

## XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - XXV ENANCIB

### GT 2 – Organização e Representação do Conhecimento

#### INDEXAÇÃO AUTOMÁTICA E OS MODELOS GENERATIVOS: UMA TAXONOMIA APLICADA

#### *AUTOMATIC INDEXING AND GENERATIVE MODELS: AN APPLIED TAXONOMY*

**Juliana Cristhine Schifino Santos Jardim Pereira** – Universidade Federal do Rio Grande do Sul (UFRGS)

**Rene Faustino Gabriel Junior** – Universidade Federal do Rio Grande do Sul (UFRGS)

#### Modalidade: Trabalho Completo

**Resumo:** a Organização do Conhecimento é função central de um bibliotecário, envolvendo identificar assuntos e traduzi-los para linguagens controladas. Para tanto, técnicas de indexação automática, em conjugação à indexação manual, vêm sendo estudadas, especialmente com o advento de modelos generativos de Inteligência Artificial, como o ChatGPT. O presente trabalho objetiva desenvolver uma taxonomia sobre Inteligência Artificial aplicada à Organização do Conhecimento. Este estudo é aplicado, qualitativo e metodológico. Partiu-se da identificação de termos nas bases Brapci, Scopus e Web of Science. Os anos de 2022 a 2024 foram escolhidos buscando dessa forma representar o cenário atual. Das 7871 palavras-chave encontradas, restaram 543 termos relacionados à Inteligência Artificial. Destes, 162 eram terminologias aptas a formarem uma taxonomia adequada ao objetivo deste estudo. Conclui-se que o potencial da Inteligência Artificial como agente sobre a Organização do Conhecimento é evidenciado por sua ampla presença na área. Além disso, a formação de uma taxonomia amplia as discussões sobre os vocabulários controlados e sua importância, possibilitando futuros estudos sobre a intersecção entre Organização do Conhecimento e Inteligência Artificial.

**Palavras-chave:** indexação automática; inteligência artificial generativa; taxonomia.

**Abstract:** Knowledge Organization is a central function of a librarian, involving identifying subjects and translating them into controlled languages. To this end, automatic indexing techniques, in conjunction with manual indexing, have been studied, especially with the advent of generative Artificial Intelligence models, such as ChatGPT. This work aims to develop a taxonomy on Artificial Intelligence applied to Knowledge Organization. This study is applied, qualitative and methodological. It started by identifying terms in the Brapci, Scopus and Web of Science databases. The years 2022 to 2024 were chosen in order to represent the current scenario. Of the 7871 keywords found, 543 terms related to Artificial Intelligence remained. Of these, 162 were terminologies capable of forming a taxonomy appropriate to the objective of this study. It is concluded that the potential of Artificial Intelligence as an agent on Knowledge Organization is evidenced by its wide presence in the area. Furthermore, the formation of a taxonomy broadens discussions on controlled vocabularies and their importance, enabling future studies on the intersection between Knowledge Organization and Artificial Intelligence.

**Keywords:** automatic indexing; generative artificial intelligence; taxonomy.

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

## **1 INTRODUÇÃO**

A identificação de assuntos e sua tradução para linguagens controladas são partes centrais do trabalho de organização do conhecimento. Porém, a sociedade e seus hábitos de consumo científico, cultural e informativo mudaram muito em pouco tempo, e com isso, foram precisas adaptações (Abadal; Anglada, 2017; Dodebei, 2014; López Yepes, 2015). Afora os múltiplos formatos físicos ou virtuais, há ainda um universo complexo de assuntos contidos num mesmo objeto informacional, que podem ser, em um mesmo momento, de interesse para diferentes campos do conhecimento (Rendón-Rojas, 2008).

Nos anos 1950, Luhn pioneiramente investigou como a indexação automática poderia auxiliar bibliotecários na organização do conhecimento especializado, hipótese que se fortaleceu na década de 1960 com o surgimento de bases de dados e, posteriormente, da Web, exigindo novas estruturas tecnológicas e necessidades informacionais. Paralelamente, Robredo e Ferreira (1980) desenvolveram linguagens controladas rigorosamente estruturadas para otimizar esse processo. Em *Indexing and Abstracting in Theory and Practice*, Lancaster (1998) sistematizou métodos de indexação manual e os contrapôs criticamente a sistemas automáticos emergentes, evidenciando seus limites e potencialidades. Por fim, Valentin (2016, p. 37) ressalta que compreender as tendências informacionais é “essencial para as bibliotecas contemporâneas”, dado seu impacto na mediação da informação.

Diversos são os estudos sobre a aplicação tecnológica por meio da utilização de sistemas de organização do conhecimento, procurando tornar a recuperação da informação mais efetiva (De Luca; Fallucchi; Morelato, 2021; Markey, 2006). Neste cenário, com a popularização dos modelos generativos, como o ChatGPT no final de 2022, inovações significativas apresentaram a capacidade de interpretar grandes volumes de texto e interagir com usuários por meio de linguagem natural. Mais recentemente, surgiram modelos generativos ainda mais avançados, que começam a se destacar com potenciais aplicações na área de organização do conhecimento.

Diante do avanço dos modelos generativos, surge a discussão sobre sua aplicabilidade à Organização do Conhecimento (OC). Eles seriam realmente capazes de lidar com a complexa tarefa de indexar automaticamente grandes volumes de dados? Frente a essa problemática, o presente estudo tem como objetivo geral propor uma representação estruturada dos conceitos de Inteligência Artificial no contexto da Organização do Conhecimento, com base

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

em análise bibliométrica e terminológica. Os objetivos específicos são: identificar palavras-chave relacionadas à Inteligência Artificial (IA) nas bases de dados Brapci, Scopus e Web of Science (WoS); analisar as palavras-chaves identificadas a fim de gerar uma taxonomia; e desenvolver uma taxonomia sobre IA aplicada à OC.

## **2 DESENVOLVIMENTO**

A OC tem considerado os aparatos tecnológicos como solução para lidar com a organização da massa informacional que permanece em constante crescimento e transformação. Os impactos tecnológicos são evidentes no fazer científico da área de OC. E segundo Abadal e Anglada (2017) é importante lembrar que muitos relatórios prospectivos sobre o futuro das bibliotecas incluíam considerações sobre tecnologia.

### **2.1 Inteligência Artificial**

Mas apesar das críticas por parte da classe intelectual da época, o campo da Inteligência Artificial (IA) ganhou destaque em 1947, com os estudos em aprendizado de máquina, algoritmos genéticos e aprendizado por reforço de Alan Turing (Russel; Norvig, 2024). Com o contínuo avanço tecnológico, aliado ao desenvolvimento da World Wide Web nos anos 1990 (e seu desenvolvimento em Web Semântica<sup>3</sup>), desencadeia-se o surgimento do fenômeno denominado *big data*<sup>4</sup> (Berners-Lee; Lassila; Hendler, 2001). A partir deste, foram desenvolvidos algoritmos de aprendizagem capazes de processar grandes conjuntos de dados disponíveis na web.

A IA generativa reúne técnicas de aprendizado profundo que, a partir de grandes volumes de dados, são capazes de compreender e gerar conteúdo em múltiplos formatos. Essa tecnologia está presente em sistemas inteligentes de perguntas e respostas, oferecendo suporte na resolução de problemas simples e complexos. Exemplos notáveis incluem modelos como GPT-4, Copilot e Ollama (Feuerriegel *et al.*, 2024), os quais, fundamentados em arquiteturas de Processamento de Linguagem Natural (PLN), funcionam como modelos matemáticos que aprendem e interagem com usuários, contribuindo também para o avanço científico no estudo das linguagens e de seus usos.

<sup>3</sup> “A Web Semântica não é uma Web à parte, mas uma extensão da atual” (Berners-Lee; Lassila; Hendler. 2001, tradução nossa).

<sup>4</sup> Grandes conjuntos de dados, nos mais diversos formatos (imagem, texto, som etc.).

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

Entre exemplos de IA generativa, temos as ferramentas *open source* (código aberto), de baixíssimo custo (Zupan; Demsar, 2008). Elas tendem a ter uma comunidade de usuários participativa e colaborativa, que constroem novos modelos de linguagem e ajudam a aprimorar os já existentes (Androutsellis-Theotokis *et al.*, 2011). Ainda, são capazes de executar modelos grandes de linguagem (LLMs<sup>5</sup>) diretamente em um computador *offline*, permitindo controle total de dados e evitando riscos de segurança. Tais ferramentas, também reduzem a dependência de servidores externos, com respostas rápidas e confiáveis.

## **2.2 Organização do Conhecimento e sua relação com a Inteligência Artificial**

Tais predicados, relacionados à IA generativa, vem ao encontro das necessidades da OC. Conforme Hjørland (2016), a OC é um campo de pesquisa, ensino e prática, que trata de atividades passíveis de serem realizadas tanto por humanos quanto por máquinas. Entre estas atividades, sistemas de classificação, tesouros, entre outras linguagens (Guimarães *et al.*, 2015), permeia o *Simple Knowledge Organization System* (SKOS). Segundo a *WORLD WIDE WEB CONSORTIUM* (2012), SKOS é uma área de trabalho que desenvolve meios de apoiar os sistemas de organização do conhecimento (SOCs) dentro da Web Semântica. Nascimento e Pinho (2019) afirmam que os principais tipos de SOCs encontrados na literatura e na prática social da OC podem ser distinguidos em taxonomias (Campos; Gomes, 2007), tesouros e ontologias, sistema de classificação, mapa conceitual, *topic maps* e folksonomia.

A taxonomia em Ciência da Informação é uma estrutura hierárquica de termos voltada à organização de domínios de conhecimento, visando otimizar agregação e recuperação de dados (Mendes; Pinto, 2019). Por meio de relações “pai-filho” (generalização > especialização) e integração com linguagens documentárias, ela sustenta sistemas de classificação, tesouros, modelos conceituais e ontologias (González, 2014), tornando a navegação em repositórios mais intuitiva. Hjørland (2008) enfatiza que a representação e organização dos objetos informacionais devem refletir os objetivos e valores dos usuários, garantindo SOCs adequados ao contexto de uso.

A elaboração de taxonomias interdisciplinares exige uma base teórica robusta, incorporando conceitos da Classificação Facetada e dos Níveis Integrativos (Novo, 2010), o que confere maior flexibilidade à modelagem conceitual. Além disso, mapas conceituais

---

<sup>5</sup> Do inglês, *Large Language Models*.

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

funcionam como recursos visuais eficazes para desenhar interfaces e estruturar hierarquias de navegação, esclarecendo as relações semânticas entre os termos (Gabriel Junior; Laipelt, 2017).

Ingetraut Dahlberg integrou a Teoria do Conceito à Teoria da Classificação, estabelecendo fundamentos sólidos para representação e recuperação da informação no âmbito das linguagens documentárias (Dodebei, 2014). Essas linguagens são vocabulários controlados, derivados de linguagens naturais e chamados de linguagens de indexação ou descritoras, cujo propósito é aprimorar o fluxo de informação entre indexadores e usuários (Nocetti; Figueiredo, 1978). No século XXI, esses vocabulários evoluíram de meros critérios linguísticos para princípios lógicos.

Desde a década de 1970, diversos modelos têm buscado automatizar a classificação e a atribuição de descritores a documentos, motivados pelo aumento do volume informacional e antecipados por Luhn (1958a, 1958b), que já imaginava máquinas capazes de substituir parte do esforço humano na indexação. Robredo (1982) e Narukawa e Sales (2021) apontam a indexação automática como alternativa viável para análise e representação do conhecimento, ao passo que Richmond (1974) criticou esquemas enumerativos, defendendo sua adaptação aos contextos digitais. F. W. Lancaster (2004), sistematizou métodos manuais e comparou-os a soluções automáticas emergentes, destacando critérios de especificidade, consistência e exaustividade que continuam a orientar a construção de vocabulários controlados, tesouros e bases de dados bibliográficas.

A indexação automática, conforme Lancaster (2004) e Hjørland (2008), emprega algoritmos que selecionam termos sem intervenção humana direta, mas o alcance dessas aplicações é condicionado ao estágio tecnológico vigente. De Luca, Fallucchi e Morelato (2021) destacam que, apesar dos avanços computacionais que viabilizam tarefas antes impraticáveis, as máquinas carecem de capacidade de compreensão textual e abstrativa; sem instruções precisas, não produzem resultados, pois sua saída consiste unicamente em metadados gerados por operações codificadas sobre os materiais catalogados.

A eficácia da indexação automática depende diretamente da qualidade dos dados de treinamento, erros, *outliers* e ruídos comprometem o desempenho dos sistemas de Machine Learning (Géron, 2021), e da rigorosidade na coleta e anotação desses dados (Gomes, 2022). Ainda assim, conforme Richmond (1974), esse recurso não substitui os métodos tradicionais, mas os complementa, ampliando precisão, profundidade e flexibilidade

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

na organização e recuperação da informação. A partir dos anos 2000, pesquisas passaram a valorizar a qualidade da representação informacional, o emprego de linguagem natural e metodologias adaptadas a contextos digitais; nesse sentido, Alcaide *et al.* (2001) demonstraram, em comparação entre sistemas automáticos e indexadores humanos, que a consistência na recuperação e a equivalência semântica dos descritores seguem sendo desafios centrais.

O desenvolvimento da indexação automática reflete os avanços em PLN e OC, evidenciando um progresso contínuo no balanceamento entre eficiência, precisão e semântica dos sistemas de recuperação, o que é fundamental tanto para o avanço científico quanto para democratizar o acesso à informação.

### **3 METODOLOGIA**

O presente estudo se caracteriza por uma natureza aplicada, pois intenta aprimorar as práticas e resultados obtidos no contexto da organização do conhecimento (Lieu; Platt, 2017), e busca a solução de problemas ou necessidades concretas e imediatas (Appolinário, 2007). Ainda, propõe abordagem qualitativa pelas análises subjetivas que serão realizadas, uma forma de pesquisa onde os dados são coletados pelo pesquisador, por meio da verificação experimental de interações sociais dinâmicas, e analisados subjetivamente (Chizzotti, 2014). Assim, se preocupa com os fenômenos, ou seja, com a interpretação subjetiva dos fatos. Quanto aos fins, a pesquisa se caracteriza como metodológica, pois objetiva desenvolver instrumentos, técnicas e métodos de investigação científica.

Para operacionalizar o desenvolvimento da pesquisa, procedeu-se ao desdobramento dos meios conforme os objetivos específicos e a metodologia de Novo (2010), baseada na Classificação Facetada e dos Níveis Integrativos, que posteriormente foi revista em 2017 por Gabriel Junior e Laipelt (2017). Para o primeiro objetivo, foi realizada busca bibliográfica nas bases Brapci, Scopus e Web of Science, com o propósito de extrair palavras-chave e compor o corpus de análise. Adotou-se estratégia booleana bilíngue, com os termos: (“Inteligência Artificial” OR “Artificial Intelligence”) AND (“Organização do Conhecimento” OR “Knowledge Organization”), restrita a documentos em português e inglês, publicados entre 2022 e 2024, período considerado representativo das investigações mais recentes no campo da Ciência da Informação. Dessa forma, garantiu-se a seleção de registros alinhados ao escopo temporal e linguístico definidos pelo estudo.

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

A análise das palavras-chave encontradas nas bases, ocorreu através dos seguintes passos: a) separação dos termos, com a retirada dos que fossem sobre outros assuntos não relacionados à IA; b) deduplicação, retirando termos de grafia igual ou semelhante com o mesmo significado; e c) separando termos de mesmo sentido, mas em outro idioma, que possuísem equivalência em português. Todas essas etapas foram realizadas na ferramenta Excel.

A seguir, os termos restantes foram em parte traduzidos e em parte reagrupados com suas equivalências já encontradas entre os dados iniciais, passando a ser apresentados concomitantemente nas línguas portuguesa, inglesa e espanhola (ver Quadro 2). Após isso, foram subdivididos em grandes categorias de termos correlacionados, que foram nomeadas conforme os títulos e/ou assuntos de seus respectivos termos subordinados. Desta forma, formou-se uma hierarquização no sentido de estabelecer uma taxonomia sobre IA aplicada à OC.

#### **4 RESULTADOS**

Os objetivos específicos são: identificar palavras-chave relacionadas à Inteligência Artificial (IA) nas bases de dados Brapci, Scopus e Web of Science (WoS); analisar as palavras-chaves identificadas a fim de gerar uma taxonomia; e desenvolver uma taxonomia sobre IA aplicada à OC.

Para identificar palavras-chave relacionadas à Inteligência Artificial (IA), efetuaram-se buscas nas bases BRAPCI (3827 termos), Scopus (2198 termos) e WoS (1846 termos). A constituição do corpus inicial demonstra a abrangência das bases consultadas e sua relevância para capturar a terminologia atual relacionada à IA aplicada à OC. Após a remoção de duplicatas, o conjunto consolidado reduziu-se inicialmente para 543 termos e, em seguida, para 299. A exclusão de variações ortográficas e sinônimos resultou em 162 termos finais, (representando aproximadamente 4,2% do conjunto inicial) que compõem a taxonomia de domínio sugerida (Campos; Gomes, 2007), apresentada no Quadro 1 e detalhada em português, inglês e espanhol no Quadro 2. Tal índice de filtragem (162 termos) ressalta tanto a riqueza vocabular inerente ao tema quanto a necessidade de criteriosa seleção para garantir coerência metodológica

A distribuição dos 162 termos em categorias temáticas (por exemplo, “Algoritmos e Técnicas Relacionadas”, “Aprendizado de Máquina” e “Ética na Inteligência Artificial”) reflete,



**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

semântica entre repositórios. Essa decisão metodológica fortalece o alinhamento com padrões internacionais, como o SKOS, e potencializa o uso da taxonomia em ambientes Web Semântica, onde a multilinguagem é fator crítico para a recuperação eficaz da informação.

Os resultados confirmam o potencial da Inteligência Artificial como agente catalisador na OC, sobretudo por meio de modelos generativos que, conforme evidenciado no corpus, ocupam posicionamento de destaque (por ex., “Inteligência artificial generativa”, “ChatGPT”, “Grandes modelos de linguagem”).

Ao mapear tais termos, a taxonomia não apenas evidencia áreas consolidadas, como também aponta lacunas emergentes, por exemplo, em domínios de “Interação humano–IA” e “Política de Regulação de Inteligência Artificial”, sinalizando oportunidades para pesquisas futuras que aprofundem essas temáticas.

A preparação manual dos termos (limpeza, deduplicação e tradução via Excel) garante controle e confiabilidade, mas impõe limitações em termos de escalabilidade e reprodutibilidade do processo. Para superar tais desafios, recomenda-se o desenvolvimento de pipelines automatizados de extração e normalização de termos, utilizando técnicas de NLP e algoritmos de correspondência lexical. Essa abordagem poderá ampliar o escopo temporal e de bases consultadas, além de reduzir o viés humano inerente à filtragem manual.

**Quadro 1 – Termos da IA e suas variações**

<b>Termos Gerais (General terms; Términos generales)</b>	
	Inteligência artificial (IA) ( <i>Artificial intelligence (AI); Inteligencia artificial (IA)</i> )
	Variações do termo “Inteligência Artificial” ( <i>Variations of the term “Artificial Intelligence”; Variaciones de la terminología</i> )
	“Inteligencia Artificial”
	IA (AI; IA)
	Inteligência artificial centralizada ( <i>Centralized artificial intelligence (CAI); Inteligencia artificial centralizada</i> )
	Inteligência artificial confiável ( <i>Trustworthy artificial intelligence; Inteligencia artificial confiable</i> )
	Inteligência artificial das coisas ( <i>Artificial intelligence of things; Inteligencia artificial de las cosas</i> )
	Inteligência artificial generativa ( <i>Generative artificial intelligence (GenAI); Inteligencia artificial generativa</i> )
	Inteligência artificial pessoal ( <i>Personal artificial intelligence (PAI); Inteligencia artificial personal</i> )
	Inteligência artificial preditiva ( <i>Predictive artificial intelligence; Inteligencia artificial predictiva</i> )
	Inteligência artificial romântica ( <i>Romantic artificial intelligence; Inteligencia artificial romántica</i> )
	a)
	b)
	c)
	d)
	<b>Algoritmos (Algorithms; Algoritmos)</b>
	Algoritmo ( <i>Algorithm; Algoritmo</i> )
	e)
	f)
	Algoritmo de clusterização hierárquica ( <i>Hierarchical clustering algorithm; Algoritmo de agrupamiento jerárquico</i> )
	g)
2.	Clustering
	Clusterização automática ( <i>Automatic clustering; Agrupación automática</i> )
	Clusters ( <i>Clusters, Clústeres</i> )
	<b>Análise em IA (AI analysis; Análisis de IA)</b>
1.	Análise de citação ( <i>Citation analysis; Análisis de citas</i> )

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

2. Análise de conteúdo (*Content analysis; Análisis de contenido*)
3. Análise de dados (*Data analysis; Análisis de datos*)
4. Análise de enunciados (*Statement analysis; Análisis de enunciados*)
5. Análise de grandes dados (*Big data analysis; Análisis de grandes datos*)
6. Análise de mapas (*Maps analysis; Análisis de mapas*)
7. Análise de redes neurais (*Neural network analysis; Análisis de redes neuronales*)
8. Análise de sentimento (*Sentiment Analysis; Análisis de sentimiento*)

**Aprendizado de Máquina (*Machine learning; Aprendizaje automático*)**

1. Aprendizagem de máquina (*Machine learning; Aprendizaje automático*)
  - a) Aprendizado ativo profundo (*Deep active learning; Aprendizaje activo profunda*)
  - b) Aprendizado profundo (*Deep learning; Aprendizaje profundo*)
  - c) Aprendizagem profunda (*Deep learning; Aprendizaje profundo*)
2. Bibliotecas de aprendizado de máquina (*Machine learning libraries; Bibliotecas de aprendizaje automático*)

**Extração de Dados (*Data extraction; Extracción de datos*)**

1. Extração de conhecimento (*Knowledge extraction; Extracción de conocimiento*)
2. Extração semi-automática de termo (*Semi-automatic term extraction; Extracción de términos semiautomática*)
3. Extrativismo de dados (*Data extractivism; Extractivismo de datos*)

**Linguística (*Linguistics; Lingüística*)**

1. Aprendizado de ontologia (*Ontology learning; Aprendizaje de ontología*)
2. Árvore de decisão (*Decision tree; Árbol de decisión*)
3. Linguística computacional (*Computational linguistics; Lingüística computacional*)
4. Lógica difusa (*Fuzzy logic; Lógica difusa*)
5. Ontologia (*Ontology, Ontología*)

**Mineração em IA (*AI Mining; Minería de IA*)**

1. Mineração de dados (*Data mining; Minería de datos*)
  - a) Big Data (*Big data; Big data*)
  - b) Ferramentas de mineração de dados (*Data mining tools; Herramientas de minería de datos*)
  - c) Mineração de conceito (*Concept mining; Minería conceptual*)
2. Mineração de textos (*Text mining; Minería de textos*)

**Modelos de Linguagem (*Language models; Modelos de lenguaje*)**

1. BERT (*BERT; BERT*)
2. Gemini (*Gemini; Gemini*)
3. GPT (*GPT; GPT*)
4. Grandes modelos de linguagem (*Large language models; Grandes modelos de lenguaje*)
5. Grandes modelos multimodais (*Large multimodal models; Grandes modelos multimodales*)
6. OpenAI (*OpenAI; OpenAI*)

**Organização do Conhecimento (*Knowledge organization; Organización del conocimiento*)**

1. Indexação (*Indexing; Indexación*)
  - a) Critérios de indexação automática (*Automatic indexing criteria; Criterios de indexación automática*)
  - b) Indexação automática (*Automatic indexing; Indexación automática*)
  - c) Indexação automática por atribuição (*Automatic indexing by assignment; Indexación automática por atribución*)
  - d) Indexação automática por extração (*Automatic indexing by extraction; Indexación automática por extracción*)
  - e) Indexação semiautomática (*Semi-automatic indexing; Indexación semiautomática*)
2. Catalogação (*Cataloging; Catalogación*)
  - a) Catalogação automatizada (*Automated cataloging; Catalogación automatizada*)

**Processamento de Linguagem Natural (*Natural language processing; Procesamiento del lenguaje natural*)**

1. PLN (*NLP; PLN*)
2. Sistemas de processamento de linguagem natural (*Natural language processing systems; Sistemas de procesamiento del lenguaje natural*)
3. TensorFlow (*TensorFlow; TensorFlow*)

**Redes Neurais (*Neural networks; Redes neuronales*)**

1. Rede adversarial generativa (*Generative adversarial network (gan); Red adversarial generativa*)
2. Rede neural artificial (*Artificial neural network; Red neuronal artificial*)
3. Rede neural de grafos (*Graph neural network (GNN); Graficar red neuronal*)
4. Rede neural profunda (*Deep neural network; Red neuronal profunda*)
5. Redes neurais artificiais (*Artificial neural networks; Redes neuronales artificiales*)
6. Redes neurais convolucionais (*Convolutional neural networks; Redes neuronales convolucionales*)
7. Redes neurais generativas (*Generative neural networks; Redes neuronales generativas*)

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

**Responsabilidade Ética (Ethical responsibility; Responsabilidad ética)**

1. Ética da informação relacionada à IA (*Information ethics related to AI; Ética de la información relacionada con IA*)
2. Ética na ciência (*Ethics in science; Ética en la ciencia*)
3. Ética na inteligência artificial (*AI ethics; Ética en inteligencia artificial*)
4. Explicabilidade (*Explainability; Explicabilidad*)
5. IA, aspectos éticos e legais (*AI, ethical and legal aspects; IA, aspectos éticos y legales*)
6. IA, aspectos jurídicos (*AI, legal aspects; IA, aspectos jurídicos*)
7. Inteligência artificial responsável (*Responsible artificial intelligence; Inteligencia artificial responsable*)
8. Interpretabilidade algorítmica (*Algorithmic interpretability; Interpretabilidad algorítmica*)
9. Modelo explicável (*Explainable model; Modelo explicable*)
10. Política de dados de pesquisa (*Research data policy; Política de datos de investigación*)
11. Política de Regulação de Inteligência Artificial (*Artificial Intelligence Regulatory Policy; Política de regulación de la inteligencia artificial*)
12. Políticas de IA (*AI policies; Políticas de IA*)
13. Uso de IA (*Use of AI; Uso de IA*)
14. Uso de IA comunicativa (*Use of communicative AI; Uso de IA comunicativa*)
15. Uso de IA por pesquisadores (*Use of AI by researchers; Uso de IA por investigadores*)
16. Vigilância de dados (*Data surveillance; Vigilancia de datos*)

**Sistemas Inteligentes (Smart systems; Sistemas inteligentes)**

1. Assistentes de voz (*Voice assistants; Asistentes de voz*)
2. Assistentes virtuais (*Virtual assistants; Asistentes virtuales*)
3. Automação Inteligente (*Smart Automation; Automatización inteligente*)
  - a. Agentes inteligentes (*Intelligent agents; Agentes inteligentes*)
  - b. Robôs (*Robots; Robots*)
  - c. Robótica (*Robotics; Robótica*)
  - d. Sistemas conversacionais (*Conversational systems; Sistemas conversacionales*)
  - e. Sistema de resposta a perguntas (*Question answering systems; Sistemas de respuesta a preguntas*)
  - f. Sistemas de apoio à decisão (*Decision support systems; Sistemas de apoyo a la decisión*)
  - g. Sistemas especialistas (*Expert systems; Sistemas expertos*)
  - h. Sistemas inteligentes (*Intelligent systems; Sistemas inteligentes*)
  - i. Sistemas multiagentes (*Multi-agent systems; Sistemas multiagente*)
  - j. Yolov5 (*Yolov5; Yolov5*)
4. BARD (*BARD, BARD*)
5. Bots (*Bots; Bots*)
6. Chatbot (*Chatbot; Chatbot*)
7. ChatGPT (*ChatGPT; ChatGPT*)
8. Gêmeo Digital Dinâmico (*Dynamic Digital Twin; Gemelo digital dinámico*)
9. Gêmeos digitais urbanos (*Urban digital twins; Gemelos digitales urbanos*)
10. Gemação digital (*Digital Twins; Hermanamiento digital*)
11. LangChain (*LangChain; LangChain*)
12. Ollama (*Ollama; Ollama*)

**Sistemas Inteligentes em Bibliotecas (Smart Systems; Sistemas inteligentes)**

1. Adoção de Inteligência Artificial (*Artificial intelligence Adoption; Adopción de inteligencia artificial*)
2. Biblioteca Inteligente (*Smart Library; Biblioteca inteligente*)
3. Blockchain aplicado à IA (*Blockchain applied to AI; Blockchain aplicada a la IA*)
4. Computação cognitiva (*Cognitive computing; Computación cognitiva*)
5. Humanidades digitais (*Digital humanities; Humanidades digitales*)
6. Inteligência artificial em bibliotecas (*Artificial intelligence in libraries; Inteligencia artificial en bibliotecas*)
7. Jogos eletrônicos (*Electronic games; Juegos electrónicos*)
8. Métodos de recuperação de informações (*Information retrieval methods; Métodos de recuperación de información*)
9. Serviços de gestão do conhecimento (*Knowledge Management Services; Servicios de gestión del conocimiento*)
10. Verificação de fatos (*Fact Check; Verificación de hechos*)

**Visão Computacional (Computer vision; Visión por computadora)**

1. Compreensão de leitura de máquina (*Machine reading comprehension; Comprensión de lectura automática*)
2. Detecção de objetos (*Object detection; Detección de objetos*)
3. Escuta de máquina (*Machine listening; Escucha de máquina*)
4. Processamento de imagens (*Image processing; Procesamiento de imágenes*)
5. Reconhecimento de padrões (*Pattern recognition; Reconocimiento de patrones*)
6. Reconhecimento facial (*Facial recognition; Reconocimiento facial*)
7. Tecnologias de Informação Digital (*Digital Information Technologies; Tecnologías de la información digital*)

)

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

IA)	Tecnologias orientadas por IA ( <i>AI-Driven technologies; Tecnologías impulsadas por IA</i> )
	Tecnologias semânticas ( <i>Semantic technologies; Tecnologías semánticas</i> )
<b>Web Semântica (<i>Semantic Web; Web semántica</i>)</b>	
Interação ( <i>Interaction; Interacción</i> )	Interação humano - computador ( <i>Human-computer interaction; Interacción humano - computadora</i> )
	Interação humano - IA ( <i>Human-AI interaction; Interacción humano - IA</i> )
	Linguagem ( <i>Language; Idioma</i> )
	Linguagem de busca ( <i>Search language; Idioma de búsqueda</i> )
indexación)	Linguagem de indexação ( <i>Indexing language; Lenguaje de indexación</i> )
documental)	Linguagem documentária ( <i>Documentary language; Lenguaje documental</i> )
	Semântica ( <i>Semantics; Semántica</i> )

Fonte: Elaborado pelos autores (2025).

## 5 CONSIDERAÇÕES FINAIS

A presente proposta de estudo teve como principal objetivo a formação de uma taxonomia de domínio voltada à intersecção entre a IA e a OC. Conforme apontado por Campos e Gomes (2007), a disponibilidade de uma taxonomia fornece ao usuário não apenas um guia sobre o assunto a ser investigado, mas também uma ferramenta de ensino, que pode revelar campos de interesse antes inexplorados pelo pesquisador.

Os resultados deste estudo confirmam o papel da Inteligência Artificial como agente transformador na Organização do Conhecimento, destacando-se o surgimento de ferramentas generativas e de código aberto que ampliam o acesso à tecnologia e viabilizam novas práticas de representação e recuperação da informação.

A análise demonstrou que a IA permeia diversos aspectos da Organização do Conhecimento, da automação de processos de indexação à análise semântica e ao desenvolvimento de interfaces inteligentes, ao mesmo tempo em que revela lacunas críticas em torno da ética, da confiabilidade dos modelos, da curadoria de dados e da manutenção da compreensão contextual, áreas nas quais a intervenção humana permanece necessárias.

A construção dessa taxonomia também permitiu ampliar o debate teórico sobre o papel dos vocabulários controlados e das linguagens documentárias em ambientes mediados por IA. Em uma era marcada pelo crescimento exponencial de dados, torna-se essencial que a OC se reinvente, sem abdicar de seus princípios fundamentais, mas adaptando-se às novas lógicas de produção e circulação de conhecimento.

A taxonomia proposta constitui instrumento tanto para o ensino e formação de usuários na área de Organização do Conhecimento como para a implementação de sistemas de indexação automática em bibliotecas e repositórios digitais. Sua aplicação prática pode

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

impactar positivamente a consistência e a exaustividade da indexação, ao fornecer vocabulários padronizados que alimentem mecanismos de RAG (Retrieval-Augmented Generation) e outras soluções baseadas em IA.

Por fim, acredita-se que este trabalho pode servir como base para futuras pesquisas que pretendam explorar as interseções entre a IA, a Organização do Conhecimento e a Web Semântica, contribuindo tanto para a inovação tecnológica quanto para o fortalecimento da atuação dos profissionais da informação em contextos digitais cada vez mais complexos.

#### **REFERÊNCIAS**

ABADAL, E.; ANGLADA, L. Tic e Bibliotecas: situação atual e perspectivas. *In*: RIBEIRO, A. C. M. L.; GONÇALVES, P. C.(Org.). **Biblioteca do século XXI**: desafio e perspectivas. Brasília: IPEA, 2017.

ALCAIDE, G. S. *et al.* Análise comparativa e de consistência entre representação automática e manual de informações documentárias. **Transinformação**, v. 13, n. 1, 2001.

ANDROUTSELLIS-THEOTOKIS, S.; SPINELLIS, D.; KECHAGIA, M. *et al.* Open source software: a survey from 10,000 feet. **Technology, Information and Operations Management**, v. 4, n. 3–4, p. 187-347, 2011. Disponível em: DOI: 10.1561/02000000026. Acesso em: 23 abr. 2025.

BERNERS-LEE, T.; LASSILA, O.; HENDLER, J. The Semantic Web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific America**, v. 284, n. 5, 2001.

CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e classificação: a categorização como princípio. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais [...]**Salvador: ANCIB, 2007.

CHIZZOTTI, A. **Pesquisa qualitativa em ciências humanas e sociais**. Petrópolis: Vozes, 2014.

CURRÁS, E. **Ontologias, taxonomia e tesouros em teoria de sistemas e sistemática**. Brasília: Thesaurus, 2010.

DE LUCA, E.W.; FALLUCCHI, F.; MORELATO, R. Teaching an Algorithm How to Catalog a Book. **Computers** 10, n. 11, p. 155, 2021. Disponível em: 10.3390/computers10110155. Acesso em: 23 abr. 2025.

DODEBEI, V. L. D. **Tesouro**: linguagem de representação da memória documentária. Niterói; Rio de Janeiro: Intertexto; Ed. Interciência, 2014.

EUROPEAN COMMISSION (EU-U.S.), **Terminology and Taxonomy for Artificial Intelligence**. Brussels: EU-U.S, 2024. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence-second-edition> Acesso em: 14 abr. 2025.

FEUERRIEGEL, S. *et al.* Generative AI. Business and Information Systems Engineering, Heidelberg, v.66, p. 111-126, 2024. Disponível em: 10.1007/s12599-023-00834-

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

7. Acesso em: 23 abr. 2025.

GABRIEL JUNIOR, R. F.; LAIPELT, R. C. Thesa: ferramenta para construção de tesouro semântico aplicado interoperável. **Revista P2P and Inovação**, Rio de Janeiro, v. 3, n. 2, p.124- 145, 2017.

GÉRON, A. **Mãos à obra**: aprendizado de máquina com Scikit-Learn, Keras e TensorFlow: Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes. 2. ed. Rio de Janeiro: Alta Books. 2021.

GOMES, L. I. E. Transformação digital e Inteligência Artificial nos serviços de informação: inovação e perspectivas para a Ciência da Informação no mundo pós-pandemia. **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 15, n. 1, p. 148–166, 2022. Disponível em: 10.26512/rici.v15.n1.2022.41490. Acesso em: 23 abr. 2025.

GONZÁLEZ, J. A. M. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador: EDUFBA, 2014.

GUIMARÃES, J. A. C.; SANTOS, A. Y. dos; SALES, R. de; MATOS, D. F. de O. A dimensão conceitual da organização do conhecimento nos congressos da NASKO: análise de conteúdo Bardiana. In: GUIMARÃES, J. A. C.; DODEBEI, V. (org.). **Organização do conhecimento e diversidade cultural**. Marília: ISKO-Brasil; FUNDEPE, 2015. p. 55-72.

HJØRLAND, B. Automatic Indexing. In: HJØRLAND, B. **Lifeboat for Knowledge Organization**. [S. l.: s. n.], [2008].

HJØRLAND, B. Knowledge Organization (KO). **Knowledge Organization: Reviews of Concepts in Knowledge Organization**, v. 43, n. 6, p. 475-48, 2016.

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

LANCASTER, F. W. **Indexing and Abstracting in Theory and Practice**. 2. ed. Illinois, EUA: University of Illinois, 1998.

LIEU, T. A.; PLATT, R. Applied research and development in health care: meshift. **The New England Journal of Medicine**, Waltham, v. 376, n. 8, p. 710–713, 2017.

LÓPEZ YEPES, J. **La ciencia de la información documental**: el documento, la disciplina y el profesional en la era digital. México: Universidad Panamericana. 2015.

LUHN, H. P. A Business Intelligence System. **IBM Journal of Research and Development**, New York, v. 2, n. 4, p. 314-319, 1958a.

LUHN, H. P. The Automatic Creation of Literature Abstracts. **IBM Journal of Research and Development**, New York, v. 2, n. 2, p. 159, 1958b.

MARKEY, K. **Forty Years of Classification Online**: Final Chapter or Future Unlimited? *Cataloging & Classification Quarterly*, 42:3-4, 1-63, 2006. DOI: 10.1300/J104v42n03\_01  
Acesso em: 07 fev. 2024.

MENDES, I.; BENTES, V. Taxonomia nas áreas da Biblioteconomia e da Ciência da Informação:

**Comentado [A1]:** Verifique se o ano de publicação é 2008 mesmo.

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

uma revisão sistemática. **Páginas A&B**, Arquivos e Bibliotecas (Portugal), v. 12, p. 36–47, 2019.

MENDES, I.; PINTO, V. B. Taxonomia nas áreas da biblioteconomia e da Ciência da Informação: uma revisão sistemática. **Páginas A&B**, Arquivos e Bibliotecas (Portugal), n. 12, 2019.

NARUKAWA, C. M.; SALES, R. A relevância e influência de linguagens documentárias na política de tratamento da informação. *In*: GIL-LEIVA, I.; FUJITA, M. S. L. **Política de indexação**. São Paulo (SP): Cultura Acadêmica; Marília: Oficina Universitária, 2021.

NASCIMENTO, F. M. S.; PINHO, F. A. Sistemas de organização do conhecimento: semelhanças e diferenças. **Convergências em Ciência da Informação**, Aracaju, v. 2, n. 3, p. 104- 122, set./dez. 2019.

NOCETTI, M. A. *et al.* Avaliação dos pacotes bibliográficos do serviço automatizado de disseminação seletiva da informação da empresa brasileira de pesquisa agropecuária sdi/embrapa. **Revista de Biblioteconomia de Brasília**, Brasília, v. 6, n. 2, 1978.

NOVO, H. F. A taxonomia enquanto estrutura classificatória: uma aplicação em domínio de conhecimento interdisciplinar. **Ponto de Acesso**, v. 4, n. 2, 2010.

RENDÓN-ROJAS, M. A. La ciencia de la información en el contexto de las ciencias sociales y humanas. ontología, epistemología, metodología e interdisciplina. **DataGramZero**, v. 9, n. 4, 2008.

RICHMOND, P. A. A reconsideration of enumerative classification for current information needs. **Ciência da Informação**, v. 3, n. 1, 1974.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. **Estudos Avançados em Biblioteconomia e Ciência da Informação**, v. 1, n. 1, 1982.

ROBREDO, J.; FERREIRA, J. A. P. Conceituação de um programa para indexação automática de textos. **Revista de Biblioteconomia de Brasília**, v. 8, n. 2, 1980.

RUSSEL, S. J.; NORVIG, P. **Inteligência Artificial: uma abordagem moderna**. 4. ed. [2a Reimp.]. Rio de Janeiro: LTC., 2024.

TELLA, A. Robots are coming to the libraries: are librarians ready to accommodate them? **Library Hi Tech News**, v. 37, n. 8, 2020. Disponível em: 10.1108/lhtn-05-2020-0047. Acesso em: 23 abr. 2025.

TEUBNER, T. *et.al.* **Welcome to the era of ChatGPT**. Business and Information Systems Engineering, v. 65, n.2, p. 95–101, 2023.

**AGRADECIMENTO**

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

Este trabalho contou com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, por meio da Bolsa de Produtividade em Pesquisa (Processo nº 133119/2025-4), o que tem contribuído significativamente para o desenvolvimento desta linha de investigação.

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB**  
**Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

VALENTIM, M. L. P. O perfil das bibliotecas contemporâneas. *In*: RIBEIRO, A. C. M. L.; FERREIRA, P. C. G. (orgs.). **Biblioteca do século XXI**: desafios e perspectivas. Brasília: Ipea, 2016.

WORLD WIDE WEB CONSORTIUM. **SKOS Simple Knowledge Organization System**. [Home Page](#). 2012.

ZUPAN, B.; DEMSAR, J. **Open-Source Tools for Data Mining**. *Clinics in Laboratory Medicine*, v. 28, Issue 1, 2008.

**Comentado [A2]:** Informe o link e a data de acesso.