

GT 11 – Informação e Saúde

INTEGRAÇÃO DA LILACS AO POLYGLOT SEARCH TRANSLATOR: ESTUDO METODOLÓGICO SOBRE AUTOMAÇÃO EM ESTRATÉGIAS DE BUSCA

INTEGRATION OF LILACS INTO THE POLYGLOT SEARCH TRANSLATOR: METHODOLOGICAL STUDY ON AUTOMATION IN SEARCH STRATEGIES

Camila Belo Tavares Ferreira – Instituto Nacional de Câncer (INCA)

Mayara Rodrigues Batista – Instituto Nacional de Câncer (INCA)

Justin Clark – Bond University (BU)

Irene Priya Jose – Bond University (BU)

Modalidade: Resumo Expandido

Resumo: Este trabalho apresenta a integração da base de dados LILACS à ferramenta de automação Polyglot Search Translator, desenvolvida no *Institute for Evidence-Based Healthcare* da Bond University, como parte do sistema *Translating Evidence Review Accelerator (TERA)*. O objetivo foi ampliar a cobertura da ferramenta e valorizar a literatura científica latino-americana no processo de busca de evidências. O procedimento incluiu o mapeamento da sintaxe de busca da LILACS, testes e alinhamento técnico com o envolvimento de três instituições. Os testes indicaram bom desempenho e impactaram positivamente o serviço de busca em uma instituição especializada em câncer baseado na autonomia dos usuários. A integração fortalece o uso da LILACS na prática baseada em evidências em saúde.

Palavras-chave: armazenamento e recuperação da informação; interoperabilidade da informação em saúde; LILACS.

Abstract: This work presents the integration of the LILACS database into the automated tool Polyglot Search Translator, developed at the *Institute for Evidence-Based Healthcare* at Bond University, as part of the *Translating Evidence Review Accelerator (TERA)*. The objective was to expand the tool's coverage and highlight Latin American scientific literature in the evidence search process. The procedure included mapping the LILACS search syntax, conducting tests, and carrying out technical alignment involving three institutions. The tests demonstrated good performance and positively impacted the new search service at a cancer-specialized institution. The integration strengthens the use of LILACS in evidence-based health practice.

Keywords: information storage and retrieval; health information interoperability; LILACS.

1 INTRODUÇÃO

O acesso eficaz à informação científica é fundamental para a prática da Saúde Baseada em Evidências (SBE), especialmente no contexto das metodologias propostas pela Cochrane e pelo Joanna Briggs Institute (JBI). A construção de estratégias de busca sensíveis, específicas

e reproduzíveis é uma etapa crítica no desenvolvimento de revisões sistemáticas e outras formas de síntese do conhecimento. No entanto, a adaptação manual dessas estratégias entre diferentes bases de dados é uma realidade no Brasil e trata-se de um processo que demanda tempo, conhecimento técnico e atenção aos detalhes — especialmente em bases que utilizam sintaxes distintas.

O Polyglot Search Translator, ferramenta de automação desenvolvida pelo Institute for Evidence-Based Healthcare da Bond University - parte integrante do sistema *Translating Evidence Review Accelerator (TERA)* (<https://tera-tools.com/>) - tem se mostrado um recurso útil nesse processo, permitindo a tradução automatizada de estratégias entre bases como PubMed, Embase, Cochrane Library, entre outras, com evidências de economia de tempo e redução de erros (Clark *et al.*, 2020). Contudo, até 2025, a base Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS) ainda não havia sido integrada à ferramenta.

O problema de pesquisa que orienta este estudo é a ausência da LILACS no Polyglot Search Translator, o que limita a interoperabilidade da ferramenta e dificulta a integração da literatura científica latino-americana em estratégias de busca para revisões sistemáticas.

Considerando a relevância dessa fonte de informação no contexto das políticas públicas de saúde e da produção científica regional — especialmente para um centro de referência em oncologia no Brasil —, propôs-se a inclusão da base LILACS no Polyglot como parte de uma parceria internacional. Essa ação alinha-se ao processo de modernização do serviço de estratégias de busca do sistema integrado de bibliotecas dessa unidade de saúde, promovendo maior visibilidade da literatura regional e fortalecendo a Biblioteca Virtual em Saúde (BVS) como fonte de evidência para a tomada de decisão. Para isso, foi elaborado um estudo metodológico que descreve o processo até a disponibilização da LILACS no Polyglot.

O objetivo geral deste trabalho é descrever o processo de integração da base de dados LILACS ao Polyglot Search Translator, discutindo seu impacto potencial na prática da busca de evidências em saúde.

2 REFERENCIAL TEÓRICO

Apesar da ampla oferta de informação em saúde, pesquisadores e usuários frequentemente têm dúvidas sobre quais fontes consultar, quais as diferenças entre elas e como realizar buscas de qualidade (Abdala; Figueiró, 2017). De acordo com a JBI (Aromataris *et al.*, 2024, s.p.), “as pesquisas precisam ser traduzidas para mais de uma base de dados e

raramente podem ser executadas exatamente ‘como estão’”. Isso porque cada base de dados possui seu próprio vocabulário controlado, muitas vezes proprietário, e a tradução permite que você aproveite ao máximo todos os recursos de cada uma.

Alguns processos de automação em revisões sistemáticas estão bem difundidos no Brasil. Ferramentas como Covidence e Rayyan são utilizadas na deduplicação e seleção dos estudos, enquanto Zotero, Mendeley e Endnote são adotados para o gerenciamento de referências (Guimarães *et al.*, 2022).

No que se refere à automação na construção de estratégias de busca, destacam-se duas ferramentas recomendadas pela Cochrane: MEDLINE Transpose e Polyglot (Lefebvre *et al.*, 2024). A ferramenta é utilizada para traduzir automaticamente estratégias de busca entre diferentes bases de dados científicas, poupando tempo e reduzindo erros manuais em revisões sistemáticas e outros estudos. A JBI também recomenda o Polyglot como ferramenta útil para tradução de estratégias (Aromataris *et al.*, 2024).

O Polyglot é amplamente utilizado na América Latina, com destaque para Lima (Peru), entre as 10 cidades com maior número de acessos, e o Brasil entre os 10 países que mais utilizam a ferramenta. O Systematic Review Accelerator, plataforma que hospeda o Polyglot, já foi acessado por mais de 12 mil usuários da América Latina.

Diversos países mantêm bases bibliográficas voltadas à produção científica local, muitas vezes com periódicos e documentos não indexados em bases internacionais. Existem também bases temáticas, como AMED (medicina complementar), CINAHL (enfermagem e saúde aliada) e APA PsycInfo (psicologia e psiquiatria). É altamente recomendável incluir bases nacionais, regionais e temáticas nas estratégias de busca — como a LILACS, no contexto da América Latina e Caribe (Higgins *et al.*, 2023).

A LILACS é a principal base de dados em saúde da América Latina e Caribe, reunindo a produção científica da região. Integrada ao portal da BVS e mantida pela BIREME/OPAS/OMS, tem como objetivo dar visibilidade à literatura local e promover o acesso equitativo à informação. A busca é realizada por meio da plataforma iAHx, que permite o uso de filtros, operadores booleanos e campos específicos como título, resumo, assunto e autor, proporcionando recuperação precisa de documentos.

A inclusão da LILACS em estratégias de busca pode melhorar significativamente a qualidade das revisões sistemáticas. Um estudo mostrou que, em 60% dos casos analisados,

a LILACS continha artigos relevantes que não haviam sido incluídos nas revisões publicadas (Clark; Castro; Atallah, 1998).

Nas revisões sistemáticas e de escopo, é necessário adaptar a estratégia da base principal para outras bases relevantes, além de incluir fontes de literatura cinzenta quando previsto. Não há uma única fonte para síntese de evidências (Solomons; Hinton, 2021), nem um número fixo de bases recomendadas, embora a maioria das revisões inclua ao menos três — número que tem crescido ao longo dos anos (Lam; McDiarmid, 2016).

A adaptação entre bases exige atenção às diferenças de vocabulário, operadores e sintaxes (Lefebvre *et al.*, 2025). Diferenças em truncamentos e campos de data podem impactar os resultados. Por isso, recomenda-se apoio técnico de bibliotecários especializados para revisar ou adaptar estratégias corretamente.

3 METODOLOGIA

Foi realizado um estudo metodológico conforme definição de (Mbuagbaw *et al.* 2020), que considera como tal qualquer estudo que descreva ou analise métodos de pesquisa publicados ou não, abrangendo desde a formulação de perguntas até a análise de credibilidade.

A proposta de integração da LILACS ao Polyglot surgiu da necessidade de ampliar o uso de estratégias automatizadas para bases regionais. A ausência da LILACS limitava a interoperabilidade da ferramenta com a literatura latino-americana. Diante disso, a equipe brasileira entrou em contato com o responsável técnico pelo Polyglot para discutir a incorporação da base.

A metodologia adotada foi qualitativa e descritiva, com base em análise documental detalhada da sintaxe da LILACS, incluindo a identificação dos campos de busca, operadores booleanos, filtros disponíveis e especificidades do vocabulário controlado, conforme documentação oficial (BIREME, [2021]).

Os critérios utilizados nos testes práticos de validação incluíram: (a) verificação da correta tradução dos campos de busca; (b) correspondência de operadores booleanos e de proximidade; (c) preservação da lógica da estratégia original; e (d) análise de eventuais perdas ou alterações de resultados decorrentes da tradução automática. Por exemplo, ao utilizar no PubMed o descritor de assunto “Hypertension”[MeSH], o Polyglot traduzirá automaticamente

para a LILACS o equivalente MH: “Hipertensão”, preservando a lógica da busca e permitindo a recuperação adequada dos registros indexados.

Em seguida, a equipe preencheu a planilha técnica de integração fornecida pela Bond University, com suporte da equipe da BIREME/OPAS para esclarecer aspectos técnicos da interface e compatibilidade com o PubMed.

Trata-se de um processo extenso, que exige a identificação de diversos elementos. O Polyglot utiliza uma tabela de correspondência dividida em seis áreas: códigos de campo, descritores, traduções de termos MeSH, comandos de busca, intervalos de data e combinação entre linhas. A tabela contém 2.500 células de dados que garantem a tradução entre dezesseis bases utilizadas em revisões sistemáticas — sendo a LILACS agora uma delas.

A tabela é mantida por especialistas em busca, que testam cuidadosamente as traduções antes de convertê-las para JavaScript e integrá-las à versão web do Polyglot.

Após o envio da planilha, a equipe de desenvolvimento implementou a codificação e disponibilizou a funcionalidade em ambiente de testes. Foram então realizadas buscas simuladas e validações com estratégias reais da BVS Prevenção e Controle do Câncer.

Quanto à reprodutibilidade, a metodologia empregada permite que outras instituições realizem processo semelhante, desde que tenham acesso à documentação técnica da base de dados, à planilha de integração e a uma equipe com conhecimento especializado em estratégias de busca. As principais limitações metodológicas incluem a dependência da validação manual por bibliotecário e analista de sistemas para garantir a precisão, bem como a necessidade de ajustes posteriores caso a sintaxe da LILACS ou do Polyglot seja atualizada.

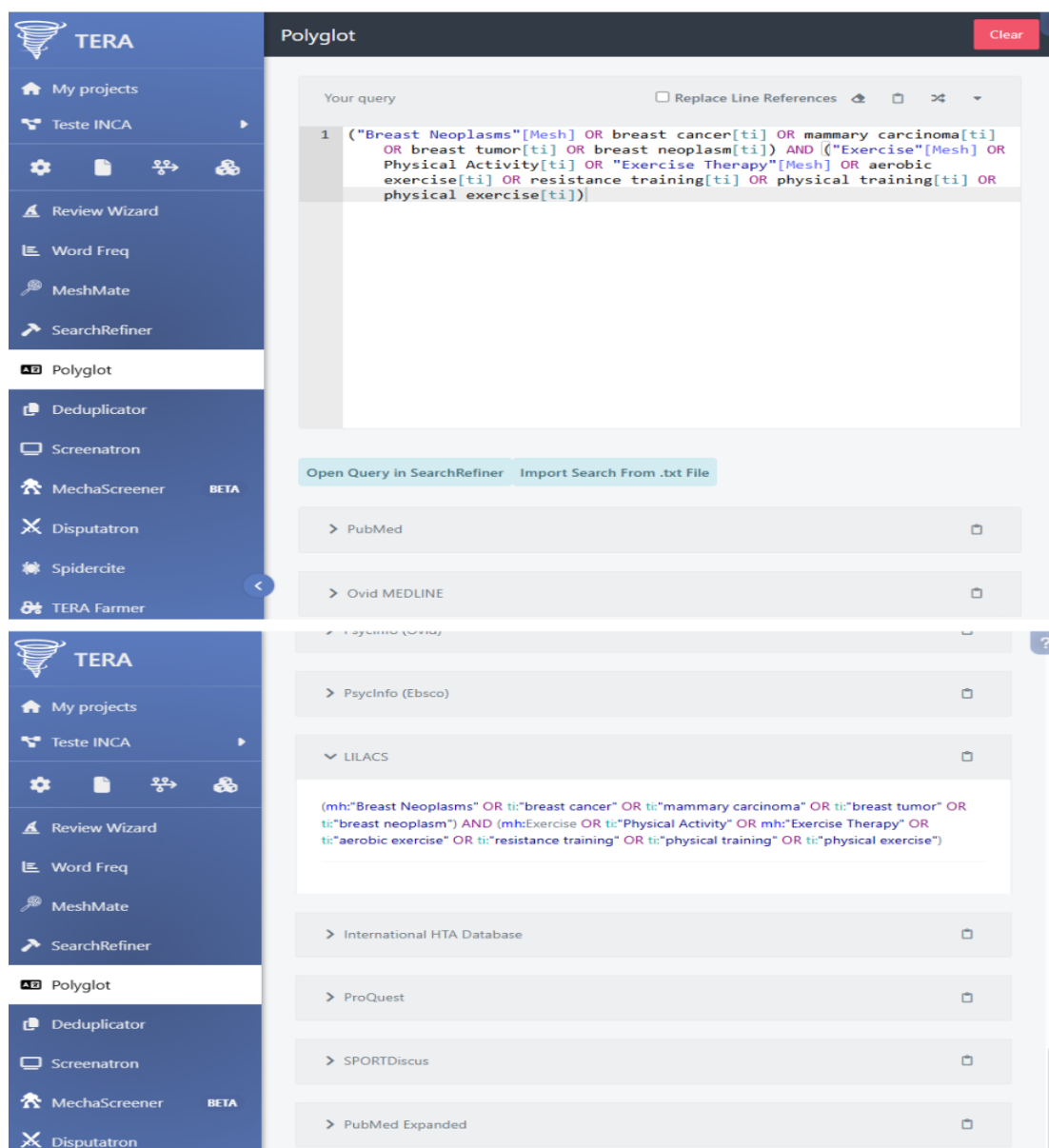
4 RESULTADOS

Os testes mostraram que a tradução automática para a LILACS funcionou adequadamente quanto aos marcadores de campo e operadores básicos. As limitações práticas identificadas incluem a necessidade de ajustes manuais em casos de descritores hierárquicos, operadores de proximidade específicos e estratégias complexas com linhas numeradas. Vale ressaltar que essas limitações decorrem da base de dados LILACS, e não do Polyglot. Essas questões foram comunicadas para futuras melhorias. Ainda assim, a integração já se mostrou funcional como ponto de partida, desde que acompanhada por revisão bibliotecária, conforme já ocorre com outras bases no Polyglot.

O impacto na prática informacional foi observado principalmente na agilidade em incluir a LILACS em estratégias de busca, reduzindo o tempo de adaptação manual e ampliando a visibilidade das revisões sistemáticas que utilizam essa base de dados.

Após três meses do início do trabalho, a LILACS foi incorporada à ferramenta, estando disponível na nova versão da plataforma. A figura a seguir apresenta o funcionamento do Polyglot com a LILACS a partir da inserção de uma estratégia de busca no Pubmed (figura 1).

Figura 1 - Funcionamento do Polyglot com a LILACS a partir da inserção de uma estratégia de busca no Pubmed



Descrição: A imagem apresenta o sistema TERA, com o módulo Polyglot aberto para traduzir estratégias de busca entre bases de dados. No topo, há uma estratégia de busca para a base PubMed sobre câncer de mama e exercícios físicos, utilizando vocabulário controlado, sinônimos e operadores booleanos (AND, OR). Abaixo,

aparece um exemplo de tradução automática para outras bases de dados, incluindo a LILACS. **Fonte:** Captura de tela do sistema TERA (2025)

A estratégia utilizada no Pubmed foi: ("Breast Neoplasms"[Mesh] OR breast cancer[ti] OR mammary carcinoma[ti] OR breast tumor[ti] OR breast neoplasm[ti]) AND ("Exercise"[Mesh] OR Physical Activity[ti] OR "Exercise Therapy"[Mesh] OR aerobic exercise[ti] OR resistance training[ti] OR physical training[ti] OR physical exercise[ti]).

Na LILACS, por meio do Polyglot, a estratégia de busca ficou traduzida automaticamente para: (mh:"Breast Neoplasms" OR ti:"breast cancer" OR ti:"mammary carcinoma" OR ti:"breast tumor" OR ti:"breast neoplasm") AND (mh:Exercise OR ti:"Physical Activity" OR mh:"Exercise Therapy" OR ti:"aerobic exercise" OR ti:"resistance training" OR ti:"physical training" OR ti:"physical exercise").

5 CONSIDERAÇÕES FINAIS

A integração da LILACS ao Polyglot Search Translator representa um avanço no fortalecimento da infraestrutura de busca por evidências em saúde na América Latina. A iniciativa contribui para a modernização do serviço de estratégias de busca do centro de referência em oncologia do Brasil, promovendo mais agilidade, qualidade e visibilidade à produção científica regional — e certamente poderá beneficiar outros serviços similares na região.

Embora algumas limitações da LILACS exijam ajustes manuais ou revisão especializada, a funcionalidade atual da ferramenta é robusta o suficiente para apoiar usuários e bibliotecários na elaboração inicial de estratégias. Estudos futuros poderão aprofundar o impacto da integração no contexto das revisões que incluam a LILACS como fonte de evidência.

Além disso, essa ação reforça o papel da LILACS na BVS e amplia seu reconhecimento internacional como fonte confiável de informação científica. Espera-se que essa experiência inspire outras iniciativas de integração entre ferramentas automatizadas e bases regionais, promovendo maior equidade no acesso à informação em saúde baseada em evidências.

REFERÊNCIAS

ABDALA, V.; FIGUEIRÓ, M. Bases de dados de literatura científica e estratégias de busca. *In*: TOMA, T. S.; PEREIRA, T. da V.; VANNI, T.; BARRETO, J. O. M. (org.). **Avaliação de tecnologias**

de saúde & políticas informadas por evidências. São Paulo: Instituto de Saúde, 2017. Acesso em: 19 maio 2025.

AROMATARIS, E; LOCKWOOD, C.; PORRITT, K.; PILLA, B.; JORDAN, Z. **JBIManual for Evidence Synthesis**. [S.l.]: JBI, 2024. Disponível em: <https://synthesismanual.jbi.global>. Acesso em: 20 maio 2025.

BIREME. **Tabela 2.1: ampos de dados do formato utilizados para pesquisa no iAHx**. [S. l.]: BIREME, [2021]. Disponível em: <https://bvsalud.org/searchtutorial/EN/documents/tabela.2.1.pdf>. Acesso em: 20 maio 2025.

CLARK, J. M. *et al.* Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. **Journal of the Medical Library Association**, Baltimore, v. 108, n. 2, p. 195–207, Apr. 2020. DOI: <https://doi.org/10.5195/jmla.2020.834>.

CLARK, O. A. C.; CASTRO, A. A.; ATALLAH, A. N. Searching LILACS database improves systematic reviews. *In*: COCHRANE COLLOQUIUM, 6., 1998, Baltimore. **Anais [...]**. Baltimore: [s. n.], 1998. Disponível em: <https://abstracts.cochrane.org/1998-baltimore/searching-lilacs-database-improves-systematic-reviews>. Acesso em: 19 maio 2025.

GUIMARÃES, N. S. *et al.* Deduplicating records in systematic reviews: there are free, accurate automated ways to do so. **Journal of Clinical Epidemiology**, Amsterdam, v. 152, p. 110–115, dez. 2022. DOI: <https://doi.org/10.1016/j.jclinepi.2022.10.009>.

HIGGINS, J. P. T.; LASSERSON, T.; THOMAS, J.; FLEMYNG, E.; CHURCHILL, R. Methodological Expectations of Cochrane Intervention Reviews. **Cochrane**: London, 2023. Disponível em: <https://community.cochrane.org/mecir-manual>. Acesso em: 19 maio 2025.

LAM, M. T.; McDIARMID, M. Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014. **Journal of the Medical Library Association**, Baltimore, v. 104, n. 4, p. 284–289, 2016. DOI: 10.3163/1536-5050.104.4.006.

LEFEBVRE, C. *et al.* Searching for and selecting studies. *In*: HIGGINS, J. P. T. *et al.* (ed.). **Cochrane handbook for systematic reviews of interventions version 6.5.1**. [S.l.]: Cochrane, 2025. Disponível em: <https://www.cochrane.org/handbook>. Acesso em: 19 maio 2025.

LEFEBVRE, C. *et al.* Technical supplement to chapter 4: searching for and selecting studies. *In*: HIGGINS, J. P. T. *et al.* (ed.). **Cochrane Handbook for Systematic Reviews of Interventions version 6.5**. [S.l.]: Cochrane, 2024. Disponível em: <https://www.cochrane.org/handbook>. Acesso em: 19 maio 2025.

MBUAGBAW, L.; LAWSON, D.O.; PULJAK, L.; ALLISON, D. B.; THABANE, L. A tutorial on methodological studies: the what, when, how and why. **BMC Medical Research Methodology**, [s.l.], v. 20, p. 226, 2020. cap. 4. DOI: <https://doi.org/10.1186/s12874-020-01107-7>

SOLOMONS, T.; HINTON, E. Federated searches: why a one-stop shop approach to literature searching falls short for evidence synthesis. **JBIM Evidence Synthesis**, Adelaide, v. 19, n. 6, p. 1259–1262, 2021. DOI: <https://doi.org/10.11124/JBIES-21-00177>