

10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



A INDEXAÇÃO AUTOMÁTICA NA CIÊNCIA DA INFORMAÇÃO BRASILEIRA:

uma análise bibliométrica comparativa (2013–2025)

Raimundo Nonato Macedo dos Santos

<https://orcid.org/0000-0002-9208-3266>
raimundo.macedo@ufpe.br
Universidade Federal de Pernambuco
(UFPE)

Remi Correia Lapa

<https://orcid.org/0009-0000-7333-0121>
remi.lapa@ufpe.br
Universidade Federal de Pernambuco
(UFPE)

Cíntia Maria Silva de Holanda

<https://orcid.org/0009-0005-7217-3546>
cinthia.holanda@ufpe.br
Universidade Federal de Pernambuco
(UFPE)

Renato Fernandes Corrêa

<https://orcid.org/0000-0002-9880-8678>
renato.correa@ufpe.br
Universidade Federal de Pernambuco
(UFPE)

Resumo:

Este estudo analisa a produção científica brasileira sobre indexação automática na Ciência da Informação entre 2013 e 2025, comparando-a ao período 1973–2012. Mediante investigação bibliométrica de 65 documentos da BRAPCI e Google Acadêmico, identificam-se estabilidade produtiva, fortalecimento de redes interinstitucionais e transição metodológica para abordagens híbridas baseadas em linguística computacional e aprendizado de máquina.

Palavras-Chave: Indexação automática. Ciência da Informação. Bibliometria. Inteligência artificial. Produção científica brasileira.

EIXO TEMÁTICO:

Produtividade e Colaboração Científica

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1031

COMO CITAR ESTE ARTIGO:

SANTOS, Raimundo Nonato Macedo dos; HOLANDA, Cíntia Maria Silva de; LAPA, Remi Correia; CORRÊA, Renato Fernandes. A indexação automática na Ciência da Informação brasileira: uma análise bibliométrica comparativa (2013-2025). In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. Anais [...]. Curitiba, 2026. DOI: 10.22477/x.ebbc.1031. Disponível em: <https://doi.org/10.22477/x.ebbc.1031>.



1 INTRODUÇÃO

A indexação automática consolidou-se como resposta às limitações da indexação intelectual frente ao crescimento da produção científica, focando em celeridade, padronização e escalabilidade (Nascimento; Martins; Albuquerque, 2023; Borges, 2010; Gil-Leiva, 1997). As abordagens iniciais fundamentaram-se em métodos estatísticos e coocorrência de termos (Stevens, 1970), evoluindo para aplicação de aprendizado de máquina e técnicas de linguística computacional que hoje centralizam os debates sobre organização e acesso ao conhecimento (Lancaster, 2004).

No Brasil, um panorama de 1973 a 2012, realizado por Corrêa e Lapa em 2013, identificou padrões de institucionalização e tendências temáticas que hoje exigem atualização frente aos avanços das Tecnologias da Informação e Comunicação. Este estudo propõe uma análise bibliométrica comparativa (2013–2025) para verificar a reconfiguração do campo, revisitando os resultados do estudo publicado em 2015 para preencher lacunas sobre a obsolescência tecnológica e a mudança de paradigma — da estatística clássica para aplicação de técnicas de linguística computacional, arquiteturas de aprendizado de máquina e Modelos de Linguagem de Grande Escala (LLMs). Busca-se verificar se a maturidade da área resultou em autonomia sistêmica ou na persistência de métodos híbridos, oferecendo uma avaliação longitudinal da eficácia dessas tecnologias no cenário nacional.

2 METODOLOGIA

Esta pesquisa é um estudo quali-quantitativo e exploratório que utiliza a análise bibliométrica para mapear a produção brasileira sobre indexação automática, em diálogo comparativo com o panorama de 1973 a 2012 de Corrêa e Lapa (2013). A abordagem exploratória permite explicitar tendências e reconfigurações do cam-

po (Gil, 2010). Trata-se de uma pesquisa bibliográfica baseada no levantamento sistemático de artigos e anais de eventos brasileiros em Ciência da Informação escritos em língua portuguesa (Marconi; Lakatos, 2010). Optou-se pela exclusão de teses, livros e literatura cinzenta para garantir paridade comparativa e o rigor da avaliação por pares (Glänzel, 2003). Embora essa delimitação possa ocultar nuances da evolução epistemológica inicial, a concentração em periódicos e eventos assegura o mapeamento do conhecimento consolidado.

O *corpus* foi composto via BRAPCI e Google Acadêmico, utilizando os termos “indexação automática”, “automatização da indexação” e “indexação semi-automática” entre 2013 e 2025. A base PERI foi excluída por incompatibilidade temporal. Reconhece-se que o uso da BRAPCI e do Google Acadêmico impõe desafios de cobertura desigual e duplicidade; contudo, realizou-se depuração manual rigorosa para normalização de autorias e exclusão de redundâncias. O *corpus* final, de 65 documentos, foi exportado em formatos *txt* e *bibtex* para análise bibliométrica de autores, instituições, fontes e sistemas.

3 RESULTADOS E DISCUSSÕES

A análise dos 65 documentos, publicados nos anos de 2013–2025, revela, frente aos resultados da análise de documentos publicados entre 1973–2012, elementos de continuidade e deslocamentos metodológicos rumo a maior aplicação de técnicas de Inteligência Artificial. Autores como Nascimento, Martins e Albuquerque (2023), Fujita e Sousa (2025), Sousa e Fujita (2025), Pereira e Gabriel Junior (2025), Costa, Arakaki e Nascimento (2025) e Sousa (2025) discutem essa integração progressiva.

3.1 EVOLUÇÃO TEMPORAL DA PRODUÇÃO CIENTÍFICA

A distribuição anual indica manutenção do interesse acadêmico, com média de 5,38 trabalhos por ano. A regularidade sugere estabilização do campo, refletindo a recontextualização de problemas clássicos de seleção de termos diante da aplicação de novos recursos computacionais.

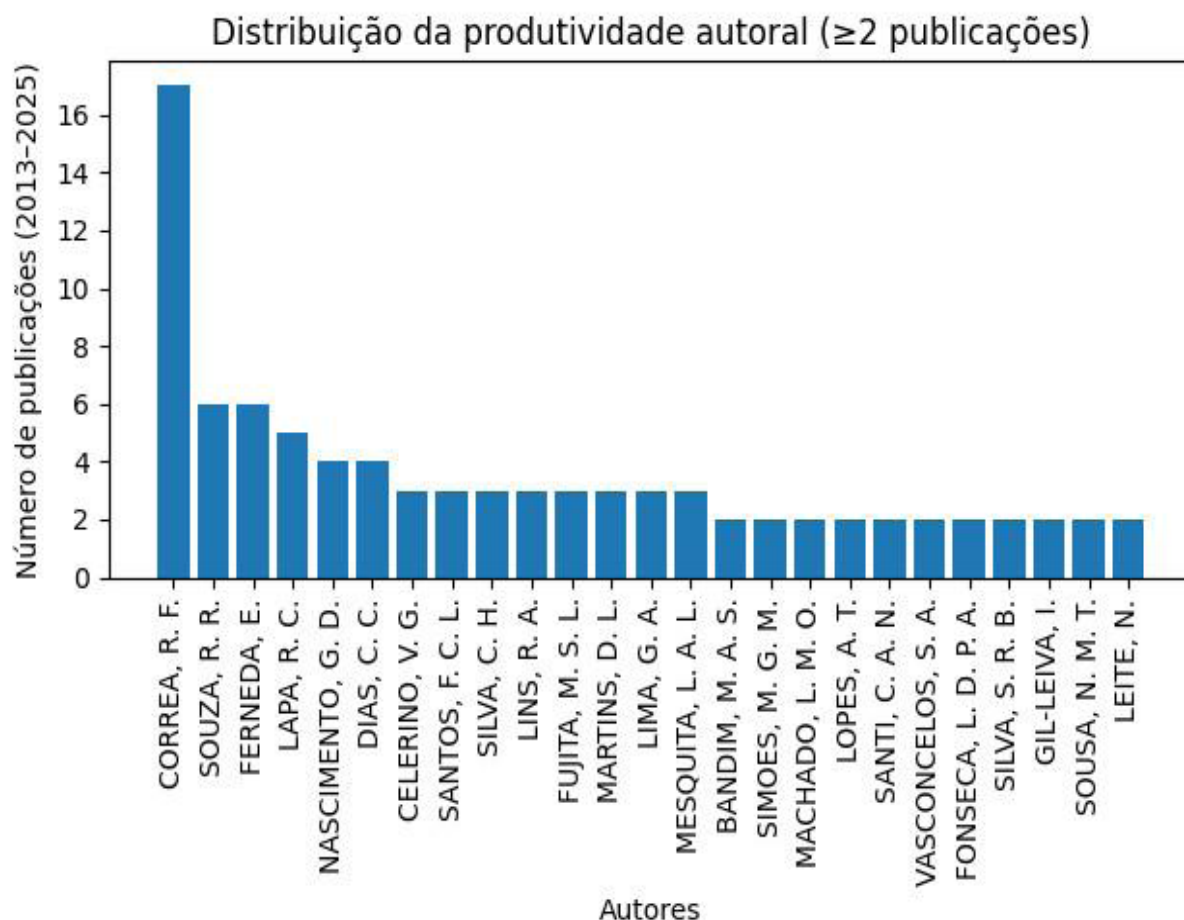
3.2 CANAIS DE COMUNICAÇÃO CIENTÍFICA

Observa-se a predominância de artigos em periódicos (67,69%), seguidos por anais de eventos (32,31%). Essa distribuição mantém proximidade com o recorte anterior, indicando estabilidade nas formas de comunicação adotadas na área.

3.3 AUTORES MAIS PRODUTIVOS E RELAÇÕES DE COLABORAÇÃO (2013–2025)

Identificou-se 89 autores e 163 ocorrências autorais, Figura 1. A produtividade adere à Lei de Lotka, com forte concentração em poucos pesquisadores. O índice de hápax reduziu para 71,9%, enquanto o núcleo de continuidade subiu para 28,1% (contra 22% no período anterior), sugerindo maior permanência no campo. Destaca-se CORREA, R. F. (17 publicações), seguido por SOUZA, R. R. e FERNEDA, E. (6 cada) e LAPA, R. C. (5). A coautoria é elemento estruturante, evidenciando integração interinstitucional e entre níveis de experiência.

Figura 1 – Distribuição da produtividade autorial



Fonte: Os autores (2026).

3.4 PERIÓDICOS E EVENTOS COM MAIOR CONCENTRAÇÃO DE PUBLICAÇÕES

O Encontro Nacional de Pesquisa em Ciência da Informação ENANCIB é o principal canal (23,1%), seguido pelos periódicos Encontros Bibli e Perspectivas em Ciência da Informação (A1), que somam 16,9%, Quadro 1. O evento ISKO Brasil também se destaca por ter o mesmo número de trabalhos publicados que o periódico Perspectivas em Ciência da Informação. A dispersão em fontes interdisciplinares (30,8%) aponta para a inserção da temática em circuitos científicos mais amplos, como a Computação.

Quadro 1 – Principais fontes de publicação (2013–2025)

FONTE DE PUBLICAÇÃO	F	%	QUALIS
ENANCIB	15	23,1	-
Encontros Bibli	6	9,2	A1
ISKO BRASIL	5	7,7	-
Perspectivas em Ciência da Informação	5	7,7	A1
Ciência da Informação	4	6,2	A2
TransInformação	2	3,1	A1
Informação & Tecnologia	2	3,1	B2
Em Questão	2	3,1	A1
Ciência da Informação em Revista	2	3,1	A3
ASKLEPION Informação em Saúde	2	3,1	B1
Outras fontes (≤ 1 ocorrência)	20	30,8	-
TOTAL	65	100	-

Fonte: Os autores (2026).

3.5 INSTITUIÇÕES PRODUTORAS E REDES DE COOPERAÇÃO

Houve expansão institucional e descentralização geográfica, com aumento de parcerias e fortalecimento de programas. A UNESP consolida-se como núcleo articulador central de uma rede que inclui UnB, IFPR, IFRJ, UC, UM, UC3M e UFPB, Figura 2. A colaboração interinstitucional é superior ao período precedente, integrando universidades nacionais com internacionais como a Universidade de Murcia na Espanha e a Universidade de Coimbra em Portugal.

The graph illustrates a network of connections between various Brazilian universities. The central node is UNESP. Other nodes include UFPA, UC, UFMG, UP, FAP, BT, FID, CRUZ, CERETIMG, UFCC, UFPE, UFAL, UM, IFRR, UNB, UFV, IFSP, IFPR, UC3M, UFPA, UES, and UFPA.

3.6 SISTEMAS, MÉTODOS E PROGRAMAS UTILIZADOS (2013-2025)

10º EBBC, Curitiba - PR, Jul. 2026

Quadro 2 – Sistemas, métodos e programas (2013–2025)

SISTEMA / PROGRAMA / MÉTODO	FREQUÊNCIA	CATEGORIA PREDOMINANTE
SISA	6	Sistema especialista / Atribuição supervisionada
Annif	4	Aprendizado de máquina / Framework
OGMA	2	Linguístico (extração de sintagmas nominais)
MAUI	2	Aprendizado de máquina
FintoAI	2	Inteligência Artificial
Kratt	2	Inteligência Artificial
Leis de Zipf / Ponto T de Goffman	1	Estatístico clássico
PyPLN	1	Processamento de Linguagem Natural
b2	1	Computação evolucionária
Portal Semântico	1	Sistema baseado em ontologia
Ollama	1	Modelos de Linguagem de Grande Escala (LLM)

Fonte: Os autores (2026).

4 CONCLUSÕES

A análise da produção científica brasileira sobre indexação automática (2013–2025) revela um campo em plena consolidação e reconfiguração tecnológica. Confrontando-se os dados atuais com os reportados no levantamento de 1973–2012, observa-se que o interesse acadêmico se manteve estável, com média de 5,38 trabalhos anuais, sugerindo a integração definitiva da temática na agenda de pesquisa da Ciência da Informação no Brasil. Contudo, as dinâmicas de autoria e as bases metodológicas sofreram transformações que evidenciam o amadurecimento da área.

O fortalecimento das redes de colaboração e das lideranças científicas é um dos principais achados. Verificou-se maior permanência de pesquisadores e a intensificação da concentração produtiva em polos de excelência, com destaque para CORREA, R. F. como autor mais produtivo. Paralelamente, ocorreu expansão institucional e descentralização geográfica, com a UNESP consolidada como núcleo articulador de uma rede interinstitucional que envolve diversas universidades tanto nacionais quanto internacionais.

Metodologicamente, o período marca um deslocamento de paradigma: o cenário centrado em sistemas “fechados” e estatísticos deu lugar a abordagens modulares e híbridas, integrando modelos de aprendizado de máquina da Inteligência Artificial, como por exemplo o *framework* Annif e modelos de linguagem de grande escala, como o Ollama. Essa transição reflete uma sofisticação técnica, que instiga a necessidade de que tais processos automáticos sejam avaliados sob rigoroso controle experimental via métricas de precisão e revocação.

Sob o prisma prático, os achados oferecem subsídios para a otimização de fluxos em repositórios institucionais de larga escala, demonstrando que a transição para métodos híbridos é estratégica para viabilizar o tratamento documental diante do volume crescente de dados científicos no país. Em suma, a indexação automática evoluiu para um estágio de investigação interdisciplinar robusto e alinhado às transformações globais. Este panorama subsidia futuras pesquisas sobre os impactos técnicos e éticos das novas tecnologias aplicadas em soluções eficazes para a crescente produção científica nacional.

REFERÊNCIAS

BORGES, G. B. Indexação automática de documentos textuais: proposta de critérios essenciais. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 11, 2010, Rio de Janeiro. **Anais** [...]. Rio de Janeiro: ANCIB, 2010.

Disponível em: <https://sl1nk.com/repositorio-ufmg-br-borges>. Acesso em: 16 fev. 2026.

COSTA, B. C. C.; ARAKAKI, A. C. S.; NASCIMENTO, R. L. S. Aprimoramento do Annif para indexação automática de publicações em português. *In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA*, 8, 2025, Marília, SP, p. 27-29. **Anais [...]**. Marília, SP: WIDaT, 2025. DOI: 10.22477/viii.widat.273. Disponível em: <https://widat.ibict.br/widat2025/article/view/273>. Acesso em: 16 fev. 2026.

CORRÊA, R. F.; LAPA, R. C. Panorama de estudos sobre indexação automática no âmbito da ciência da informação no Brasil (1973-2012). **Ciência da Informação**, v. 42, n. 2, 2013. DOI: <https://doi.org/10.18225/ci.inf.v42i2.1385>. Disponível em: <https://revista.ibict.br/ciinf/article/view/1385/1563>. Acesso em: 06 mar. 2026.

FUJITA, M. S. L.; SOUSA, N. M.T. Vocabulário controlado e inteligência artificial na indexação: uma revisão bibliográfica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 30, Fluxo Contínuo, 2025. DOI: 10.1590/1981-5344/56745. Disponível em: <https://www.scielo.br/j/pci/a/br3MytQmVck88FTJPgTJvvc/?lang=pt>. Acesso em: 16 fev. 2026.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GIL-LEIVA, I. **La automatización de la indización, propuesta teórico-práctica: aplicación al área de Biblioteconomía y Documentación**. 1997. 268 f. Tese (Doutorado) – Universidad de Murcia, Murcia, 1997. Disponível em: <https://webs.um.es/isgil/resources/PhDissertation%20Gil-Leiva.pdf>. Acesso em: 15 fev. 2026.

GLÄNZEL, W. **Bibliometrics as a research field: a course on theory and application of bibliometric indicators**. Belgica: [s.n.], 2003. Disponível em: https://nsdl.niscpr.res.in/bitstream/123456789/968/1/Bib_Module_KUL.pdf. Acesso em: 15 fev. 2026.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos, 2004.

MARCONI, M. A.; LAKATOS, E. M. **Metodologia do trabalho científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos**. 7. ed., 5. reimpr. São Paulo: Atlas, 2010.

NASCIMENTO, G. D.; MARTINS, G. K.; ALBUQUERQUE, M. E. B. C. Automação da indexação: evidências e tendências da produção científica indexada na Brapci. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**. Florianópolis/SC, Brasil, v. 28, p. 1-20, 2023. DOI: /10.5007/1518-2924.2023.e91956. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/91956>. Acesso em: 15 fev. 2026.



PEREIRA, J. C. S. S. J.; GABRIEL JUNIOR, R. F. Inteligência artificial aplicada à organização do conhecimento: uma taxonomia. *In*: INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION, 8, 2025, Canela, RS. **Anais [...]**. Canela: ISKO Brasil, 2025. DOI: [https://doi.org/ 10.12303/ISKO25.112](https://doi.org/10.12303/ISKO25.112) . Disponível em: <https://isko.org.br/ojs/index.php/iskobrasil/article/view/112>. Acesso em: 16 fev. 2026.

SOUSA, N. M. T. A organização do conhecimento na era da inteligência artificial: uma revisão sistemática e análise bibliométrica. **Brazilian Journal of Information Science: research trends**, v. 19, 2025, e025032. DOI: 10.36311/1981-1640.2025.v19.e025032. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/17322>. Acesso em: 16 fev. 2026.

SOUSA, N. M. T.; FUJITA, M. S. L. Utilização da inteligência artificial como apoio aos bibliotecários. *In*: INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION, 8, 2025, Canela, RS. **Anais [...]**. Canela: ISKO Brasil, 2025. DOI: 10.22477/ISKO25.23. Disponível em: <https://isko.org.br/ojs/index.php/iskobrasil/article/view/23>. Acesso em: 16 fev. 2026.

STEVENS, M. E. **Automatic indexing**: a state-of-the-art report. Washington, D.C.: National Bureau of Standards Monograph 91, 1970. DOI: 10.6028/NBS.MONO.91. Disponível em: <https://l1nq.com/nvlpubs-nist-gov>. Acesso em: 15 fev. 2026.