



VIII ISKO BRASIL 2025



Revisão narrativa da literatura sobre o uso de IA generativa na construção e manutenção de taxonomias

Amanda J. C. de Oliveira Souza*
Elisângela Cristina Aganette**

*Universidade Federal de Minas Gerais; amanda.jercika@gmail.com

**Universidade Federal de Minas Gerais; elisangelaaganette@gmail.com



Na sociedade atual, onde o conhecimento é um dos ativos estratégicos, a forma como as organizações estruturam e acessam sua informação se tornou um diferencial competitivo.

Nesse cenário, as **taxonomias** cumprem um papel fundamental: elas organizam o conhecimento, categorizam conteúdos e ajudam a tornar a informação acessível, **mas só são eficazes quando bem adaptadas à realidade da organização.**

O problema é que, diante do volume massivo de dados com que lidamos hoje, **manter essas estruturas atualizadas de forma manual se tornou um grande desafio.**

É nesse ponto que surge a pergunta central desta pesquisa:

de que maneira a inteligência artificial generativa pode ser empregada na otimização da construção e reformulação de taxonomias?



Este trabalho investiga como a IA generativa pode ser aplicada nesse contexto, identificando **potenciais contribuições** e **desafios reais**.

Conectar duas frentes: **os fundamentos da Ciência da Informação e as capacidades emergentes da IA generativa**, explorando suas aplicações teóricas e práticas na organização do conhecimento.



- Este estudo tem por objetivo analisar, a partir da literatura recentemente publicada, como os modelos e técnicas de inteligência artificial generativa (GAI) podem contribuir para construção e reformulação de taxonomias.



Metodologia



VIII ISKO BRASIL 2025



- Pesquisa qualitativa.
- Natureza exploratória.
- Revisão narrativa.
- Estruturada em três fases.

Revisão narrativa

- (1) Definição do problema e questão de orientadora.
- (2) Scopus e WoS
- (3) Definição palavras-chaves (inglês)
- (4) Aplicação de filtro (2016-2025)
- (5) Restrição do idioma (EN - ES - PT) e publicações Open Access.
- (6) Leitura exploratória e seleção de artigos significativos (título, resumo e palavras-chave).

Análise dos métodos e técnicas da GAI aplicadas na construção e reformulação de taxonomias

Compreender de que maneira essas abordagens podem ser aplicadas para a construção e reformulação de taxonomias, destacando vantagens, desafios e oportunidades

1ª

2ª

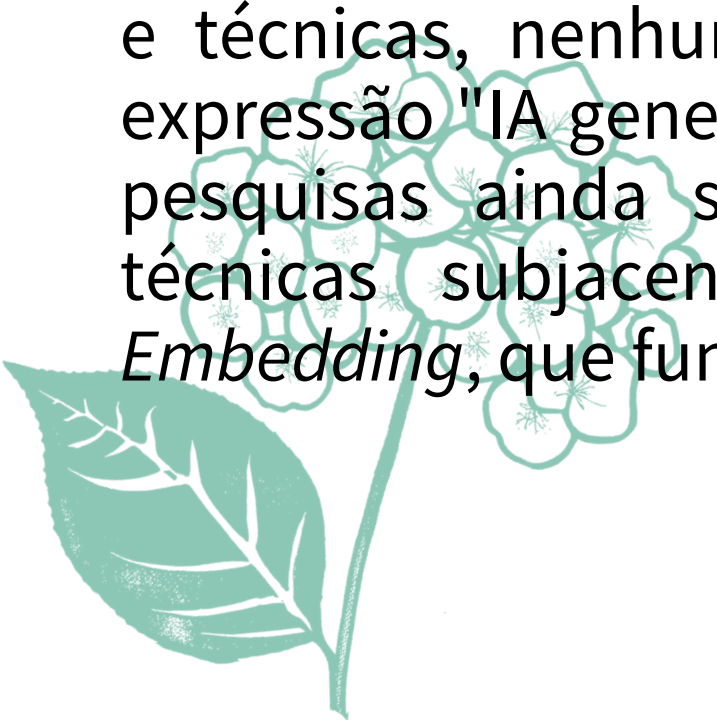
Consolidação dos resultados e análise de conteúdo

- Análise de conteúdo para seleção de estudos relevantes.
 - Identificação dos principais métodos e técnicas utilizado na criação e reformulação de taxonomias por meio de GAI.
 - Critérios de inclusão e exclusão dos documentos analisados consideraram a pertinência ao tema e solidez metodológica.

3ª

Fase 1

- A análise dos artigos revelou o uso de IA generativa com diferentes níveis de sofisticação, abrangendo desde LLMs contemporâneos até modelos clássicos, como seq2seq e sistemas estatísticos, a exemplo do Moses.
- Embora o avanço da IA generativa seja evidente nas conexões entre termos e técnicas, nenhum dos trabalhos analisados utilizou explicitamente a expressão "IA generativa" como palavra-chave. Esse achado sugere que as pesquisas ainda se concentram na exploração e desenvolvimento de técnicas subjacentes, como NLP, *Machine Learning*, LLMs e *Word Embedding*, que fundamentam esse novo paradigma tecnológico.

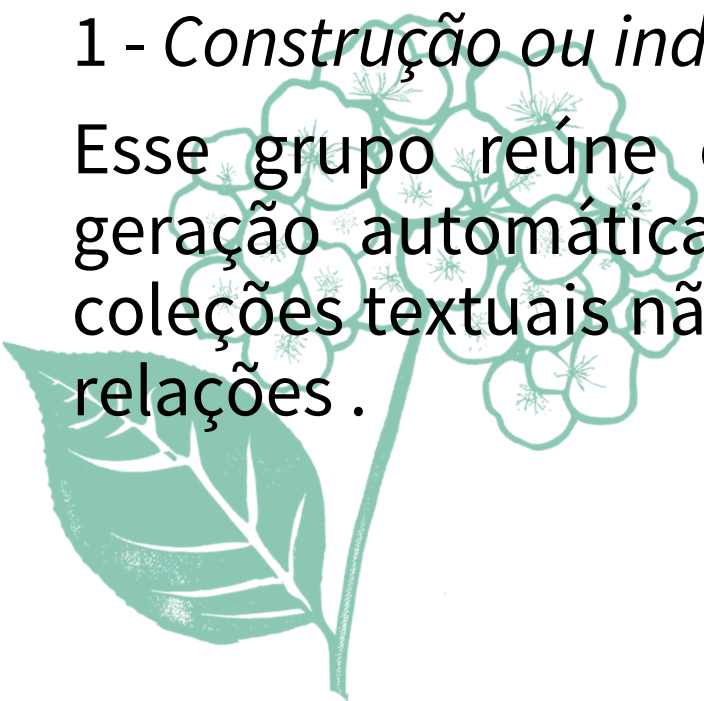


Fase 2

A partir da análise sistemática observou-se que os artigos se distribuem em quatro grandes enfoques metodológicos principais, que definem o papel da IA generativa no ciclo de vida das taxonomias:

1 - *Construção ou indução de taxonomias a partir de dados textuais.*

Esse grupo reúne estudos, que utilizam modelos de linguagem para a geração automática de hierarquias taxonômicas, geralmente a partir de coleções textuais não estruturadas. A IA generativa é empregada para prever relações.



2 - *Expansão e enriquecimento de taxonomias existentes*

A IA generativa é utilizada para propor novos termos ou sugerir inserções hierárquicas em taxonomias preexistentes. Os métodos adotados variam entre o uso de *embeddings* semânticos e abordagens baseadas em modelos seq2seq. **Validação humana frequente** em contextos institucionais.

3 - *Avaliação e refinamento automatizado de taxonomias*

Modelos generativos como avaliadores da coerência semântica e estrutural de taxonomias. Modelos como GPT-4, Claude-3 e Qwen-14B são empregados para estimar a adequação de tópicos, identificar duplicatas ou outliers e sugerir refinamentos estruturais, como a eliminação de ciclos, a unificação de sinônimos e a reestruturação de vínculos taxonômicos.

4- Geração de conceitos e rótulos taxonômicos

Embora menos frequente na literatura revisada, essa abordagem é também uma das mais inovadoras, pois foca na geração autônoma de conceitos linguísticos a partir de contextos hierárquicos. Utilizando modelos baseados em grafos, os sistemas são capazes de nomear nós intermediários ausentes na estrutura taxonômica com base em informações distribuídas entre conceitos vizinhos.





USO DA GAI APLICADO À TAXONOMIAS

POTENCIALIDADES

DESAFIOS

Geração automática de conceitos e relações hierárquicas

A IA gera novos nós, nomes e relações parônimas com fluidez semântica e linguística.

Risco de incoerência semântica, tal como conceituações ambíguas, termos genéricos ou relações inadequadas podem comprometer a integridade da taxonomia.

Enriquecimento com linguagem natural contextualizada

Inclusão de termos emergentes a partir de descrições coletadas da web e de usuários.

Introdução de ruído semântico, a informalidade, ambiguidade ou superficialidade dos dados textuais podem degradar a qualidade conceitual da estrutura.

Reformulação estrutural automatizada

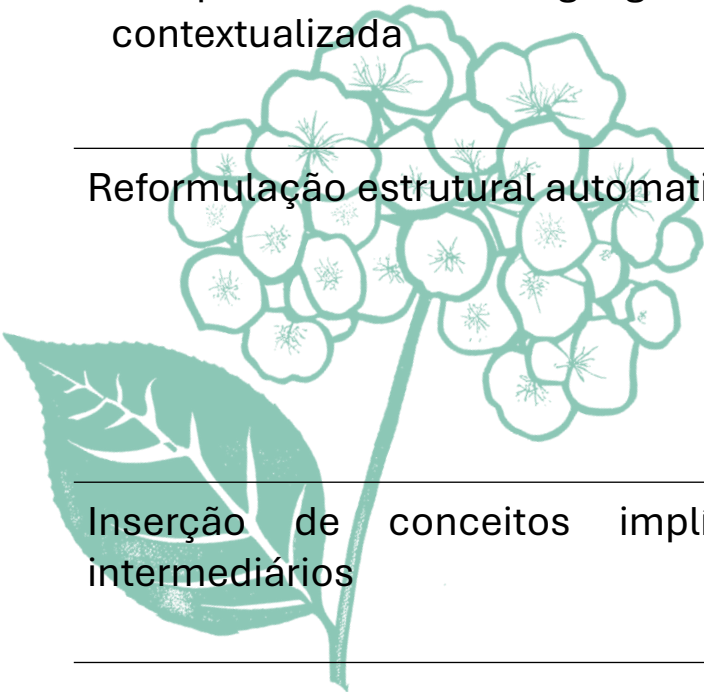
Construção de sistemas que refinam árvores complexas, removem redundâncias e reorganizam hierarquias.

Ausência de justificativa das mudanças, impactam na rastreabilidade e visibilidade de informações claras sobre as transformações aplicadas. Acarretando a dificuldade de corrigir e auditar as taxonomias.

Inserção de conceitos implícitos e intermediários

Preenchimento de lacunas não explícitas, fortalecendo a coesão textual.

Modelos podem inserir conceitos não validados ou irrelevantes, criando ruído ou duplicidade na taxonomia.



GOETHE
INSTITUT



Instituto Brasileiro de Informação
em Ciência e Tecnologia



FAPERGS



CNPq

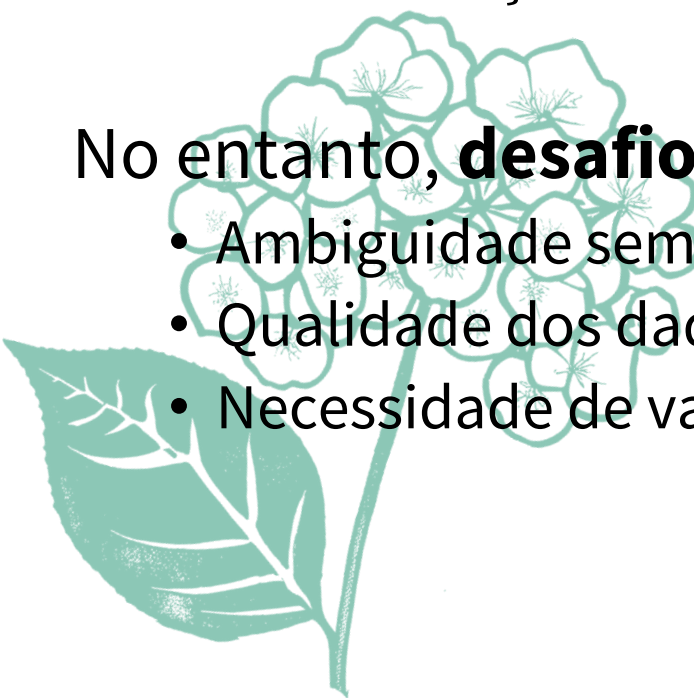


CAPES

- A IA generativa mostra grande potencial para **automatizar a construção e o refinamento de taxonomias**, com técnicas eficazes para organizar grandes volumes de dados. Entre os avanços, destacam-se:
 - Geração automática de conceitos e relações.
 - Enriquecimento com linguagem contextualizada.
 - Reformulação estrutural de forma escalável.

No entanto, **desafios persistem**, como:

- Ambiguidade semântica e vieses algorítmicos.
- Qualidade dos dados de entrada.
- Necessidade de validação humana.



- A pesquisa também apontou para uma lacuna importante na literatura sobre o papel dos profissionais da Ciência da Informação, como bibliotecários e taxonomistas, na curadoria dos resultados gerados por sistemas de IA. Essa lacuna representa uma oportunidade significativa para futuras investigações, que podem explorar formas de colaboração entre inteligência artificial e especialistas humanos, visando garantir a qualidade, a confiabilidade e a relevância das taxonomias aplicadas.



Agradecimentos



VIII ISKO BRASIL 2025



FAPEMIG



Programa de Pós Graduação
GESTÃO & ORGANIZAÇÃO DO
CONHECIMENTO

U F *m* G



- ADRIANI, M. L. *Um método para a construção de taxonomias utilizando a DBPEDIA*. 149 fls. 2017. Dissertação (Mestrado em Engenharia do Conhecimento) - Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina.
- AGANETTE, E. C.; ALVARENGA, L.; SOUZA, R. R. *Taxonomia Corporativa: Estudo sobre Definições e Etapas de Construção com Base na Literatura Publicada*. Our Knowledge Publishing, 2024.
- AGANETTE, E. C.; MACULAN, B. C. M. S. *A complexidade da classificação: Representação no ensino de ciência da informação*. Seven Editora, 2023. Disponível em: <https://sevenpublicacoes.com.br/editora/article/view/2346>. Acesso em: 24 fev. 2025.
- ANDERSON, L. W.; KRATHWOHL, D. R. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, 2001. Disponível em: <https://www.worldcat.org/title/taxonomy-for-learning-teaching-and-assessing>. Acesso em: 24 fev. 2025.
- BAILEY, K. D. *Typologies and Taxonomies: An Introduction to Classification Techniques*. SAGE Publications, 2012. Disponível em: <https://us.sagepub.com/en-us/nam/typologies-and-taxonomies/book52363>. Acesso em: 24 fev. 2025.
- CHEN, B; VARRÓ, D., "Prompting or Fine-tuning? A Comparative Study of Large Language Models for Taxonomy Construction," 2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C), Västerås, Sweden, 2023, pp. 588-596.
- BLOOM, B. S. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Longman, 1956. Disponível em: <https://eric.ed.gov/?id=ED056243>. Acesso em: 24 fev. 2025.
- BROWN, A.; SILVA, P. *Advances in Digital Taxonomies and AI-Driven Classification*. Journal of Information Science, 2023. Disponível em: <https://journals.sagepub.com/home/jis>. Acesso em: 24 fev. 2025.
- CASAR CORREDERA, J. R. Inteligência artificial generativa. *Anales... Anales de la Real Academia de Doctores de España*. Volumen 8, número 3 – 2023, páginas 475-489. Disponível em: <https://www.rade.es/imageslib/PUBLICACIONES/ARTICULOS/V8N3%20-%20001%20-%20ED%20-%20CASAR.pdf>. Acesso em: 8 fev. 2025.
- COZMAN, F. G.; NERI, H. O que, afinal, é inteligência artificial? In: COZMAN, F. G.; PLONSKI, G. A.; NERI, H. (org.) *Inteligência artificial: avanços e tendências*. São Paulo: Instituto dos estudos avançados, 2021. Disponível em: https://www.mpsp.mp.br/portal/page/portal/documentacao_e_divulgacao/doc_biblioteca/bibli_servicos_produtos/BibliotecaDigital/BibDigitalLivros/TodosOsLivros/Inteligencia-artificial-avancos-e-tendencias.pdf. Acesso em: 12 fev. 2025.
- D'AMICO, S.; et al. Enriching Skill Taxonomies through Vector Space Models, 2024 *IEEE International Conference on Big Data (BigData)*, Washington, DC, USA, 2024, pp. 2297-2302, doi: 10.1109/BigData62323.2024.10825415. Acesso em: 03 mar. 2025.

- GONZÁLEZ, R.; PEREIRA, J. *Machine Learning Applications in Taxonomy Development*. Information Systems Review, 2022. Disponível em: <https://www.infosysreview.com/articles/machine-learning-taxonomy>. Acesso em: 14 mar. 2025.
- GRUBER, T. R. *A translation approach to portable ontology specifications*. Knowledge Acquisition, v. 5, n. 2, p. 199-220, 1993. Disponível em: <https://tomgruber.org/writing/ontology-approach.pdf>. Acesso em: 24 fev. 2025.
- LINNAEUS, C. *Systema Naturae*. 1735. Disponível em: <https://www.biodiversitylibrary.org/item/15311>. Acesso em: 24 fev. 2025.
- MAYR, E. *Principles of Systematic Zoology*. McGraw-Hill, 1969. Disponível em: <https://archive.org/details/principlesofsystematiczoology>. Acesso em: 24 fev. 2025.
- MCCARTHY, J. *Formalizing Common Sense: Papers by John McCarthy*. Ablex Publishing, 1995. Disponível em: <https://www-formal.stanford.edu/jmc/common-sense.html>. Acesso em: 24 fev. 2025.
- MORAES, D. Using Zero-shot Prompting in the Automatic Creation and Expansion of Topic Taxonomies for Tagging Retail Banking Transactions. *arXiv*, v1, 8 jan. 2024. Disponível em: <https://arxiv.org/abs/2401.06790>. Acesso em: 20 mar. 2025.
- MOSKVORETSKII, V. et al. TaxoLLaMa: WordNet-based model for solving multiple lexical semantic tasks. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 62, 2024, Bangkok. *Proceeding [...]*. Arxiv: New York, 2024. Disponível em: <https://arxiv.org/pdf/2403.09207>. Acesso em: 29 abr. 2025.
- NICKERSON, R. C.; MUNTERMANN, J.; VARSHNEY, U. *Taxonomy Development in Information Systems: Developing a Taxonomy of Mobile Applications*. European Conference on Information Systems (ECIS), 2009. Disponível em: <https://aisel.aisnet.org/ecis2009/158/>. Acesso em: 25 mar. 2025.
- PADILLA, C. *Hybrid Taxonomies for Digital Knowledge Management*. Knowledge Organization Journal, 2021. Disponível em: <https://www.knoworg.org/articles/hybrid-taxonomies>. Acesso em: 26 fev. 2025.
- PENG, Y.; BONALD, T.; ALAM, M. Refining Wikidata taxonomy using large language models. In: ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 33, 2024, Boise. *Proceedings [...]*. ACM: New York, 2024. Disponível em: <https://dl.acm.org/doi/10.1145/3627673.3679156>. Acesso em: 29 abr. 2025.
- PRODANOV, C. C.; FREITAS, E. C. *Metodologia do trabalho científico*. 2. ed. 2013.
- RANGANATHAN, S. R. *Colon Classification*. Madras Library Association, 1933. Disponível em: <https://www.isko.org/cyclo/colon-classification>. Acesso em: 26 fev. 2025.
- ROWLEY, J. *Organizing Knowledge: An Introduction to Information Retrieval*. Gower Publishing, 1992. Disponível em: <https://www.worldcat.org/title/organizing-knowledge-an-introduction-to-information-retrieval>. Acesso em: 26 fev. 2025.
- RUSSELL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. 4. ed. Hoboken: Pearson, 2021. Acesso em: 12 fev. 2025.

- SANTO, A. de. Taxonomy enrichment: a framework for automatic updates and labor market analysis. In: INTERNATIONAL CONFERENCE OF THE ITALIAN ASSOCIATION OF ARTIFICIAL INTELLIGENCE BOLZANO, 23, 2024, Italy. *Proceedings* [...]. [S. n.; s.l.], 2024. Disponível em: <https://ceur-ws.org/Vol-3914/short71.pdf>. Acesso em: 29 abr. 2025.
- SHI, J. et al. Taxonomy completion via implicit concept insertion. In: ACM WEB CONFERENCE, 2024, Singapore. *Proceeding* [...] ACM: New York, 2024. Disponível em: <https://dl.acm.org/doi/10.1145/3589334.3645584>. Acesso em: 29 abr. 2025.
- SHIRKY, C. *Ontology is Overrated: Categories, Links, and Tags*. 2005. Disponível em: http://www.shirky.com/writings/ontology_overrated.html. Acesso em: 24 fev. 2025.
- SMITH, J.; JOHNSON, L. *Dynamic Taxonomies in the Age of Big Data*. Data Science Review, 2023. Disponível em: <https://www.datasciencereview.com/articles/dynamic-taxonomies-big-data>. Acesso em: 18 mar. 2025.
- TREVISAN, L. C.; DAMIAN, I. P. M. Gestão do conhecimento: diretrizes e práticas recomendadas às organizações. *Ci.Inf.*, Brasília, DF, v.47 n.2, p.21-34, maio/ago. 2018. Disponível em: <https://www.brapci.inf.br/v/99232>. Acesso em: 17 fev. 2025.
- WELLER, M.; ANDERSON, T. *Taxonomy and Digital Learning: The Future of Classification Systems*. Educational Technology Research, 2020. Disponível em: <https://www.edtechresearch.org/taxonomy-digital-learning>. Acesso em: 24 fev. 2025.
- XU, H. et al. TacoPrompt: a collaborative multi-task prompt learning method for self-supervised taxonomy completion. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2023, Singapore. *Proceedings* [...]. ACL: Singapore, 2023. Disponível em: <https://aclanthology.org/2023.emnlp-main.979.pdf>. Acesso em: 29 abr. 2025.
- ZENG, Q. et al. Enhancing taxonomy completion with concept generation via fusing relational representations. In: KDD, 2021, Singapore. *Proceedings* [...]. Arxiv: New York, 2021. Disponível em: <https://arxiv.org/abs/2106.02974>. Acesso em: 29 abr. 2025.
- WOODS, E. The corporate taxonomy: creating a new order. *KMWorld*, Camden, v.13, n. 7, July/Aug. 2004.

