mesclagem, que já se mostrou eficiente. Pretende-se instrumentar separadamente consulta, leitura, padronização, deduplicação e gravação. Só então serão avaliadas medidas como reaproveitamento temporário de resultados de buscas repetidas e processamento

assíncrono de consultas extensas, preservando atualização e rastreabilidade dos dados.

CONCLUSÃO

O fluxo completo do NBVIZ, executado online pela interface e pelo backend com apoio da biblioteca de processamento, manteve resultados de deduplicação consistentes nos três volumes. A mesclagem e a gravação permaneceram abaixo de 0,4 s mesmo com 60.000 registros, enquanto a requisição completa apresentou maior tempo e variação. Os resultados indicam que a prioridade é diagnosticar as etapas externas à mesclagem e não substituir o algoritmo atual.

REFERÊNCIAS

PRIEM, J.; PIWOWAR, H.; ORR, R. OpenAlex: a fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv, 2022. Disponível em: <https://arxiv.org/abs/2205.01833>. Acesso em: 10 setembro 2026.

VISSER, M.; VAN ECK, N. J.; WALTMAN, L. Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. Quantitative Science Studies, v. 2, n. 1, p. 20-41, 2021. Disponível em: https://doi.org/10.1162/qss_a_00112.

Desenvolvimento de recursos pedagógicos digitais para o ensino de lógica matemática em cursos superiores de tecnologia

Development of digital pedagogical resources for teaching mathematical logic in higher education technology programs

Lucia Maria Reis Braga – luciamrbraga@gmail.com
Bianca Savegnago de Mira – biancamira@unimar.br

INTRODUÇÃO

A lógica matemática constitui fundamento essencial à formação de profissionais de TI, sustentando a programação, os algoritmos e o raciocínio computacional (Setti, 2009; Raeder et al., 2016). A transição do raciocínio intuitivo para o formal impõe obstáculos epistemológicos recorrentes (Setti; Cifuentes, 2006), agravados em campos como a Inteligência Artificial. Pensar matematicamente exige articular conhecimento explícito e tácito (Ernest, 1998; Schoenfeld, 1998), o que orienta o uso de recursos pedagógicos digitais, interativos e contextualizados às aplicações de TI.

Este projeto integra um projeto guarda-chuva voltado ao desenvolvimento colaborativo de materiais didáticos entre docente e discentes dos cursos de tecnologia da UNIMAR, reunidos em livro digital disponibilizado em plataforma web de acesso público. O objetivo geral é desenvolver algoritmos educacionais fundamentados em lógica matemática como apoio ao ensino. Como objetivos específicos: sistematizar, via pesquisa bibliográfica, a relação entre lógica matemática e aplicações de TI; implementar algoritmos educacionais sobre lógica proposicional e de predicados; e contribuir para a interface visual da plataforma. A originalidade do estudo reside na escassez de materiais didáticos em português que articulem, de forma interativa, os fundamentos da lógica matemática às aplicações de TI. A metodologia compreende pesquisa bibliográfica e desenvolvimento técnico-pedagógico de algoritmos e da interface da plataforma.

DESENVOLVIMENTO

O desenvolvimento organiza-se em duas frentes complementares. A primeira consiste no levantamento e análise de literatura sobre lógica matemática, raciocínio computacional e ensino de programação, articulando três eixos: a fundamentação filosófico-matemática da lógica formal; contribuições da Educação Matemática sobre conhecimento tácito e explícito e sobre o “pensar matematicamente” (Ernest, 1998; Schoenfeld, 1998); e estudos da Ciência da Computação sobre a transição do raciocínio intuitivo para o formal (Setti, 2009; Setti; Cifuentes, 2006; Raeder et al., 2016). Adota-se a revisão bibliográfica narrativa, cujo produto será sistematizado em capítulo textual do livro digital.

A segunda frente compreende a concepção e implementação de três algoritmos educacionais interativos — um gerador de tabela-verdade, um validador de proposições lógicas (simples e compostas) e um validador de estruturas lógicas com quantificadores — além da construção do front-end da plataforma web.

O projeto se encontra em fase de fundamentação teórica e planejamento técnico, sem resultados de implementação a relatar. As próximas etapas são: finalização do levantamento bibliográfico, definição da arquitetura dos algoritmos e prototipação de interfaces, implementação, e testes de validação funcional com casos representativos de lógica proposicional e de predicados. Resultados parciais serão reportados em relatório subsequente.

CONCLUSÃO

Espera-se, ao final do projeto, a produção de um capítulo de contextualização da lógica matemática para cursos de TI, o desenvolvimento da interface visual da plataforma — layout, homepage, menu principal e organização das páginas — e a implementação dos três algoritmos educacionais descritos, integrados à plataforma e disponíveis aos estudantes.

REFERÊNCIAS

ERNEST, P. Mathematical knowledge and context. In: WATSON, A. (Ed.). Situated cognition and the learning of mathematics. Oxford: University of Oxford Department of Educational Studies, 1998. cap. 1, p. 13-29.

RAEDER, M. et al. L2pm: relato de uma experiência sobre o ensino integrado de lógica, programação e matemática para computação. In: Workshop sobre Educação em Computação (WEI). SBC, 2016. p. 2186-2195.

SETTI, M. de O. G. O processo de discretização do raciocínio matemático na tradução para o raciocínio computacional. 2009. Tese (Doutorado em Educação) – UFPR, Curitiba, 2009.

SETTI, M. G.; CIFUENTES, J. C. Os obstáculos epistemológicos no pensamento computacional. In: SIMPÓSIO DE EDUCAÇÃO MATEMÁTICA, 8., 2006, Buenos Aires. Anais [...]. Buenos Aires: [s.n.], 2006.

Desenvolvimento de um Modelo Preditivo para Identificação de Fatores de Risco a partir de Dados Viários e de Acidentes de Trânsito

Development of a Predictive Model for Risk Factor Identification from Road and Traffic Accident Data

Diego Bernardo Artero – diego.bernardo.artero@gmail.com
Rafael Gutierrez Castanha – rafaelcastanha@unimar.br

INTRODUÇÃO

Os acidentes de trânsito figuram entre as principais causas de mortes evitáveis no Brasil, evidenciando a necessidade de antecipar fatores de risco. O problema da pesquisa consiste em identificar, a partir de dados públicos de sinistros viários do estado de São Paulo, os fatores mais associados à gravidade fatal de um acidente. Objetivou-se construir um pipeline de coleta e tratamento dos dados do Detran-SP, treinar um modelo em LightGBM e desenvolver uma rede neural em TensorFlow/Keras, permitindo comparar as duas abordagens. A originalidade do estudo reside na combinação entre geohash, features cíclicas de tempo e embeddings de entidade.

DESENVOLVIMENTO

Para a coleta dos dados do Infosiga – portal do Detran-SP – foi feito um web scraping, utilizando a biblioteca Selenium para Python. Enquanto o `etl_scraper.py` tem por função executar o web scraping na página do Infosiga obtendo o arquivo `dados_infosiga.zip`, o `etl_parser.py` tem a função de extrair arquivos e sintetizar os dados em 3 arquivos .csv: `Pessoas.csv`, `Sinistros.csv` e `Veículos.csv`.

A Análise Exploratória de Dados foi conduzida com as bibliotecas Pandas e Seaborn; foram descobertas inconsistências críticas da base de dados, com grande destaque à unicidade da gravidade dos acidentes anteriores a 2019 que eram classificados como Não Fatal. Vista a abundância de exemplares, todos os registros de 01/01/2015 até 31/12/2018 foram dispensados para a modelagem e as duplicidades foram tratadas via remoção dos exemplares idênticos, resultando em um dataset que abrange de 01/01/2019 até 31/08/2025.

Durante a Engenharia de Features, foram utilizados conceitos de seno e cosseno para categorizar dados cíclicos tais como dia, dia da semana, mês e hora. Além desse processo, foram aglutinadas ambas latitude e longitude em um geohash, por definição, um método de indexação espacial que transforma um ponto contínuo no espaço geográfico R^2 (latitude, longitude) em uma sequência discreta de símbolos, tal que prefixos comuns do código correspondem a regiões espaciais contíguas, permitindo consultas de proximidade e agregação espacial eficientes tratáveis como categoria, portanto, iteráveis pelo LightGBM.

Foi treinado um classificador LightGBM (Scikit-learn). Dado o desequilíbrio de classes ($\approx 25:1$), aplicou-se ajuste de threshold por maximização do F1, obtendo-se limiar ótimo de 0,1536 para a classe “Não Fatal”. Complementarmente, foi desenvolvida uma rede neural com tratamento híbrido das categóricas: embeddings de entidade (Guo; Berkahn, 2016) para as de alta cardinalidade (município, geohash, circunscrição, região administrativa) e one-hot para as de baixa cardinalidade (tipo_via, administração, conservação, turno, tipo_local); as numéricas são normalizadas internamente ao grafo; as representações são concatenadas em camadas densas com Batch Normalization e Dropout, ponderação de classe equivalente ao `scale_pos_weight` do LightGBM e o

mesmo critério de threshold por F1.

CONCLUSÃO

O projeto resultou em um pipeline funcional de coleta, tratamento, modelagem e visualização de dados de acidentes. A modelagem com LightGBM e uma rede neural em TensorFlow/Keras, tratando as categóricas via embeddings e one-hot conforme a cardinalidade, amplia o escopo comparativo do estudo e permite avaliar se a representação aprendida pelos embeddings supera a codificação nativa do LightGBM. Futuramente, recomenda-se comparar formalmente os dois modelos, ajustar hiperparâmetros e realizar novas validações cruzadas.

REFERÊNCIAS

GUO, C.; BERKHAHN, F. Entity embeddings of categorical variables. arXiv, 2016. Disponível em: <https://arxiv.org/abs/1604.06737>. Acesso em: 12 setembro 2026.

JINDAL, R. K.; AGARWAL, A. K.; SAHOO, A. K. Data analytics for analysing traffic accidents. Test Engineering and Management, v. 83, p. 14796-14803, mar./abr. 2020.

MORETTIN, P. A.; BUSSAB, W. de O. Estatística básica. São Paulo: Saraiva, 2017.

SANTOS, D. et al. Machine learning approaches to traffic accident analysis and hotspot prediction. Computers, v. 10, n. 12, p. 157, 2021.

Do RISC ao RISC-V: uma abordagem prática para o estudo de arquiteturas RISC e programação assembler

From RISC to RISC-V: a practical approach to studying RISC architectures and assembler programming

Murilo Silva Rêgo – murilosrego@gmail.com
Cesar Giacomini Penteado – cesargiacomini@gmail.com

INTRODUÇÃO

Arquiteturas RISC (Reduced Instruction Set Computer) influenciam diretamente desempenho e consumo energético. Este trabalho investiga, de forma prática, a programação em assembler em microcontroladores RISC legados (PIC16FXXX) e a transposição desses conhecimentos para a arquitetura mais moderna e modular RISC-V, aplicando-os no desenvolvimento de um jogo. O objetivo é entender como projetar programas eficientes em tempo de execução e consumo de energia, justificando a escolha pela arquitetura RISC.

DESENVOLVIMENTO

O estudo começou utilizando a documentação técnica dos microcontroladores (Microchip Technology Inc., [s.d.]) e desenvolvimento prático de programas simples que exploram recursos básicos (flags, rotinas, jumps, registradores, endereçamento e manipulação de bits) em miSim DE (Feersum Technology, [s.d.a], [s.d.b]). Após a familiarização com a base, um programa mais complexo de um jogo Tetris (Gunee, [s.d.]) foi estudado e modificado para colocar em prática o domínio obtido. Por fim, com a experiência obtida, documentação (RISC-V International, [s.d.]) e alguns cursos introdutórios ao RISC-V na Linux Foundation ([s.d.]), um jogo de campo minado foi desenvolvido para execução no Venus Simulator (HM.RISCV, [s.d.]), um ambiente que simula um RISC-V de 32 bits e módulos adicionais.

Durante o estudo dos microcontroladores mais antigos, o trabalho se desenvolveu inicialmente com algoritmos básicos para o PIC16F84, como operações aritméticas com controle de fluxo através de flags, que consolidaram as bases necessárias que serão usadas futuramente. A seguir ocorreu o estudo e modificação do jogo de Tetris de Rickard Gunee, em que foram identificadas e alteradas estruturas de armazenamento de peças, buffers e rotinas de renderização, rotinas de colisão e física geral entre as peças, rotinas de rotação e geração de peças para suportarem maior tamanho, tudo sob restrição de respeitar o limite de operações por ciclo de renderização de linhas de vídeo em saída NTSC.

No RISC-V, os principais desafios foram implementar rotinas em assembler para conversão de strings em HEX, iterações aninhadas para manipulação de malha de LEDs, geração e contagem de bombas, controle de I/O via terminal e validação de coordenadas, resultando em um jogo de campo minado funcional.

CONCLUSÃO

O percurso prático de RISC clássico ao RISC-V mostrou que o estudo de assembler em microcontroladores legados fornece compreensão sólida de conceitos essenciais para arquiteturas mais modernas. Resultados principais incluem alguns algoritmos autorais para PIC16F84 simulado, PIC16F628A real e para um RISC-V

simulado, um Tetris fortemente modificado e executado em hardware real no PIC16F628A e um Campo Minado em RISC-V que serve como caso de estudo para assembler moderno. Como continuidade, o próximo objetivo é simular um RISC-V real em um FPGA e executá-lo em uma matriz de LEDs física conectada à placa.

REFERÊNCIAS

FEERSUM TECHNOLOGY. Feersum Technology: página institucional. [S.l.], [s.d.a]. Disponível em: <https://www.feertech.com/>. Acesso em: 31 agosto 2026.

FEERSUM TECHNOLOGY. miSim DE: software de simulação de microcontrolador. [S.l.], [s.d.b]. Disponível em: <https://web.archive.org/web/20071208064027/http://www.feertech.com/misim/software.html>. Acesso em: 31 agosto 2026.

GUNEE, R. Tetris (PIC16F84): código do projeto. [S.l.], [s.d.]. Disponível em: <https://web.archive.org/web/20080208093946/http://www.rickard.gunee.com/projects/video/pic/tetris.php>. Acesso em: 31 agosto 2026.

HM.RISCV. Venus Simulator: plugin para VSCode. [S.l.], [s.d.]. Disponível em: <https://marketplace.visualstudio.com/items?itemName=hm.riscv-venus>. Acesso em: 31 agosto 2026.

LINUX FOUNDATION. Introduction to RISC-V (LFD110). [S.l.], [s.d.]. Disponível em: <https://training.linuxfoundation.org/training/introduction-to-riscv-lfd110/>. Acesso em: 31 agosto 2026.

MICROCHIP TECHNOLOGY INC. PIC16F84 datasheet. [S.l.], [s.d.]. Disponível em: <https://ww1.microchip.com/downloads/aemDocuments/documents/MCU08/ProductDocuments/DataSheets/30430D.pdf>. Acesso em: 31 agosto 2026.

RISC-V INTERNATIONAL. RISC-V: página oficial. [S.l.], [s.d.]. Disponível em: <https://riscv.org/>. Acesso em: 31 agosto 2026.

Engenharia social e phishing: proposta de um framework para análise dos impactos da exploração de vulnerabilidades humanas na segurança de dados pessoais sensíveis

Social engineering and phishing: a proposed framework for analyzing the impacts of exploiting human vulnerabilities on the security of sensitive personal data

Letícia Oliveira Corrêa – leticiaocor@gmail.com
Douglas Rodrigues – douglasrodrigues@unimar.br

INTRODUÇÃO

A crescente dependência das tecnologias digitais tem ampliado a exposição de indivíduos e organizações a ameaças cibernéticas. Entre essas ameaças, o phishing e as técnicas de engenharia social destacam-se por explorar vulnerabilidades humanas em vez de falhas exclusivamente tecnológicas. A manipulação psicológica de usuários representa um dos principais desafios para a segurança da informação, uma vez que pode comprometer sistemas mesmo quando mecanismos técnicos de proteção estão presentes. Este trabalho tem como objetivo investigar os impactos do phishing e da engenharia social sobre indivíduos e organizações, considerando aspectos técnicos, financeiros, sociais e psicológicos. Busca-se compreender os danos decorrentes dessas práticas e discutir estratégias de prevenção fundamentadas na educação digital e na conscientização em segurança da informação. A pesquisa adota abordagem qualitativa e exploratória, baseada em revisão bibliográfica e análise crítica da literatura especializada sobre phishing, engenharia social e segurança da informação, fundamentando-se em relatórios técnicos do setor, como os da APWG (2024) e de Kosinski (2024), e em análises especializadas sobre os mecanismos psicológicos da manipulação, como a de Andersen (2025). Os detalhes metodológicos e a construção do framework proposto serão apresentados na seção de Desenvolvimento. Espera-se evidenciar que os impactos do phishing e da engenharia social extrapolam o campo técnico, alcançando as dimensões financeira, social e psicológica, o que reforça a relevância de estratégias de prevenção fundamentadas na educação digital e na conscientização em segurança da informação.

DESENVOLVIMENTO

Com base na revisão bibliográfica e na análise crítica dos relatórios técnicos citados, propõe-se um framework original, composto por cinco categorias de impacto (técnica, financeira, social, psicológica e organizacional), que relaciona o tipo de dado exposto ao perfil da vítima e aos vetores comportamentais envolvidos, como os diferentes tipos de ataque descritos por Kosinski (2024). Os critérios de criticidade serão definidos com base na combinação de três dimensões: (a) extensão do dano, medida pelo volume de dados comprometidos ou pelo número de vítimas afetadas; (b) reversibilidade, considerando a possibilidade de recuperação dos ativos ou da reputação; e (c) impacto residual, avaliando consequências persistentes como sequelas psicológicas ou exposição prolongada de dados. A escala resultante (baixo, moderado, alto e crítico) permitirá cruzar indicadores quantitativos dos relatórios setoriais com o referencial teórico de engenharia social.

CONCLUSÃO

Ao final da pesquisa, o framework proposto deverá oferecer uma ferramenta sistemática para classificar e priorizar riscos associados à exploração do fator humano, reforçando que os impactos do phishing e da engenharia social extrapolam o prejuízo financeiro imediato e subsidiando ações de conscientização em segurança da informação. Como próximos passos, prevê-se a validação do framework a partir do cruzamento dos dados dos relatórios técnicos com o referencial teórico adotado, além da redação de um artigo científico para submissão a evento ou periódico da área.

REFERÊNCIAS

- ANDERSEN, G. The psychology of social engineering. [S.l.]: Coalition, 2025. Disponível em: <https://www.coalitioninc.com/blog/security-labs/the-psychology-of-social-engineering>. Acesso em: 9 setembro 2026.
- ANTI-PHISHING WORKING GROUP (APWG). Phishing activity trends report: 3rd quarter 2024. [S.l.]: APWG, 2024.
- KOSINSKI, M. O que é phishing? [S.l.]: IBM, 2024. Disponível em: <https://www.ibm.com/br-pt/think/topics/phishing>. Acesso em: 10 setembro 2026.

MeuPonteiro: ferramenta web educacional para visualização e geração de código no ensino de estruturas de dados em C

MyPointer: an educational web tool for visualizing and generating code in teaching data structures in C

Ronaldo Tavares Cano – ronaldotavarescano@gmail.com
Douglas Rodrigues – douglasrodrigues@unimar.br

INTRODUÇÃO

O aprendizado de algoritmos e estruturas de dados é um dos maiores obstáculos no início da graduação em Ciência da Computação, exigindo compreender conceitos abstratos, como alocação dinâmica e manipulação de ponteiros, a partir exclusivamente de códigos textuais (La Rocca, 2021). A ausência de representações visuais que sejam diretamente alinhadas à sintaxe da linguagem C agrava essa dificuldade de compreensão, evidenciando a necessidade de ferramentas específicas para complementar opções generalistas já existentes (VisuAlgo, 2026). Diante disso, este projeto tem como objetivo propor o desenvolvimento da ferramenta web meuPonteiro e avaliar o seu impacto no processo de ensino-aprendizagem de estruturas de dados em C, aproveitando a eficácia já comprovada do uso de representações visuais no estudo de algoritmos (Hundhausen; Douglas; Stanko, 2002). A ferramenta visa preencher uma lacuna metodológica ao permitir o controle interativo de simulações, onde o estudante pode parametrizar componentes e observar o rearranjo imediato dos ponteiros na tela. Para alcançar esses objetivos, o projeto adota uma abordagem metodológica baseada no desenvolvimento iterativo de software associada à validação empírica por meio de testes de usabilidade com turmas de graduação.

DESENVOLVIMENTO

A metodologia baseia-se na construção iterativa da aplicação utilizando o ecossistema React e TypeScript. A interface interativa e sincronização reativa foram desenvolvidas com a biblioteca React Flow para a renderização de grafos. Para garantir a reprodutibilidade e o isolamento de dependências, o ambiente foi consolidado com Vite e o gerenciador de pacotes Nix, com o código versionado em um repositório no GitHub (<https://github.com/ronaldotavcano/IniciacaoCientifica>).

Como resultados parciais, o projeto conta com um protótipo visual funcional que renderiza nós e arestas em tempo real, simulando com êxito a representação de estruturas como pilhas. Testes preliminares em ambiente de desenvolvimento confirmam que a sincronização reflete imediatamente a alocação de memória e o redirecionamento de ponteiros. A próxima etapa consistirá na validação empírica via testes de usabilidade com turmas experimentais de graduação, visando monitorar a evolução e o engajamento dos alunos na disciplina.

CONCLUSÃO

O desenvolvimento inicial do projeto meuPonteiro resultou na entrega de um protótipo visual funcional capaz de renderizar e sincronizar, em tempo real, a alocação de memória e a manipulação de ponteiros. Essa entrega preliminar atende ao objetivo proposto de reduzir a distância entre os conceitos abstratos de estruturas de dados e a sua codificação na linguagem C, oferecendo uma representação interativa que preenche

a lacuna metodológica das abordagens puramente textuais. Como próximos passos para a continuidade da pesquisa, o sistema será submetido a testes empíricos de usabilidade com turmas de graduação. Essa etapa de validação permitirá mensurar o nível de engajamento dos alunos, consolidando a ferramenta como um recurso didático de alto impacto com potencial para melhorar a compreensão prática de tópicos complexos, como listas encadeadas, e impactar positivamente as taxas de aprovação na disciplina.

REFERÊNCIAS

HUNDHAUSEN, C. D.; DOUGLAS, S. A.; STANKO, J. T. A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing*, v. 13, 2002.

LA ROCCA, M. Entendendo estruturas de dados. 1. ed. São Paulo: Novatec, 2021.

VISUALGO. Visualising data structures and algorithms through animation. [S.l.], 2026. Disponível em: <https://visualgo.net/>. Acesso em: 13 setembro 2026.

Modelos Preditivos Espaço-Temporais para Antecipação de Ocorrências de Violência Doméstica em Marília-SP

Spatio-Temporal Predictive Models for Anticipating Domestic Violence Occurrences in Marília-SP

Larah Cezário Lima da Silva – cezariolarah@gmail.com
Rafaela Tavares Návía – rafaelatnunimar@gmail.com
Matheus Ferraroni Sanches – matheusferraroni@unimar.br

INTRODUÇÃO

A crescente disponibilidade de dados públicos de segurança no Brasil possibilita modelos capazes de antecipar situações de risco, diante do agravamento da violência doméstica contra a mulher. Em 2025, o país registrou recorde de 1.470 feminicídios, com São Paulo liderando em números absolutos (MJSP, 2026; CNN Brasil, 2026; SSP, 2026). A efetividade das Medidas Protetivas de Urgência (MPUs) depende da identificação de padrões de vulnerabilidade e da capacidade de agir preventivamente (Ismael; Camargo, 2018), porém a utilização desses dados para fins preditivos ainda é limitada.

Diante disso, este projeto propõe modelos preditivos espaço-temporais para Marília-SP, utilizando ocorrências registradas entre 2003 e 2022 (Freitas; Clarindo; Aguiar, 2023). A abordagem combina clusterização espacial por densidade (DBSCAN), para identificar regiões de maior concentração criminal, com modelos de séries temporais (ARIMA e Holt-Winters), capazes de representar dependências temporais e sazonalidade. O objetivo é construir indicadores que apoiem a proteção à mulher e a tomada de decisão no âmbito das MPUs, por meio da integração das bases públicas, identificação de regiões críticas, previsão do tempo-resposta, desenvolvimento de uma ferramenta de visualização e predição espaço-temporal e avaliação de sua aplicabilidade em outros municípios.

DESENVOLVIMENTO

A pesquisa parte de revisão bibliográfica sobre violência doméstica, MPUs, policiamento inteligente e bases públicas, com destaque para o dataset SPSafe (USP). A primeira etapa consistiu na limpeza e padronização da base de Marília, realizada em Python, com filtragem do período de 2003 a 2022. Como o dataset apresenta múltiplos registros para um mesmo evento, decorrentes das combinações de natureza e veículo em um mesmo boletim de ocorrência (B.O.), foi realizada a deduplicação por número de B.O., consolidando-os em uma única ocorrência. Na base tratada, aplicou-se o k-distance plot (k=20) para orientar a definição do parâmetro eps do DBSCAN, cujas configurações serão comparadas por ruído e coeficiente de silhueta. Como resultado preliminar, a deduplicação reduziu a base de 8.618 para 6.789 ocorrências únicas (-21,1%), evitando a superestimação da densidade em endereços com múltiplos registros. O k-distance plot indicou uma mudança acentuada da curva entre 300 e 350 m, faixa inicial para os testes de eps, mantendo min_samples = 20.

Paralelamente, os modelos ARIMA (Newbold, 1983) e Holt-Winters (Chatfield; Yar, 1988) serão utilizados para analisar dependência temporal e sazonalidade, com foco no tempo-resposta das MPUs. As próximas etapas incluem os testes do DBSCAN e a integração dos clusters a uma ferramenta de visualização geoespacial (Leaflet/HTML).

CONCLUSÃO

Os resultados preliminares confirmam a viabilidade da deduplicação e da calibração inicial do DBSCAN sobre a base de Marília, etapas que sustentam a construção dos clusters e a modelagem temporal via ARIMA e Holt-Winters. Espera-se que a ferramenta resultante aprimore a decisão operacional, fortaleça a proteção às mulheres em Marília e sirva de base para outros municípios.

REFERÊNCIAS

CHATFIELD, C.; YAR, M. Holt-Winters forecasting: some practical issues. *The Statistician*, v. 37, n. 2, p. 129, 1988.

CNN BRASIL. 2025 registra recorde de feminicídios com 4 mulheres mortas por dia no país. São Paulo, 20 jan. 2026. Disponível em: <https://www.cnnbrasil.com.br>. Acesso em: 21 junho 2026.

FREITAS, J. B.; CLARINDO, J. P.; AGUIAR, C. D. de. SPSafe: um dataset sobre dados de criminalidade no estado de São Paulo. In: DATA SCIENCE FOR SOCIAL GOOD WORKSHOP, 2023, Porto Alegre. Anais [...]. Porto Alegre: SBC, 2023. p. 48-57. Disponível em: <https://doi.org/10.5753/dsw.2023.233945>. Acesso em: 21 junho 2026.

ISMAEL, F.; CAMARGO, M. E. Medidas protetivas de urgência: um olhar “oficial” da demanda, desafios e soluções na região do Sertão do Araripe – PE. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, p. 21-276, 2018.

MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA (MJSP). Sistema Nacional de Informações de Segurança Pública (Sinesp). Brasília: MJSP, 2026. Disponível em: <https://www.gov.br/mj>. Acesso em: 21 junho 2026.

NEWBOLD, P. ARIMA model building and the time series analysis approach to forecasting. *Journal of Forecasting*, v. 2, n. 1, p. 23-35, jan. 1983.

PORTAL SSP. Programa SP Mulher. [S.l.], 2026. Disponível em: <https://www.ssp.sp.gov.br/institucional/programa-sp-mulher>. Acesso em: 21 junho 2026.

SECRETARIA DA SEGURANÇA PÚBLICA DE SÃO PAULO (SSP). Números sem mistério: vítimas de violência doméstica. [S.l.]: Portal SSP, 2026. Disponível em: <https://www.ssp.sp.gov.br/estatistica/violencia-domestica>. Acesso em: 21 junho 2026.

RAG sem Vetores para Domínios Técnicos: desenvolvimento, validação e análise de documentos agronômicos baseado em indexação hierárquica e raciocínio sobre árvore documental

Vectorless RAG for Technical Domains: development, validation and analysis of agronomic documents based on hierarchical indexing and reasoning over document tree

Eduardo Vinícius dos Santos Menezes – eduardo.santosm1202@gmail.com
Caio Saraiva Coneglian – caio.coneglian@gmail.com

INTRODUÇÃO

Os Modelos de Linguagem de Grande Escala (LLMs) geram afirmações plausíveis, porém factualmente incorretas, fenômeno designado alucinação (Ji et al., 2023). No domínio agronômico isso é risco operacional direto: erros em doses de corretivos ou em índices de saturação acarretam prejuízo econômico e dano à lavoura. A Geração Aumentada por Recuperação (RAG) mitiga o problema ao fundamentar a resposta em fontes externas recuperadas na inferência (Lewis et al., 2020); contudo, na formulação vetorial predominante, a fragmentação dos documentos em chunks rompe a hierarquia lógica do texto técnico e fragiliza a rastreabilidade das citações (Gao et al., 2023), requisito crítico onde o agrônomo responde tecnicamente pela prescrição. O RAG sem vetores substitui os embeddings por indexação estruturada e por recuperação conduzida pelo raciocínio do LLM sobre a estrutura do documento; o framework PageIndex (Zhang; Tang, 2025) materializa esse paradigma, mas foi validado apenas sobre relatórios financeiros em inglês. Este trabalho desenvolve, implementa e valida o Agronomy Assistant, assistente para análise de prescrições agronômicas baseado em RAG sem vetores, e o compara ao RAG vetorial denso e híbrido em corpus agronômico em português, sob a hipótese de que a indexação hierárquica oferece rastreabilidade superior sem perda de corretude. A pesquisa é aplicada, experimental e quantitativa.

DESENVOLVIMENTO

O corpus reúne documentos técnico-agronômicos em português: boletins técnicos da EMBRAPA, manuais de prescrição e laudos de análise de solo de uma propriedade parceira, em PDF com hierarquia explícita de seções. Comparam-se três arquiteturas: RAG vetorial denso (BGE-M3 e busca por vizinhos aproximados em Qdrant), RAG vetorial híbrido (BGE-M3 e BM25 fundidos por Reciprocal Rank Fusion) e RAG sem vetores (PageIndex com GPT-4o via Azure OpenAI). Nesta última, sem etapa de embedding, o documento é representado como árvore de nós com título, resumo e intervalo de páginas, percorrida de cima para baixo pelo raciocínio do modelo. Foi implantado em produção o Agronomy Assistant, sistema full-stack em FastAPI e React 19 com painel de citações e dashboard de métricas, que mantém as três arquiteturas ativas. A avaliação emprega ground truth sintético de 120 pares pergunta-fonte, gerados por modelo distinto do usado nas consultas para mitigar a circularidade avaliativa, em quatro categorias: factual, numérico-quantitativa, multi-hop e comparativa. Nos ensaios preliminares, com 30 perguntas das duas primeiras categorias sobre cinco documentos e corretude binária, o PageIndex citou documento, página e seção em 100% das respostas, enquanto as vetoriais citaram apenas o chunk recuperado; a corretude foi similar entre as três no subconjunto factual e a latência e os tokens do PageIndex ficaram em cerca do dobro dos do RAG denso. Registra-se como ameaça à validade a assimetria de infraestrutura entre os modelos geradores.

CONCLUSÃO

Os ensaios preliminares respondem parcialmente de forma afirmativa à pergunta central: o RAG sem vetores oferece rastreabilidade consistente até o nível de documento, página e seção, superando qualitativamente o RAG vetorial no critério mais crítico para a auditabilidade agronômica, ao custo de maior latência e consumo de tokens. As contribuições são o sistema em produção, um ground truth sintético de 120 perguntas e um protocolo de avaliação em duas fases. Seguem-se a avaliação completa com o framework RAGAS (Es et al., 2024), a validação do ground truth por especialistas e a unificação do modelo gerador.

REFERÊNCIAS

ES, S. et al. RAGAS: automated evaluation of retrieval augmented generation. In: CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (EACL), 2024. Proceedings [...]. [S.l.: s.n.], 2024. p. 150-158.

GAO, Y. et al. Retrieval-augmented generation for large language models: a survey. arXiv, 2023. Disponível em: <https://arxiv.org/abs/2312.10997>. Acesso em: 14 abril 2026.

Jl, Z. et al. Survey of hallucination in natural language generation. ACM Computing Surveys, v. 55, n. 12, p. 1-38, 2023.

LEWIS, P. et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, v. 33, p. 9459-9474, 2020.

ZHANG, M.; TANG, Y. PageIndex: document index for vectorless, reasoning-based RAG. [S.l.]: VectifyAI, 2025. Disponível em: <https://github.com/VectifyAI/PageIndex>. Acesso em: 1 maio 2026.

Representação Computacional de Conhecimento Tecnológico Expresso em Claims Patentárias por Meio de Estruturas Semânticas

Computational Representation of Technological Knowledge Expressed in Patent Claims Through Semantic Structures

Paulo Henrique Almeida Veloso Leite – paulohenriqueavl291106@gmail.com
Rafael Gutierrez Castanha – rafaelcastanha@unimar.br

INTRODUÇÃO

Patentes constituem importantes fontes de conhecimento tecnológico, reunindo descrições de componentes, funções, mecanismos e relações associadas a diferentes artefatos e processos técnicos. Entre seus elementos, as claims patentárias destacam-se por concentrarem a definição formal dos aspectos tecnológicos protegidos, frequentemente apresentando estruturas relacionais complexas. Nesse contexto, a pesquisa investiga formas de representar computacionalmente esse conhecimento de maneira semanticamente estruturada.

A investigação fundamenta-se em Representação do Conhecimento, ontologias, Semantic Web e Knowledge Graphs. Ontologias permitem explicitar entidades, propriedades e relações de um domínio (Gruber, 1993), enquanto tecnologias como RDF e OWL favorecem a representação interoperável do conhecimento (Antoniou; Van Harmelen, 2004). Knowledge Graphs, por sua vez, organizam informações por meio de entidades e relações explicitamente modeladas (Hogan et al., 2021). O estudo utiliza como domínio claims associadas a tecnologias robóticas patentárias relacionadas à Petrobras.

DESENVOLVIMENTO

A pesquisa compreende revisão bibliográfica, construção do corpus, análise semântica, modelagem conceitual, desenvolvimento de ontologia preliminar e avaliação qualitativa. As claims serão selecionadas por meio da plataforma PATENTSCOPE (WIPO, [s.d.]), da World Intellectual Property Organization (WIPO), considerando relevância temática, riqueza semântica e densidade relacional.

A análise será orientada pelas dimensões “o que é”, “o que faz”, “como faz” e “como se relaciona”, destinadas à identificação de entidades, funções, mecanismos e relações entre elementos tecnológicos. A partir dessa decomposição, será investigada uma modelagem conceitual para explicitar o conhecimento identificado, considerando relações como hasComponent, performsFunction, operatesThrough, communicatesWith e dependsOnClaim.

Posteriormente, será investigada a construção de uma ontologia preliminar e de um protótipo exploratório com tecnologias compatíveis com a Semantic Web, potencialmente incluindo RDF, OWL, Protégé, GraphDB, Neo4j e RDFLib.

Como uma breve demonstração da utilização dos conceitos detalhados, foi utilizada a patente BR1020120240581, intitulada “Dispositivo para limpeza de interfaces de vedação” (Petrobras, [s.d.]), de titularidade da PETROBRAS. A análise de suas reivindicações permite identificar inicialmente entidades tecnológicas, como DispositivoLimpeza, BaseSuporte, ManipuladorROV e ElementoLimpeza, que podem ser representadas como indivíduos de classes conceituais como Dispositivo e Componente. A relação hasComponent, definida entre essas classes, permite representar, por exemplo, que DispositivoLimpeza possui BaseSuporte,

ManipuladorROV e ElementoLimpeza como componentes.

CONCLUSÃO

A pesquisa busca contribuir para a representação computacional do conhecimento tecnológico presente em claims patentárias por meio de estruturas semânticas capazes de explicitar seus principais elementos e relações. Os resultados preliminares sustentam a continuidade da modelagem, formalização e desenvolvimento do protótipo, com avaliação posterior da preservação semântica, coerência conceitual, representação relacional e interoperabilidade.

REFERÊNCIAS

ANTONIOU, G.; VAN HARMELEN, F. A semantic web primer. Cambridge: MIT Press, 2004.

GRUBER, T. R. A translation approach to portable ontology specifications. Knowledge Acquisition, v. 5, n. 2, p. 199-220, 1993.

HOGAN, A. et al. Knowledge graphs. ACM Computing Surveys, v. 54, n. 4, 2021.

PETROBRAS. Dispositivo para limpeza de interfaces de vedação. BR 10 2012 024058-0. [S.l.], [s.d.]. Disponível em: <https://patentscope.wipo.int/search/pt/detail.jsf?docId=BR128734624>. Acesso em: 11 setembro 2026.

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO). PATENTSCOPE: search international and national patent collections. [S.l.], [s.d.]. Disponível em: <https://patentscope.wipo.int/>. Acesso em: 28 maio 2026.

Roteamento Inteligente e Preventivo de Viaturas Policiais para Atendimento a Ocorrências de Violência Doméstica em Marília-SP

Intelligent and Preventive Routing of Police Vehicles for Responding to Domestic Violence Occurrences in Marília-SP

Rafaela Tavares Návila – rafaelatnunimar@gmail.com
Larah Cezário Lima da Silva – cezariolarah@gmail.com
Matheus Ferraroni Sanches – matheusferraroni@unimar.br

INTRODUÇÃO

A crescente demanda por respostas ágeis às ocorrências de violência doméstica evidencia a necessidade de soluções capazes de orientar, em tempo real, a atuação das forças de segurança pública. No contexto da violência contra a mulher, essa agilidade é essencial para a efetividade das Medidas Protetivas de Urgência (MPUs). O roteamento de viaturas de emergência é abordado pelo Emergency Vehicle Routing Problem (EVRP), que busca trajetos eficientes considerando tempo, distância e condições da malha viária (Qin et al., 2017).

Este projeto propõe uma solução de roteamento inteligente, fundamentada no EVRP e nos princípios dos Sistemas de Gerenciamento de Tráfego (TMS), adaptada à segurança pública de Marília-SP e integrada à previsão de ocorrências por região. O sistema prevê áreas de maior risco para o posicionamento preventivo das viaturas e, após a confirmação da ocorrência, calcula a rota mais eficiente por meio do algoritmo de Dijkstra. Busca-se, assim, reduzir o tempo de resposta e ampliar a cobertura geográfica do município. Especificamente, pretende-se mapear o espaço urbano e os recursos policiais; aplicar Ridge Regression para prever a concentração de ocorrências por região e período; orientar o reposicionamento preventivo das viaturas; desenvolver o roteamento para ocorrências confirmadas; e comparar o modelo atual com cenários otimizados.

DESENVOLVIMENTO

A pesquisa baseia-se em revisão bibliográfica sobre EVRP, policiamento preditivo e TMS, com apoio em trabalhos correlatos (Luiz, 2022). O ambiente urbano foi modelado com dados do OpenStreetMap processados no SUMO, enquanto o histórico criminal foi obtido do SP Safe (Freitas; Clarindo; Aguiar, 2023) e da Secretaria da Segurança Pública de São Paulo (São Paulo, [s.d.]), com foco em violência doméstica contra a mulher e georreferenciamento dos recursos policiais fixos.

A metodologia é dividida em duas etapas. Na preditiva, o histórico é analisado espaço-temporalmente e utilizado em um modelo de Ridge Regression (Dorugade, 2014) para prever a concentração de ocorrências por região e janela temporal, orientando o reposicionamento preventivo das viaturas. Na reativa, após a confirmação da ocorrência, o algoritmo de Dijkstra (Dijkstra, 1959) calcula a rota mais rápida pela malha viária. O modelo atual será reconstruído no SUMO e comparado aos cenários otimizados quanto ao tempo de resposta, cobertura territorial e eficiência na alocação de recursos. Como resultado preliminar, foi obtida a malha viária completa de Marília no SUMO, abrangendo aproximadamente 47 km × 15 km, com 12.460 junções, 37.596 arestas e 52.387 lanes. Como próximos passos, estão a construção do grafo de roteamento, o mapeamento dos recursos policiais, o treinamento do modelo de Ridge Regression e a implementação do cenário-base para avaliação comparativa.

CONCLUSÃO

Os resultados preliminares confirmam a viabilidade da modelagem do ambiente urbano de Marília no SUMO, base para a futura aplicação do algoritmo de Dijkstra e integração do modelo de Ridge Regression. Espera-se que a solução contribua para maximizar a cobertura de regiões críticas, reduzir o tempo-resposta, aprimorar a alocação de recursos e fortalecer a proteção às mulheres em situação de vulnerabilidade, reforçando a efetividade das Medidas Protetivas de Urgência.

REFERÊNCIAS

DIJKSTRA, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik*, v. 1, p. 269-271, 1959.

DORUGADE, A. V. New ridge parameters for ridge regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, v. 15, n. 1, p. 94-99, abr. 2014.

FREITAS, J. B.; CLARINDO, J. P.; AGUIAR, C. D. de. SPSafe: um dataset sobre dados de criminalidade no estado de São Paulo. In: *DATA SCIENCE FOR SOCIAL GOOD WORKSHOP*, 2023, Porto Alegre. *Anais [...]*. Porto Alegre: SBC, 2023. p. 48-57. Disponível em: <https://doi.org/10.5753/dsw.2023.233945>. Acesso em: 21 junho 2026.

LUIZ, M. Proposta de policiamento inteligente: integrando modelos de localização preditiva e de roteirização eficiente de patrulhas policiais. 2022. Dissertação (Mestrado) – UFRGS, Porto Alegre, 2022.

QIN, J. et al. The emergency vehicle routing problem with uncertain demand under sustainability environments. *Sustainability*, v. 9, n. 2, p. 288, 21 fev. 2017.

SÃO PAULO (Estado). Secretaria da Segurança Pública. Estatísticas da criminalidade. [S.l.], [s.d.]. Disponível em: <https://www.ssp.sp.gov.br/estatistica>. Acesso em: 21 junho 2026.

Teoria dos Números Aplicada à Tecnologia da Informação: desenvolvimento de algoritmos educacionais e infraestrutura back-end para uma plataforma digital de ensino de matemática

Number Theory Applied to Information Technology: development of educational algorithms and back-end infrastructure for a digital mathematics teaching platform

Guilherme da Silva Santos – guilherme.svlsantos@gmail.com
Bianca Savegnago de Mira – biancamira@unimar.br

INTRODUÇÃO

A Teoria dos Números ocupa papel central na Ciência da Computação, fundamentando aplicações como criptografia, verificação de números primos e aritmética modular (Burton, 2016). As Diretrizes Curriculares do Ministério da Educação reconhecem a Matemática como ferramenta indispensável à formalização de conceitos computacionais (Menezes, 2005), reforçando a necessidade de contextualizar esses conteúdos às aplicações práticas da área (Ernest, 1998; Schoenfeld, 1992 *apud* Setti; Cifuentes, 2006).

Este trabalho de iniciação científica integra um projeto guarda-chuva conduzido pela orientadora, voltado ao desenvolvimento de uma plataforma e de um livro digital para o ensino de Matemática na UNIMAR. O objetivo geral é produzir algoritmos educacionais fundamentados em Teoria dos Números, contribuindo também para o back-end da plataforma. Os objetivos específicos são: pesquisa bibliográfica com produção de um capítulo; algoritmos sobre divisibilidade, MDC, MMC, primalidade e aritmética modular; implementação do back-end; e boas práticas com Git e GitHub.

DESENVOLVIMENTO

A metodologia compreende duas frentes. A primeira é a pesquisa bibliográfica sobre aplicações da Teoria dos Números em TI, sistematizada em um capítulo do livro digital, abordando divisibilidade, primalidade e aritmética modular como ferramentas de modelagem computacional (Cabral, 2017). A segunda é o desenvolvimento de algoritmos: calculadora de divisibilidade; cálculo do MDC e MMC pelo método de Euclides; verificador de primos; e módulo de aritmética modular. Paralelamente, será desenvolvido o back-end da plataforma no servidor da universidade, em JavaScript, com código organizado em módulos independentes e controle de versão via Git e GitHub. O cronograma de doze meses contempla pesquisa bibliográfica, desenvolvimento dos algoritmos, escrita do capítulo e implementação do back-end; atualmente, o levantamento bibliográfico já consolidou a fundamentação teórica (Freires; Santos; Sales, 2023).

CONCLUSÃO

Espera-se, ao final, disponibilizar um capítulo sobre Teoria dos Números aplicada à Tecnologia da Informação, algoritmos educacionais funcionais e a infraestrutura back-end da plataforma, integrados e disponíveis aos estudantes da UNIMAR. O projeto soma, assim, contribuição técnica e pedagógica, evidenciando a conexão entre a Teoria dos Números e sua aplicação em Ciência da Computação, com perspectiva de continuidade na plataforma do projeto guarda-chuva.

REFERÊNCIAS

BURTON, D. M. Teoria elementar dos números. 7. ed. Rio de Janeiro: LTC, 2016. E-book. ISBN 9788521631026. Disponível em:

<https://app.minhabiblioteca.com.br/reader/books/9788521631026/>. Acesso em: 14 setembro 2026.

CABRAL, R. M. P. Matemática discreta. 1. ed. Fortaleza: EdUECE, 2017. Disponível em:

https://educapes.capes.gov.br/bitstream/capes/432209/2/Livro_Matematica%20Discreta.pdf. Acesso em: 14 setembro 2026.

ERNEST, P. Mathematical knowledge and context. In: WATSON, A. (Ed.). Situated cognition and the learning of mathematics. Oxford: University of Oxford Department of Educational Studies, 1998. cap. 1, p. 13-29.

FREIRES, K. C. P.; SANTOS, M. A.; SALES, F. O. O uso de Teoria dos Números na computação: uma visão metodológica. In: COSTA, A. J. S. et al. (Org.). Matemática e suas possibilidades: ensino, pesquisa, formação docente e práticas pedagógicas. Iguatu: Quipá Editora, 2023. cap. 1. Disponível em:

<https://drive.google.com/file/d/1cl06POVCusfobl3H26l-S42p5BoFFNF/view>. Acesso em: 14 setembro 2026.

MENEZES, P. B. Matemática discreta para computação e informática. 3. ed. Porto Alegre: Bookman, 2005.

SETTI, M. G.; CIFUENTES, J. C. Os obstáculos epistemológicos no pensamento computacional. In: SIMPÓSIO DE EDUCAÇÃO MATEMÁTICA, 8., 2006, Buenos Aires. Anais [...]. Buenos Aires: [s.n.], 2006.