


VOCABULÁRIO CONTROLADO E INTELIGÊNCIA ARTIFICIAL NA INDEXAÇÃO: UMA REVISÃO BIBLIOGRÁFICA

Mariângela Spotti Lopes Fujita
 <http://lattes.cnpq.br/6530346906709462> –  <https://orcid.org/0000-0002-8239-7114>
mariangela.fujita@unesp.br
Universidade Estadual Paulista Júlio de Mesquita Filho, Campus de Marília (UNESP)
Marília, São Paulo, Brasil

Nuno Miguel Teixeira Sousa
 <https://orcid.org/0000-0002-5146-8514>
nunomsousa@fd.uc.pt
Universidade de Lisboa
Lisboa, Portugal

RESUMO

Com a inteligência artificial, os bibliotecários encontraram um aliado na indexação para organizar vastos conjuntos de dados. Com a questão: como a inteligência artificial pode favorecer os bibliotecários na indexação? – o objetivo é compreender a relação entre a inteligência artificial e a indexação. Foi adotada abordagem qualitativa em pesquisa bibliográfica que, após critérios de inclusão (título, resumo, palavras-chave e pertinência temática) e exclusão (duplicados e artigos sem acesso integral), foram analisados 27 artigos (17 WoS; 10 EBSCO). Constatou-se que: (i) a inteligência artificial e/ou a indexação automática não substituem os bibliotecários, mas funcionam como apoio; (ii) a indexação automática evoluiu de forma positiva com ganhos de precisão e consistência; (iii) existem ferramentas eficazes aplicadas em bibliotecas, como o **Annif**, o **FintoAI** e o **Kraft**; (iv) estas ferramentas desempenham um papel relevante na modernização dos processos bibliotecários; e (v) torna-se inevitável que as bibliotecas invistam em condições de formação contínua e implementação de soluções de IA. A utilização da inteligência artificial na indexação representa uma oportunidade estratégica para a eficiência e a qualidade dos serviços bibliotecários. O impacto da IA aproxima a indexação da linguagem natural dos utilizadores, preservando o papel do bibliotecário e ampliando sua atuação na gestão da informação.

Palavras-chave: Inteligência Artificial. Bibliotecários. Indexação manual. Indexação automática. Representação e recuperação da informação. Bibliotecas.

CONTROLLED VOCABULARY AND ARTIFICIAL INTELLIGENCE IN INDEXING: A LITERATURE REVIEW

ABSTRACT

With artificial intelligence, librarians have found an ally in indexing to organize vast data sets. With the question: how can artificial intelligence help librarians with indexing? – the goal is to understand the relationship between artificial intelligence and indexing. A qualitative approach was adopted in bibliographic research which, after inclusion criteria (title, abstract, keywords, and thematic relevance) and exclusion criteria (duplicates and articles without full access), analyzed 27 articles (17 WoS; 10 EBSCO). It was found that: (i) artificial intelligence and/or automatic indexing do not replace librarians, but function as support; (ii) automatic indexing has evolved positively with gains in accuracy and consistency; (iii) there are effective tools applied in libraries, such as **Annif**, **FintoAI**, and **Kraft**; (iv) these tools play an important role in the modernization of library processes; and (v) it is inevitable that libraries will invest in continuing education and the implementation of AI solutions. The use of artificial intelligence in indexing represents a strategic opportunity for the efficiency and quality of library services. The impact of AI brings indexing closer to the natural language of users, preserving the role of the librarian and expanding their role in information management.

Keywords: Artificial Intelligence. Librarians. Manual indexing. Automatic indexing. Information representation and retrieval. Libraries.

DOI <http://dx.doi.org/10.1590/1981-5344/56745>

Recebido em: 27/12/2024
Aceito em: 14/10/2025

1 INTRODUÇÃO

Ao longo do tempo, fomos assistindo a mudanças na forma como desenvolvemos a indexação com o propósito de organizarmos e recuperarmos a informação que nos chega.

A transformação de métodos de indexação leva-nos a abordar este assunto como determinante para os bibliotecários com destaque para a aplicação da inteligência artificial, aspecto de grande impacto na indexação. De fato, estamos perante uma transição qualitativa do Processamento da Linguagem Natural (PLN) entre a Indexação Manual (IM) e a Indexação Automática (IA). Conscientes que falamos, igualmente, numa diferença de recursos de há anos para atualmente, acreditamos que esta evolução veio demonstrar a necessidade de vermos presente a criação de ferramentas que venham providenciar um controlo de qualidade cada vez mais focado e adaptado à linguagem natural, ou seja, à linguagem dos utilizadores cujos termos devem ser facilmente utilizáveis e intuitivos (NISO Z39.4, 2021 citado em Chen; Bullard; Giustani, 2023).

Falamos, deste modo, num campo que continuamente tem sentido a obrigação de se verem discutidos estudos (Chu, 2010) que permitam abordar abertamente a forma como desenvolvemos a indexação e como podemos aperfeiçoá-la através da utilização de ferramentas de inteligência artificial que consigam apoiar os bibliotecários nesta tarefa onde cada vez mais se mostra ser complexa devido ao uso de diferentes tipos de linguagens de indexação que variam de acordo com a interpretação destes profissionais, daí a importância de estarmos aqui a falar sobre a utilização da inteligência artificial na indexação a fim de conseguirmos alcançar uma ferramenta comum que ajude os bibliotecários na representação e recuperação da informação revelando-se, assim, uma importante aliada (Gil-Leiva *et al.*, 2022) neste procedimento.

Neste sentido, procuramos com esta investigação responder à questão de partida que nos leva a debatermos este assunto: **como a inteligência artificial pode favorecer os bibliotecários na indexação?** Procurando, assim, compreender a relação contemporânea entre a inteligência artificial e a indexação que apoie a atuação profissional do bibliotecário. Para tal, foram tidas em conta duas bases de dados (*Web of Science* (WoS) e *Library* e

Information Science Source (EBSCO)) onde desenvolvemos uma abordagem qualitativa partindo deste modo para uma pesquisa bibliográfica que se dividiu em duas fases: uma primeira relativa ao ponto seguinte ($TS=(\text{"artificial intelligence"} \text{ AND "automatic indexing"})$) e uma segunda relativa ao nosso foco ($TS=(\text{"artificial intelligence"} \text{ AND "indexing"} \text{ OR "manual indexing"} \text{ OR "automatic indexing"} \text{ AND "librarians"} \text{ AND "libraries"})$) tendo nesta última expressão sido utilizadas como técnicas de refinação: período cronológico (2020-2024), acesso aberto e localização (Portugal, Brasil e Estados Unidos da América).

Ao explorarmos este tema observamos a inevitabilidade da aplicação de ferramentas que impulsionem novas estratégias de desenvolvimento da indexação. A IA é uma mudança natural que o avanço tecnológico possibilitou para as bibliotecas e os bibliotecários. Definitivamente, estamos perante uma nova realidade que permitirá o aprimoramento da indexação.

2 METODOLOGIA

A investigação seguiu uma abordagem qualitativa, fundamentada na pesquisa bibliográfica, por se considerar este método adequado para analisar criticamente a produção científica existente e compreender como a inteligência artificial tem sido aplicada na indexação bibliográfica. O desenho metodológico foi estruturado em **etapas sequenciais**, de modo a garantir transparência, reprodutibilidade e coerência com os objetivos definidos.

2.1 Planeamento e definição da questão de investigação

Partiu-se da questão orientadora, *como a inteligência artificial pode favorecer os bibliotecários na indexação?* que conduziu à formulação dos objetivos específicos: (i) mapear a produção científica sobre o tema, (ii) identificar abordagens e ferramentas de IA aplicadas à indexação e (iii) discutir as implicações para a prática profissional do bibliotecário.

2.2 Seleção das fontes de informação

Foram consideradas duas bases de dados de reconhecida relevância: **WoS** e **EBSCO**. A escolha deveu-se à sua cobertura internacional, indexação criteriosa e pertinência temática para os objetivos do estudo.

2.3 Estratégias de pesquisa e critérios de inclusão

A pesquisa foi conduzida em duas fases:

- **Primeira fase:** utilização da expressão $TS = ("artificial intelligence" AND "automatic indexing")$, restrita à WoS, sem refinamentos adicionais, abrangendo o período de 2004 a 2023, o que resultou em 12 artigos;
- **Segunda fase:** ampliação da expressão para $TS = ("artificial intelligence" AND "indexing" OR "manual indexing" OR "automatic indexing" AND "librarians" AND "libraries")$, com aplicação de refinamentos: período 2020–2024, acesso aberto e localizações (Portugal, Brasil e EUA). Obtiveram-se 294 artigos, reduzidos a 27 (17 WoS; 10 EBSCO) após análise de título, resumo e palavras-chave.

2.4 Organização e análise dos dados

Os artigos selecionados foram sistematizados de acordo com: ano de publicação, país de origem, base de indexação, conceitos-chave e principais contribuições. Posteriormente, procedeu-se à análise de conteúdo, identificando padrões, tendências e categorias temáticas relacionadas com a aplicação da IA na indexação.

2.5 Síntese crítica

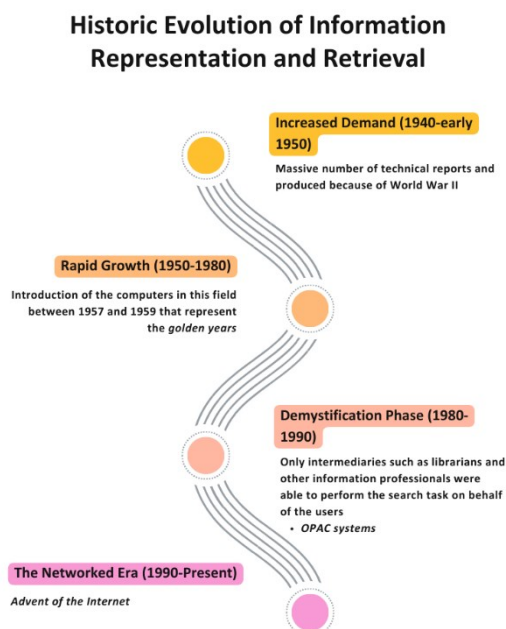
A última etapa consistiu na confrontação dos resultados obtidos com a literatura de referência, destacando convergências, lacunas e perspectivas futuras. Esta síntese permitiu não apenas descrever o estado da arte, mas também discutir as implicações práticas para bibliotecários e bibliotecas, indo ao encontro dos objetivos do estudo.

3 EVOLUÇÃO HISTÓRICA NA INDEXAÇÃO: DA INDEXAÇÃO MANUAL À INDEXAÇÃO AUTOMÁTICA

A indexação é um tema que já tem bastantes anos onde a representação e a recuperação da informação começaram a ter destaque desde o século XIX com a II Guerra Mundial sendo o ponto de partida para uma

evolução da indexação até aos dias de hoje com a indexação automática (Figura 1).

Figura 1– *Historic Evolution of the Information Representation and Retrieval*

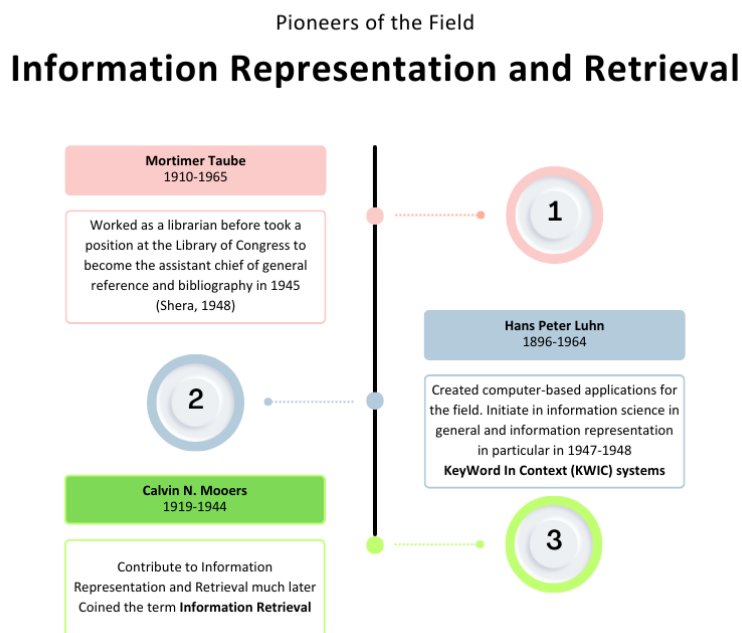


Fonte: Com base em Chu (2010).

Estamos perante três fases que caracterizam a importância que foi sendo atribuída à indexação onde a sua prática já vem sendo feita há largos anos (Vallez *et al.*, 2015) significando que fomos tendo à medida dos seus avanços, uma transição que explica a necessidade de hoje falarmos em inteligência artificial: **Indexação Manual (IM) – Indexação Automática (IA)**. Neste sentido, falamos na necessidade de vermos implementados sistemas de organização de conhecimento nas bibliotecas com o objetivo de assegurarem que um conceito é representado a partir dos seus termos de indexação (Golub, 2016).

Assim sendo, e segundo Gollub (2016), *what happens when the end user does not know which search terms to use at all?* A procura de respostas a esta questão parte sobretudo dos pioneiros da indexação (Figura 2) que à sua forma foram dando as suas perspectivas de acordo com as suas épocas tendo a noção de que, também, os recursos existentes são incomparáveis aos que existem hoje nas bibliotecas onde estas ao longo do tempo foram sofrendo com a evolução natural da representação e recuperação da informação.

Figura 2 – Pionners of the Field of Information Representation and Retrieval



Fonte: Com base em Chu (2010).

Podemos dizer que os vanguardistas apresentados na Figura 2 foram os principais responsáveis pelo desenvolvimento da representação e recuperação da informação, ou seja, a indexação, permitindo que hoje tenhamos a oportunidade de discutir o papel deste processo nas bibliotecas e a forma como os bibliotecários contribuem para o seu melhoramento. Para tal, ao longo dos anos, foram desenvolvidos diferentes tipos de indexação, que se tornaram fundamentais para a organização da informação e sua recuperação eficiente, especialmente considerando a evolução das tecnologias e as necessidades dos utilizadores.

Entre os principais tipos de indexação, destacam-se:

1. **Indexação Derivada ou Livre (Keyword Indexing):** Este tipo de indexação consiste na extração direta de palavras-chave do texto original do documento, sem recorrer a um vocabulário controlado. A seleção dos termos ocorre de maneira espontânea, refletindo o conteúdo do material, mas pode gerar inconsistências devido à falta de padronização terminológica (Chu, 2010; Vázquez *et al.*, 2015).

2. **Indexação por Atribuição (*Assignment Indexing*):** Diferente da indexação derivada, este método utiliza termos de um vocabulário controlado, conhecidos como descritores. Isso garante maior consistência na indexação e na recuperação da informação. No entanto, a utilização de vocabulários controlados tem sido debatida ao longo dos anos, pois pode limitar a flexibilidade da indexação (Golub, 2016; Park; Brenza, 2015)
3. **Indexação Automática ou por Máquina (*Automated/Automatic Indexing*):** Com o avanço tecnológico, surgiram sistemas automatizados que auxiliam na indexação, aliviando o trabalho repetitivo dos indexadores humanos. No entanto, embora os computadores realizem a parte mecânica do processo, a interpretação e atribuição intelectual dos termos ainda são maioritariamente responsabilidade dos bibliotecários (Chen; Bullard; Giustani, 2023; Lloréns *et al.*, 2004; Toepfer; Seigert, 2017).

Essa evolução da indexação, desde os métodos baseados em palavras-chave e vocabulários controlados até a aplicação da IA, evidencia a necessidade de adaptação contínua das bibliotecas e dos profissionais da informação. No contexto deste artigo, buscamos destacar a importância da adoção da IA como um recurso complementar aos processos tradicionais, sem desconsiderar as contribuições históricas desses métodos. Por isso, no Quadro 1, apresentamos uma análise comparativa das vantagens e desvantagens dessas abordagens e defendemos a ideia de uma relação colaborativa entre a inteligência artificial e os bibliotecários para otimizar a indexação e a recuperação da informação.

Quadro 1 –Vantagens e Desvantagens da IM e a IA

IM		IA	
Vállez <i>et al.</i> (2015); Park e Brenza (2015); Golub (2016)	A tarefa de obter resultados pode-se tornar dificultada, lenta e dispendiosa	Park e Brenza (2015)	Poupança de custos
Golub (2016)	Os documentos classificados manualmente, por vezes, não estão disponíveis nas	Toepfer e Seifert (2017)	Tem vários desafios: restrições legais, enviesamento de dados

IM		IA	
	diferentes áreas temáticas		terminologia dos documentos pode apresentar diferenças relativamente à terminologia corrente
		Toepler e Seifert (2017)	As bibliotecas digitais necessitam de métodos automáticos de indexação de documentos

Fonte: Elaboração própria (2025).

Observamos, desta forma, que a IA, disponível há já bastantes anos (Vállez *et al.*, 2015), apresenta-nos naturalmente desafios, mas também benefícios que a inteligência artificial veio fomentar e providenciar às bibliotecas e aos bibliotecários novas ferramentas de IA capazes de salvaguardarem uma representação e recuperação da informação próxima dos utilizadores.

Assim, vemos a importância que a evolução da indexação teve nas bibliotecas onde desde a IM até à indexação automática existiram transformações que permitiram hoje discutirmos o papel da inteligência artificial na representação e recuperação da informação sendo que embora ambas sejam relevantes para a execução desta tarefa, conseguimos verificar que estamos perante a presença de uma mudança de paradigma.

3.1 Inteligência artificial e organização do conhecimento

Nos últimos anos, a integração da IA na organização do conhecimento tem assumido um papel central no debate académico e profissional. A utilização de modelos avançados de processamento de linguagem natural, como *Transformers* e redes neurais profundas, ampliou as possibilidades de automatização da indexação e da recuperação da informação, aproximando os sistemas da linguagem natural dos utilizadores. Estudos recentes (Appleton, 2024; Chen; Bullard; Giustani, 2023; Kasprzik, 2023; Steiger, 2024; Suominen; Inkinen; Lehtinen, 2022) demonstram que ferramentas baseadas em IA são capazes de gerar descritores mais consistentes e, em alguns contextos, com precisão comparável à da IM.

Um dos contributos mais significativos da IA neste domínio é a sua capacidade de apoiar a gestão e atualização de vocabulários controlados,

taxonomias e ontologias, elementos centrais para a consistência da organização do conhecimento. A automatização deste processo contribui para reduzir a defasagem entre terminologia técnica e linguagem natural, mas levanta, ao mesmo tempo, o desafio da adaptação contínua dos modelos aos novos conceitos que emergem em cada área científica (Chandrashekara; Mulimani, 2024).

Adicionalmente, tecnologias recentes como o **Annif**, o **FintoAI** e o **Kratt** têm sido aplicadas em bibliotecas acadêmicas, evidenciando que a indexação apoiada por IA já não se limita ao plano teórico, mas integra práticas concretas. Estes sistemas operam em regime híbrido, combinando algoritmos de aprendizagem automática com mecanismos de validação humana, reforçando a complementaridade entre métodos tradicionais e modernos (Nirudi; Parichi, 2024).

Todavia, a literatura aponta também para limitações que não podem ser negligenciadas. Entre elas destacam-se os riscos de enviesamento algorítmico, a opacidade dos modelos baseados em *deep learning* e a dependência da literacia tecnológica dos bibliotecários para garantir uma utilização crítica das ferramentas (Ferreira; Corrêa, 2023; Mannheimer, 2024). Estes desafios reforçam a importância de pensar a IA não como substituto, mas como catalisador de inovação, exigindo investimentos tanto em tecnologia como em formação profissional.

Em síntese, a aplicação da IA na organização do conhecimento evidencia uma mudança de paradigma: da IM e semi-automática para sistemas híbridos, nos quais a inteligência humana e a inteligência artificial se complementam. Esta evolução aponta para um futuro em que a eficácia e a escalabilidade da IA se unem à qualidade e ao rigor da mediação bibliotecária, estabelecendo novas bases para a gestão da informação em ambientes complexos e dinâmicos.

3.2 Modelos de linguagem generativos e novos cenários para a organização do conhecimento

Nos últimos anos, os grandes Modelos de Linguagem (LLM), como o GPT-4 e outros baseados em arquiteturas *Transformers*, têm assumido crescente relevância no debate académico sobre organização do conhecimento. Estas

ferramentas destacam-se pela capacidade de compreender e gerar linguagem natural em escalas antes inimagináveis, oferecendo novas possibilidades para a indexação, a classificação e a recuperação da informação.

Estudos recentes (Appleton, 2024; Steiger, 2024; Trindade, Oliveira, 2024) salientam que a utilização de IA generativa pode aproximar a indexação da linguagem dos utilizadores, reduzindo a distância entre terminologia técnica e vocabulário natural. Este potencial contribui para tornar os sistemas mais intuitivos e inclusivos, mas levanta também preocupações. A literatura aponta riscos como o enviesamento algorítmico, a opacidade dos processos de geração textual e a dependência da validação humana para assegurar qualidade e consistência (Mannheimer, 2024).

A integração de LLM em contextos bibliotecários tem evidenciado potencial em três dimensões principais:

- **Atribuição automática de descritores** a documentos digitais em grande escala;
- **Apoio à atualização dinâmica de vocabulários controlados**, adaptando-os a novos conceitos científicos, técnicos e sociais;
- **Assistência na mediação com utilizadores**, através de interfaces conversacionais que interpretam consultas complexas e devolvem resultados ajustados ao contexto informacional.

Apesar dos avanços, o debate académico sublinha que estas aplicações exigem uma abordagem crítica e responsável. A adoção de LLM deve ser acompanhada por práticas de transparência, auditoria de resultados e programas de formação contínua para bibliotecários, de modo a assegurar que a inovação tecnológica funcione como complemento e não como substituto da mediação profissional.

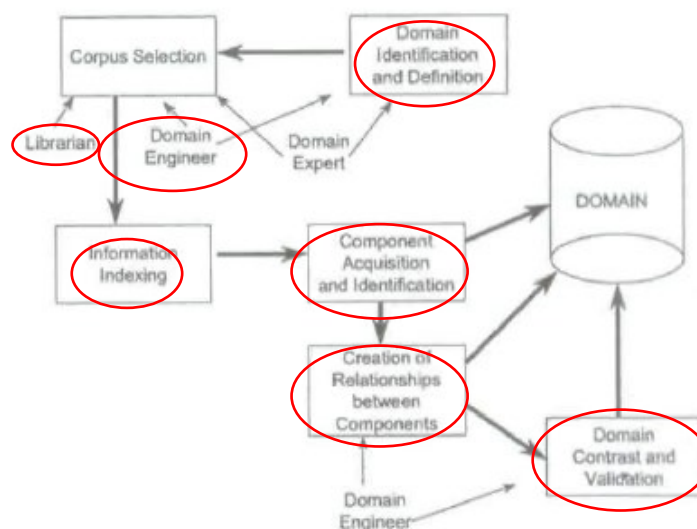
4 INTELIGÊNCIA ARTIFICIAL E INDEXAÇÃO: UMA RELAÇÃO CONTEMPORÂNEA DE APOIO AOS BIBLIOTECÁRIOS

De acordo com o ponto anterior, falamos numa evolução da indexação que hoje ganha maior destaque com a inteligência artificial onde apesar de serem escassas as evidências de sucesso da utilização de ferramentas de IA

(Golub, 2016), podemos afirmar que para ser garantido o controlo de autoridade de elevados e complexos modelos de coleções são necessárias técnicas de análise automática (Pittke *et al.*, 2015). Falamos, então, da indexação automática com o auxílio da inteligência artificial onde procuramos salientar a importância da aplicação deste processo tornando-o mais rápido e potencializando-o no âmbito da atribuição de assuntos (Ferreira; Correa, 2023) que possibilitará um maior nível de eficiência e uma maior quantidade de descritores capazes de ilustrar o conteúdo dos artigos.

Perguntamo-nos, portanto, como a IA pode apoiar os bibliotecários solucionando as suas dificuldades e respondendo aos desafios que a IA nos trouxe onde atualmente existem já ferramentas que possibilitam o desenvolvimento de um novo tipo de indexação apta para uma extração de assuntos mais eficiente sendo que para tal serão integrados diferentes tipos de técnicas (Figura 3).

Figura 3 – Integração de técnicas de desenvolvimento de indexação automática



Fonte: Com base em Lloréns *et al.* (2004, p. 847)

Reforçamos, assim, a necessidade de se verem implementadas nas bibliotecas programas de formação para os bibliotecários para se adaptarem aos novos recursos que estão a aparecer, redefinindo o processo de indexação que será extremamente relevante não só para os bibliotecários como também para a própria indexação. Certamente, a ajuda tecnológica terá um contributo diferente para a revisão e escolha de assuntos (Obaseki, 2010). Devido ao meio

envolvente em que estamos inseridos essa ajuda tecnológica adquire um novo papel de conduzir para o desenvolvimento de novas práticas que influenciem o trabalho dos bibliotecários que apesar de questionarem a qualidade da IA percebem que *"the bottom line is clear: automated indexing works! And it appears to work just as well as human indexing, just differently"* (Obaseki, 2010, p. 3).

Podemos, portanto, afirmar que a IA está pronta para assumir progressivamente um papel de grande contributo para os bibliotecários que com os recursos disponíveis têm a possibilidade de desenvolver um trabalho ainda mais destinado à satisfação das necessidades dos utilizadores alcançado, deste modo, uma aproximação maior à linguagem natural através de um controlo de qualidade rigoroso.

Deste modo, hoje podemos contrariar e discutir a posição de Tulic (2005) citado em Obaseki (2010, p. 2) que defende:

[...] to date, no one has found a way to provide computer programs with the judgement, expertise, or audience awareness that is needed to create useable indexes. Until they do automated indexing will remain a pipe dream.

Entendemos, assim, que conforme foram aparecendo ferramentas de IA foram sido relativizados determinados ceticismos acerca dos benefícios da utilização da IA no ambiente profissional (King *et al.*, 2018) onde todo este crescimento se deveu ao impacto da IA que permitiu a criação de ferramentas de IA como instrumentos de apoio aos bibliotecários no momento da indexação tais como o *Annif*¹, o *FintoAI*², o *Kratt*³, o *SISA*⁴, o *KEA*⁵ e o *MAUI*⁶. Neste sentido, estamos a lidar com novos métodos de PLN que estão a desafiar a forma como devemos encarar a indexação onde as bibliotecas têm o papel de seguir e

¹ Desenvolvido em 2017 (Suominen, 2022; Suominen; Inkinen; Lehtinen, 2022, p. 266), o *Annif* é uma ferramenta *open source* (Suominen, 2022; Suominen; Inkinen; Lehtinen, 2022 citado em Kasprzik, 2023, p. 4) que suporta e permite a utilização de várias línguas (Suominen, 2022) – annif.org

² Criado em 2020, é uma ferramenta baseada no *Annif* com uma diferença relativamente a esta: apenas aceita como línguas, o finlandês, sueco e inglês (Suominen, 2022) – ai.finto.fi

³ Iniciado em 2019 e criado pela Biblioteca Nacional de Estónia, demorou 6-7 meses para o seu completo desenvolvimento (Asula *et al.*, 2021).

⁴ A conceção do *SISA* teve início em meados do ano 1990 que tem como objetivo indexar automaticamente artigos científicos, legislação e sentenças judiciais permitindo apenas a utilização do espanhol, português e inglês (Gil-Leiva *et al.*, 2022) – <http://fcd.inf.um.es:8080/portal>

⁵ O *KEA* é um projeto desenvolvido por dois grupos de investigação da Universidade de Waikato na Nova Zelândia (*Digital Library e Machine Learning*) sendo o sistema simples, robusto e disponível em *open source* (Gil-Leiva *et al.*, 2022) – <http://community.nzdl.org/kea/>

⁶ O *MAUI* faz parte de um projeto de doutoramento concebido por Olena Medelyan em 2009 (Gil-Leiva *et al.*, 2022).

acompanhar a evolução destas mesmas ferramentas e começar a implementar na sua estrutura elementos que possibilitem a adoção de técnicas inovadoras e diferenciadoras que mantenham e fomentem a complementaridade entre a IM e a IA, pois estamos a labutar para uma mudança das práticas humanas para as automáticas (Obaseki, 2020) e necessitamos das melhores condições e apoios para podermos crescer com o impacto que a automatização veio trazer com o aparecimento da inteligência artificial com a certeza de que, tal como refere Kasprzik (2023, p. 8): "[...] one of the most essential components of quality assurance is and will remain the human element"

Defendemos, portanto, que as bibliotecas e, principalmente, os bibliotecários têm um papel exigente na resposta aos desafios que a IA e a indexação automática vieram potencializar na procura por novas abordagens de indexação capazes de chegarem aos utilizadores da mesma forma como acontece com a IM (Quadro 2). Assim, o que queremos realçar é a importância de se verem implementadas ferramentas de IA de um modo progressivo nas bibliotecas que providenciem novas técnicas e consigam dar resposta à evolução das tecnologias emergentes.

Quadro 2 - Recursos e Competências dos bibliotecários para a aplicação da indexação automática

Recursos		Competências	
Vocabulário e classificação bem definida	Dados aceitáveis para o desenvolvimento de uma formação de qualidade	Conhecimento do vocabulário e como utilizá-lo	Estar familiarizado com os processos de indexação
Profissionais com as necessárias competências	Recursos tecnológicos	Ser capaz de lidar com um grande conjunto de dados	Ter conhecimento das ferramentas de indexação automática
		Abordar o assunto com outros profissionais que trabalham diretamente com a automatização	

Fonte: Elaboração própria (2025).

Argumentamos aqui, desta forma, a noção dos bibliotecários como agentes ativos de promoção e disseminação deste tipo de ferramentas através de inquéritos de satisfação, testes de utilização, escrita de notas durante o trabalho, sugestões de novas funcionalidades e por fim, da avaliação de resultados (Suominen, 2022a) ficando, assim, com a ideia de que a interseção

entre a inteligência artificial e a indexação representa não apenas uma evolução técnica, mas também uma transformação fundamental na prática bibliotecária, capacitando profissionais a explorar novas fronteiras de acesso e gestão da informação com eficácia e inovação.

5 RESULTADOS E DISCUSSÃO

A partir da revisão bibliográfica realizada, foi possível identificar cinco eixos centrais sobre a relação entre IA e indexação no contexto bibliotecário. Para além de uma enumeração, estes resultados exigem discussão crítica, na medida em que refletem tanto avanços já consolidados como limitações que persistem na prática profissional.

5.1 A IA não substitui os bibliotecários

Os estudos convergem em afirmar que a IA não elimina o papel do bibliotecário, mas antes reforça a sua centralidade. A indexação não se reduz a processos mecânicos; envolve interpretação semântica, análise contextual e compreensão das necessidades informacionais. Embora ferramentas automáticas consigam acelerar a atribuição de descritores, carecem de sensibilidade para lidar com ambiguidades linguísticas ou conceitos emergentes. Nesse sentido, a IA deve ser entendida como apoio, libertando os profissionais de tarefas repetitivas e permitindo-lhes investir em atividades de maior valor cognitivo, como a curadoria da informação ou a mediação com utilizadores especializados;

5.2 Evolução positiva da indexação automática

Historicamente marcada por limitações técnicas, a indexação automática alcançou níveis mais sofisticados de precisão e consistência. Ferramentas de *machine learning* e modelos de linguagem natural permitiram ultrapassar parte das falhas de abordagens anteriores, reduzindo problemas de sinonímia e polissemia. Esta evolução traduz-se em maior confiança na recuperação da informação e confirma a transição de um campo experimental para uma prática em consolidação;

5.3 Ferramentas aplicadas em bibliotecas

A adoção de soluções concretas demonstra a maturidade tecnológica do campo. O **Annif**, desenvolvido pela Biblioteca Nacional da Finlândia, tem sido usado para atribuição automática de descritores a grandes coleções digitais, combinando algoritmos de aprendizagem automática com vocabulários controlados nacionais. O **Kratt**, na Estónia, integra serviços de governo eletrónico e bibliotecas públicas, oferecendo indexação e recuperação de documentos em múltiplas línguas. Já o **FintoAI** atua como extensão do serviço *FINTO*, permitindo a integração de vocabulários e ontologias em tarefas de indexação semiautomática. Estes exemplos confirmam que a IA já é aplicada em contextos reais, embora a sua eficácia dependa de fatores institucionais como infraestrutura, políticas de dados abertos e programas de capacitação;

5.4 Modernização dos processos bibliotecários

A integração de IA altera os fluxos de trabalho nas bibliotecas, promovendo um modelo híbrido que combina tarefas humanas e automáticas. Esta transformação exige o desenvolvimento de novas competências digitais e uma reorganização interna que valorize a colaboração entre bibliotecários e sistemas inteligentes. Mais do que substituir práticas manuais, a IA desafia os profissionais a redefinirem o seu papel num ecossistema de informação cada vez mais dinâmico e interdisciplinar;

5.5 Necessidade de formação e adaptação profissional

O uso crítico da IA requer literacia tecnológica sólida. Sem formação adequada, as ferramentas podem ser utilizadas de forma acrítica ou rejeitadas por percepção de ameaça. A literatura destaca a importância de programas de capacitação que combinem competências técnicas (domínio de ferramentas, avaliação de resultados) e competências éticas (compreensão de vieses algorítmicos, avaliação da qualidade e transparência dos sistemas). Essa articulação garante que a inovação tecnológica fortaleça a autonomia profissional em vez de a fragilizar.

5.6 Síntese crítica

Os resultados apontam para um cenário em que a IA redefine a indexação bibliográfica, mas não elimina a mediação humana. A análise dos casos práticos (*Annif*, *Kratt*, *FintoAI*) mostra que os métodos híbridos, em que algoritmos e bibliotecários partilham responsabilidades, apresentam maior potencial de eficácia. Persistem, contudo, desafios ligados à atualização contínua dos vocabulários controlados, à opacidade de alguns modelos de *deep learning* e à necessidade de formação crítica dos profissionais da informação.

Figura 4 – BibliotecarIA



Fonte: BibliotecarIA ([2025]).

A leitura cruzada da literatura evidencia que estes cinco tópicos não podem ser entendidos de forma isolada. Eles expressam, antes, a coexistência de diferentes modelos de indexação que se complementam e competem em função de contextos, recursos e objetivos institucionais. Para clarificar estas diferenças, apresenta-se de seguida um quadro comparativo que sintetiza as características, vantagens e limitações dos métodos manual, semi-automático e automático baseado em IA.

Quadro 3 – Comparação entre métodos de indexação

Método	Características principais	Vantagens	Limitações
Indexação manual	Realizada exclusivamente por bibliotecários, com base em vocabulários controlados e interpretação contextual.	- Elevada qualidade semântica; -Consistência conceitual; -Adaptação às necessidades do utilizador.	- Processo moroso e dispendioso; -Vulnerável a subjetividade individual; -Dificuldade em lidar com grandes volumes de dados.
Indexação semi-automática	Combina algoritmos de extração de termos com validação humana.	- Reduz o tempo de processamento;	- Requer interoperabilidade entre sistemas;

Método	Características principais	Vantagens	Limitações
		-Mantém a intervenção crítica do bibliotecário; -Maior escalabilidade que a manual.	-Qualidade depende da validação humana; -Nem sempre integra vocabulários atualizados.
Indexação automática baseada em IA	Utiliza redes neurais, <i>Transformers</i> e técnicas de <i>deep learning</i> para extrair e atribuir descritores de forma autônoma.	- Elevada eficiência e rapidez; -Escalabilidade em grandes bases de dados; -Capacidade de se aproximar da linguagem natural.	- Risco de enviesamento algorítmico; -Opacidade dos modelos (<i>black box</i>); -Dependência da literacia tecnológica dos profissionais; -Necessidade de constante atualização dos vocabulários.

Fonte: Elaboração própria com base em Chu (2010), Golub (2016), Suominen, Inkinen e Lehtinen, (2022), Chen, Bullard e Giustani (2023) e Kasprzik (2023).

O quadro confirma que nenhum método responde sozinho às exigências contemporâneas da organização do conhecimento. A IM continua a garantir qualidade interpretativa e rigor conceitual, mas carece de escalabilidade. A semi-automática mostra-se como alternativa equilibrada, ao conjugar eficiência com validação humana. Já a automática baseada em IA amplia a capacidade de processamento, mas levanta questões de transparência, viés e atualização contínua de vocabulários. A análise sugere, portanto, que o caminho mais promissor reside em modelos híbridos, nos quais a tecnologia potencia a eficiência e os bibliotecários asseguram o rigor e a mediação crítica.

6 CONCLUSÃO

A utilização da IA na indexação representa uma oportunidade estratégica para reforçar a eficiência e a qualidade dos serviços bibliotecários. Os resultados obtidos confirmam que a IA atua como tecnologia de apoio, ampliando a capacidade de análise dos profissionais e contribuindo para maior precisão e consistência na recuperação da informação.

Contudo, permanecem desafios que exigem investigação e desenvolvimento. Em primeiro lugar, a atualização contínua de vocabulários controlados coloca a necessidade de explorar modelos híbridos que articulem processos automáticos com validação profissional. Em segundo, os avanços recentes em LLM abrem novos cenários, mas ainda carecem de avaliação sistemática quanto à sua eficácia em contextos reais de indexação. Por fim, a

adoção responsável destas ferramentas depende de políticas claras de formação e capacitação, que preparem os bibliotecários para lidar com riscos de enviesamento algorítmico e falta de transparência nos sistemas.

Assim, para além de evidenciar os benefícios atuais da IA, este estudo indica três direções prioritárias para futuras investigações: (i) a análise comparativa entre métodos tradicionais, automáticos e baseados em LLM; (ii) o desenvolvimento de estratégias de integração entre IA e vocabulários controlados; e (iii) a avaliação do impacto da formação profissional na utilização crítica e responsável destas tecnologias.

REFERÊNCIAS

APPLETON, L. AI and academic libraries: what's all the fuss about? **New Review of Academic Librarianship**, [s.l.], v. 30, n. 3-4, p. 281-295, 2024. DOI <https://doi.org/10.1080/13614533.2024.2356474>.

ASULA, M. *et al.* Kratt: developing an automatic subject indexing tool for the National Library of Estonia. **Cataloguing e Classification Quarterly**, [s.l.], v. 59, n. 8, p. 775-793, 2021. DOI <https://doi.org/10.1080/01639374.2021.1998283>.

BIBLIOTECARIA. **Avanzando en la intersección de la tecnología y el conocimiento**. [s.l.]: Bibliotecar1A, ([2025]). Disponível em: <https://www.bibliotecaria.es/>. Acesso em: 31 out. 2025.

CHANDRASHEKARA, G. S.; MULIMANI, M. The impact of artificial intelligence on library and information science (LIS) services. **International Journal of Innovative Practice and Applied Research (IJIPAR)**, [s.l.], v. 14, n. 5, p. 50-56, 2024. DOI <http://dx.doi.org/10.2139/ssrn.4856459v>.

CHEN, E.; BULLARD, J.; GIUSTANI, D. Automated indexing using NLM's Medical Text Indexer (MTI) compared to human indexing in Medline: a pilot study. **Journal of the Medical Library Association**, Chicago, v. 111, n. 3, p. 684-694, 2023. DOI <https://dx.doi.org/10.5195/jmla.2023.1588>.

CHU, H. **Information Representation and Retrieval in the Digital Age**. 2. ed. Medford: American Society for Information Science and Technology : Information Today, 2010.

GIL-LEIVA, I. *et al.* Extracción de información de documentos PDF para su uso en la indización automática de e-books. **Transinformação**, Campinas, v. 34, [s.n.], p.1-11, 2022. DOI <https://doi.org/10.1590/2318-0889202234e210069>.

GOLUB, K. Potential and challenges of subject access in libraries today on the example of swedish libraries. **International Information e Library Review**, [s.l.], v. 48, n. 3, p. 204-210, 2016. DOI <https://doi.org/10.1080/10572317.2016.1205406>.

FERREIRA, M. H. W.; CORREA, R. F. Sistematização da obtenção de indicadores temáticos de informação científica. **Encontros Bibli**, Florianópolis, v. 28, [s.l.], p. 1-30, 2023. DOI <https://doi.org/10.5007/1518-2924.2023.e92070>.

KASPRZIK, A. Automatic subject indexing at ZBW: making research results stick in practice. **Journal of the Association of European Research Libraries**, [s.l.], v. 33, n. 1, p. 1-17, 2023. DOI <https://doi.org/10.53377/lq.13579>.

KING, S. *et al.* **Revisiting indexing and abstracting in the digital era**. Texas: University of North Texas, 2018. Disponível em: https://digital.library.unt.edu/ark:/67531/metadc1164546/m2/1/high_res_d/Revisiting_Indexing_and_Abstracting_in_the_Digital_Era.pdf. Acesso em: 31 out. 2025.

LLORENS, J. *et al.* Automatic generation of domain representations using thesaurus structures. **Journal of the American Society for Information Science and Technology**, [s.l.], v. 55, n. 10, p. 846-858, 2004. DOI <https://doi.org/10.1002/asi.20039>.

MANNHEIMER, S. Responsible AI practice in libraries and archives. **Information Technology and Libraries**, Ann Arbor, v. 43, n. 2, p. 1-20, 2024. Disponível em: <https://ital.corejournals.org/index.php/ital/article/view/17245>. Acesso em: 31 out. 2025.

NIRUDI, Y.; PARICHI, R. Artificial intelligence in libraries: an overview. **SSRN Electronic Journal**, [s.l.], [s.n.], [s.n.], nov. 2024. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5080670. Acesso em: 31 out. 2025.

OBASEKI, T. I. Automated indexing: the key to information retrieval in the 21st century. **Library Philosophy and Practice**, Nebraska, v. 338, [s.n.], p. 1-4, 2010. Disponível em: <https://abrir.link/CMYCh>. Acesso em: 31 out. 2025.

OLIVEIRA, R. **O uso da inteligência artificial em bibliotecas universitárias: aplicações em catalogação e indexação**. Porto Alegre: UFRGS, 2024. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio Grande do Sul, Porto Alegre. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/290598/001244292.pdf>. Acesso em: 31 out. 2025.

PARK, J.; BRENZA, A. Evaluation of semi-automatic metadata generation tools: a survey of the current state of the art. **Information Technology and Libraries**, Ann Arbor, v. 34, n. 3, p. 22-42, 2015. DOI <https://doi.org/10.6017/ital.v34i3.5889>.

PITKE, F.; LEOPOLD, H.; MENDLING, J. Automatic detection and resolution of lexical ambiguity in process models. **IEEE Transactions on Software Engineering**, [s.l.], v. 41, n. 6, p. 526-544, 2015. DOI <https://doi.org/10.1109/TSE.2015.2396895>.

STEIGER, K. Artificial Intelligence in higher education and academic libraries: a literature review. **Endnotes**, [s.l.], v. 125, n. 1, p. 1-15, 2024. Disponível em: <https://journals.ala.org/index.php/endnotes/article/view/8235>. Acesso em: 31 out. 2025.

SUOMINEN, O. **Supporting subject librarians with AI solutions**. Finland: IFLA, 2022. Disponível em: https://www.ifla.org/wp-content/uploads/1.Suominen_Supporting-Subject-Librarians--IFLA-AI-webinar.pdf. Acesso em: 31 out. 2025.

SUOMINEN, O.; INKINEN, J.; LEHTINEN, M. Annif and Finto AI: developing and implementing automated subject indexing. **JLIS.it**, [s.l.], v. 13, n. 1, p. 265-282, 2022. DOI <https://doi.org/10.4403/jlis.it-12740>.

TOEPFER, M.; SEIFERT, C. Descriptor-invariant fusion architectures for automatic subject indexing. ACM/IEEE Joint Conference on Digital Libraries (JCDL), 1., 2017, Toronto. **Proceedings** [...]. Toronto: ACM : IEEE, 2017. p. 1-10. DOI 10.1109/JCDL.2017.7991557.

TRINDADE, Alessandra Stefane Cândido Elias da; OLIVEIRA, Henry Poncio Cruz de. Inteligência Artificial (IA) Generativa e Competência em Informação: habilidades informacionais necessárias ao uso de ferramentas de ia generativa em demandas informacionais de natureza acadêmica-científica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 29, n. 2, p. 201-219, 2024. DOI <http://dx.doi.org/10.1590/1981-5344/47485>.

VALLEZ, M. *et al.* Updating controlled vocabularies by analysing query logs. **Online Information Review**, [s.l.], v. 39, n. 7, p. 1-24, 2015. DOI <http://dx.doi.org/10.1108/OIR-06-2015-0180>.

CONTRIBUIÇÕES DAS AUTORIAS

Informa-se nesta seção as funções de cada pessoa autora, de acordo com a [taxonomia CRediT](#), conforme orientado na página da revista PCI:

Função	Definição
Conceituação	Mariângela Spotti Lopes Fujita; Nuno Miguel Teixeira Sousa.
Curadoria de dados	—
Análise Formal	Nuno Miguel Teixeira Sousa.
Obtenção de financiamento	—
Investigação	Nuno Miguel Teixeira Sousa.
Metodologia	Nuno Miguel Teixeira Sousa.
Administração do projeto	—
Recursos	—
Software	—
Supervisão	Mariângela Spotti Lopes Fujita.
Validação	—
Visualização [de dados (infográfico, fluxograma, tabela, gráfico)]	Nuno Miguel Teixeira Sousa.
Escrita – primeira redação	Mariângela Spotti Lopes Fujita; Nuno Miguel Teixeira Sousa.
Escrita – revisão e edição	Mariângela Spotti Lopes Fujita.

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS data-available-upon-request – Os dados de pesquisa só estão disponíveis mediante solicitação.

Editora Chefe: Lorena Tavares de Paula 