UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Relações semânticas na terminologia da Organização do Conhecimento:
Um estudo aplicado ao Diccionario de Organización del Conocimiento

Juliana Oliveira Mauch

PORTO ALEGRE
2026

Juliana Oliveira Mauch

Relações semânticas na terminologia da Organização do Conhecimento: Um estudo aplicado ao Diccionario de Organización del Conocimiento

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Mestre em Ciência da Informação.

Orientadora: Rita do Carmo Ferreira Laipelt

Porto Alegre
2026

JULIANA OLIVEIRA MAUCH

Relações semânticas na terminologia da Organização do Conhecimento:
Um estudo aplicado ao Diccionario de Organización del Conocimiento

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de Mestre em Ciência da Informação.

Orientadora: Rita do Carmo Ferreira Laipelt

Aprovada em:

BANCA EXAMINADORA:

_____
Profª. Dr.ª Rita do Carmo Ferreira Laipelt (PPGCIN - UFRGS)

_____
Prof. Dr. Mario Barité (FIC - Universidad de la Republica)

_____
Prof. Dr. Thiago Henrique Bragato Barros (PPGCIN - UFRGS)

_____
Profª. Drª. Luciana Monteiro Krebs (UFRGS)

AGRADECIMENTOS

Agradeço à Universidade Federal do Rio Grande do Sul (UFRGS), por ter sido, em diferentes momentos da minha trajetória, espaço de formação, transformação e permanência. Foi nesta universidade pública, gratuita e de qualidade que pude realizar minha graduação em Letras, anos depois retornar para a Biblioteconomia e, agora, concluir o Mestrado em Ciência da Informação. Poder construir essa trajetória em uma instituição pública reafirma, para mim, não apenas a excelência da UFRGS, mas também a importância da universidade pública como lugar de produção de conhecimento, de formação crítica e de possibilidade concreta de continuidade acadêmica.

Expresso minha profunda gratidão à professora Rita do Carmo Ferreira Laipelt, minha orientadora, por sua presença firme, sensível e generosa ao longo deste percurso. Sua orientação foi fundamental não apenas pelo rigor intelectual e pela seriedade acadêmica com que conduziu este trabalho, mas também pelo apoio humano.

Agradeço também aos professores que participaram da minha banca de qualificação, cujas contribuições foram decisivas para o desenvolvimento desta dissertação. Ao professor Mario Barité, registro um agradecimento muito especial, porque foi a partir de sua sugestão de trabalhar metodologicamente com um único termo, indexação, que esta pesquisa encontrou um recorte mais claro, viável e produtivo. Ao professor Thiago Henrique Bragato Barros, agradeço pela atenção dedicada ao trabalho e pelas contribuições criteriosas, que foram importantes para dar mais clareza e consistência a esta dissertação.

Por fim, agradeço ao meu marido, Frederico, por tudo o que foi possível viver, suportar e atravessar ao longo desses dois anos de mestrado. Agradeço por ter permanecido ao meu lado em meio ao desgaste, ao desânimo, às crises, ao excesso e a tudo aquilo que este processo teve de difícil. Obrigada pelo apoio, pela paciência, pelo cuidado, pela escuta e pela companhia constante.

“Todos os dicionários juntos não contêm nem metade dos termos de que precisaríamos para nos entendermos uns aos outros.”
(José Saramago, O Homem Duplicado)

RESUMO

Esta dissertação investiga o conceito de indexação na Organização do Conhecimento (OC), buscando descrever seu *frame* e suas relações associativas tipadas em literatura especializada da área. Com base em OC, Ciência da Informação, Teoria de *Frames*, Terminologia Baseada em *Frames* e Análise de Domínio, o estudo parte da hipótese de que a indexação não é uma operação neutra, mas um ponto de condensação de disputas epistemológicas, tecnológicas e políticas. Metodologicamente, foram construídos dois *corpora* de artigos indexados na *Web of Science*: um *corpus* histórico, organizado por décadas (1970–2025), composto por 28 textos com o termo "*indexing*" no título, e um *corpus* temático ampliado, também com 28 artigos, recuperados por "*indexing*" em tópico associado a "*information science*" e "*knowledge organization*". Os artigos de ambos os *corpora* foram analisados com apoio do modelo de linguagem NotebookLM, a partir do mesmo *prompt* analítico e com os textos segmentados em blocos de cinco, permitindo extrair, para cada artigo, o *frame* "*indexing*", as relações associativas tipadas e excertos ilustrativos. Os resultados evidenciam que o *frame* de indexação se transforma ao longo do tempo: de modelos probabilísticos e estatísticos de recuperação, passando por abordagens sintático-lexicais, sistemas baseados em conhecimento e indexação semântica latente, até chegar a configurações marcadas por métricas bibliométricas, arquiteturas de processamento distribuído, *folksonomias* e infraestruturas de descoberta acadêmica. No *corpus* ampliado, a indexação aparece ainda articulada a teorias de domínio, Sistemas de Organização do Conhecimento (SOC), pesquisas feministas, estudos de proveniência e políticas de indexação em repositórios institucionais. A tipificação das relações associativas, interpretadas por meio de papéis Qualia, permite representar de forma mais fina os vínculos entre processos, agentes, instrumentos, propriedades e objetivos envolvidos em indexação, oferecendo subsídios para a construção de um tesauro-dicionário da OC ancorado em *frames* e para o uso metodologicamente controlado de Inteligência Artificial (IA) na análise semântica de conceitos em domínios especializados.

Palavras-chave: Organização do Conhecimento; Indexação; Teoria de *Frames*; Relações associativas tipadas; Inteligência artificial.

ABSTRACT

This dissertation investigates the concept of indexing within Knowledge Organization (KO), seeking to describe its frame and its typed associative relations in specialized literature of the field. Based on KO, Information Science, Frame Theory, Frame-Based Terminology, and Domain Analysis, the study stems from the hypothesis that indexing is not a neutral operation, but rather a condensation point of epistemological, technological, and political disputes. Methodologically, two corpora of articles indexed in the Web of Science were constructed: a historical corpus organized by decades (1970–2025), comprising 28 texts with the term "indexing" in the title, and an expanded thematic corpus, also with 28 articles, retrieved by searching "indexing" as a topic associated with "information science" and "knowledge organization". The articles from both corpora were analyzed with the support of the NotebookLM language model, using the same analytical prompt and with the texts segmented into blocks of five, allowing the extraction of the "indexing" frame, typed associative relations, and illustrative excerpts for each article. The results show that the indexing frame transforms over time: from probabilistic and statistical retrieval models, moving through syntactic-lexical approaches, knowledge-based systems, and latent semantic indexing, to configurations marked by bibliometric metrics, distributed processing architectures, folksonomies, and scholarly discovery infrastructures. In the expanded corpus, indexing also appears articulated with domain theories, Knowledge Organization Systems (KOS), feminist research, provenance studies, and indexing policies in institutional repositories. The typification of associative relations, interpreted through Qualia roles, allows for a finer representation of the links between processes, agents, instruments, properties, and objectives involved in indexing, offering insights for the construction of a KO frame-based thesaurus-dictionary and for the methodologically controlled use of Artificial Intelligence (AI) in the semantic analysis of concepts within specialized domains.

Keywords: Knowledge Organization; Indexing; Frame Theory; Typed associative relations; Artificial intelligence.

LISTA DE QUADROS

Quadro 1 – Anos 70
Quadro 2 – Anos 80
Quadro 3 – Anos 90
Quadro 4 – Anos 2000
Quadro 5 – Anos 2010
Quadro 6 – Anos 2020 até 2025
Quadro 7 – Artigos mais citados do *corpus* 2 e seus *frames* de indexação
Quadro 8 – Indexação semântica, folksonomias e epistemologias da indexação
Quadro 9 – Indexação manual, automática e "*social tagging*" em contextos documentais distintos
Quadro 10 – Dimensões políticas, infraestruturas terminológicas e indexação na *Web Semântica*
Quadro 11 – Indexação colaborativa, feminista, facetada e políticas de indexação em repositórios
Quadro 12 – Artigos sobre ética, viés e avaliação de sistemas de indexação
Quadro 13 – Síntese das relações associativas tipadas do *frame* "*indexing"*
Quadro 14 – Relações de "*indexing*" na árvore de domínio do dicionário

LISTA DE ABREVIATURAS E SIGLAS

AD – Análise de Domínio

CI – Ciência da Informação

IA – Inteligência Artificial

ISKO – International Society for Knowledge Organization

ISO – International Organization for Standardization

RC – Representação do Conhecimento (*Knowledge Representation*)

SOC – Sistemas de Organização do Conhecimento

OC – Organização do Conhecimento

SUMÁRIO

1 INTRODUÇÃO **9**
2 PROBLEMA E JUSTIFICATIVA **15**
2.1 Objetivo geral 19
2.2 Objetivos específicos 19
3 REFERENCIAL TEÓRICO **20**
3.1 Organização do Conhecimento 21
3.2 Relações semânticas e relações associativas nos SOC 24
3.3 Semântica de *Frames* 27
3.4 Organização do Conhecimento, Inteligência Artificial e convergência com a Representação do Conhecimento 29
4 METODOLOGIA **33**
4.1 Construção do *corpus* 1 - Busca por décadas 34
4.2 Construção do *corpus* 2 - Busca temática ampliada 38
4.3 Extração do *frame* "*indexing*" com apoio do NotebookLM 45
5 RESULTADOS **48**
6 ANÁLISE **82**
6.1 Análise do *corpus* por décadas **82**
6.1.1 Anos 1970: a promessa probabilística e o retorno do "about" 82
6.1.2 Anos 1980: entre o indexador cognitivo e os padrões sintáticos 83
6.1.3 Anos 1990: latência semântica, imagens e novas fronteiras cognitivas 85
6.1.4 Anos 2000–2010: bibliometria, escalabilidade e o acoplamento à infraestrutura. 86
6.1.5 Anos 2010–2020: visibilidade algorítmica, participação distribuída e infraestrutura de indexação 87
6.1.6 Anos 2020–2025: modelos combinados de IA e a resistência das Humanidades 90
6.1.7 Consequências epistemológicas e implicações para um tesauro-dicionário 91
6.2 Eixos analíticos do *corpus* ampliado **92**
6.2.1 Hjørland e a indexação como problema epistemológico 93
6.2.2 Tudhope, Golub e a indexação orientada por SOC e ontologias 94
6.2.3 Guimarães, Samuelsson e a dimensão ética-discursiva da indexação 95
6.2.4 Folksonomias, categorias de nível básico e práticas híbridas 96
6.2.5 Integração: do ecossistema ampliado ao modelo relacional do tesauro-dicionário 97
6.3 Papéis Qualia **99**
6.4 "*Indexing*" e a costura entre ramos da árvore de domínio **106**
7 CONCLUSÕES **110**
REFERÊNCIAS **114**
APÊNDICE 1 ..... **121**

1 INTRODUÇÃO

A terminologia, enquanto disciplina voltada à sistematização dos termos técnicos de um campo do saber, ocupa um lugar de destaque na Organização do Conhecimento (OC). Conforme o Dicionário Priberam da Língua Portuguesa, entende-se terminologia como "o tratado dos termos técnicos ou específicos de uma arte, ciência ou área de atividade", assim como o "conjunto desses termos técnicos". No campo da linguística, a terminologia é definida como o ramo dedicado ao estudo dos itens lexicais que correspondem a conceitos especializados. Embora essas definições sejam fundamentais, não esgotam a complexidade da terminologia no contexto da OC, cuja prática ultrapassa a simples reunião de vocábulos para abarcar as relações conceituais que os sustentam.

Para aprofundar essa compreensão, recorre-se ao “Diccionario de Organización del Conocimiento”, de Mario Barité *et al.* (2015, p. 155), que conceitua terminologia como o “âmbito do saber que se ocupa da recopilação, do estudo sistemático, da normalização e da inter-relação dos termos e das estruturas conceituais que se desenvolvem a partir deles e em torno deles.”[1]. Trata-se, portanto, não apenas de reunir termos, mas de organizar seus significados em um sistema de interdependências semânticas. A terminologia, nesse contexto, configura-se como uma instância mediadora entre o discurso técnico e a representação formal do conhecimento, essencial para o desenvolvimento de linguagens documentárias e para a viabilização de sistemas de recuperação da informação.

A área de OC tem na terminologia um de seus pilares, sobretudo no que tange à construção de vocabulários controlados, tesauros e ontologias. A explosão da produção informacional e a emergência de novos ambientes tecnológicos impuseram à terminologia desafios inéditos, exigindo instrumentos mais sofisticados para garantir a precisão conceitual e a eficiência na busca e recuperação de informações. Como salienta Dias (2000, p. 92), o avanço dos sistemas de organização da informação demanda o desenvolvimento de metodologias

[1] “ámbito del saber que se ocupa de la recopilación, el estudio sistemático, la normalización y la interrelación de los términos y de las estructuras conceptuales que se desarrollan a partir de, y en torno a ellos”.

terminológicas capazes de acompanhar a dinâmica do conhecimento científico e tecnológico.

Dentro desse panorama, a construção de vocabulários especializados demanda não apenas a definição acurada dos termos, mas também o mapeamento rigoroso das relações que os unem em redes semânticas coerentes. A diferenciação entre dicionários gerais e dicionários terminológicos, como ilustrado por Vézina *et al*. (2009), evidencia que, enquanto os primeiros apresentam definições de uso comum, os segundos estruturam conceitos dentro de um sistema de ideias interligadas. Essa distinção reforça a necessidade de um olhar atento às relações entre conceitos, pois não basta definir isoladamente, é preciso contextualizar semanticamente.

A relação intrínseca entre termo e conceito é sublinhada por Almeida (2021, p. 28), que ressalta a complexidade das unidades terminológicas, passíveis de múltiplas abordagens disciplinares e cognitivas. Cabré (2013), por sua vez, destaca que a Ciência da Informação (CI), ao adotar princípios terminológicos na indexação de documentos, reconhece a centralidade dos conceitos como núcleos organizadores do discurso especializado. Nesse sentido, a terminologia, enquanto instrumento de representação, transcende o mero arranjo lexical, configurando sistemas de significação que espelham a dinâmica dos campos científicos.

Os Sistemas de Organização do Conhecimento (SOC) emergem, nesse cenário, como instrumentos fundamentais da OC, possibilitando a codificação, a representação e a disseminação do conhecimento. Inicialmente eram chamados de Linguagens Documentárias e concebidos no âmbito da Documentação, evoluíram a partir das influências da linguística estrutural, particularmente das dicotomias propostas por Ferdinand de Saussure, que ressaltam a organização relacional dos signos. Como aponta Vogel (2007), a estruturação semântica dessas linguagens, materializada em vocabulários controlados, tesauros e taxonomias, assegura a estabilidade conceitual e facilita a interoperabilidade entre sistemas de informação.

A organização dos conceitos em estruturas hierárquicas, associativas e equivalentes (ISO 25964) não apenas garante a eficiência dos processos de indexação e recuperação, mas também reflete a lógica interna dos domínios científicos. Medeiros (1986, p. 136) reforça essa perspectiva ao destacar que as linguagens especializadas demandam a criação de elos precisos entre termos e conceitos, de modo a assegurar a exatidão na comunicação científica e tecnológica.

Green (2001) destaca que, enquanto as relações hierárquicas e de equivalência são classes "fechadas" (mais fáceis de definir), as relações associativas são classes "abertas" e praticamente infinitas. Isto significa que as relações associativas são o coração dos tesauros e sistemas de classificação. No presente trabalho, contudo, as relações hierárquicas e de equivalência permanecem como contexto descritivo; o foco analítico recai exclusivamente sobre as relações associativas.

Um dos pontos críticos para o tratamento de relações em SOC reside no fato de que as funções sintagmáticas não constituem propriedades intrínsecas dos conceitos, mas emergem do contexto em que estes são postos em relação. Em termos linguísticos, um mesmo nome pode desempenhar papel de agente ou de paciente/objeto, a depender da construção em que ocorre, o que desloca a determinação da função do conceito para o nível da estrutura predicado–argumento. Essa dependência do contexto torna as chamadas relações associativas particularmente voláteis e difíceis de estabilizar em tesauros, uma vez que combinações idênticas de termos podem ser interpretadas como relações de tipo agente–processo, instrumento–processo ou tema–processo, entre outras possibilidades. Essa instabilidade ajuda a explicar, como argumentam Bräscher (2014) e Green (2008), a heterogeneidade observada na forma como relações associativas são representadas em sistemas distintos, ainda que voltados para um mesmo domínio.

A proposta de Bräscher (2014) procura enfrentar essa dificuldade por meio da restrição e sistematização das relações sintagmáticas a partir de fundamentos linguísticos explícitos. Em vez de aceitar o rótulo genérico “associativa” como categoria residual para todo vínculo não hierárquico nem equivalente, a autora sugere ancorar a identificação de relações em teorias consolidadas, como a teoria da valência e a gramática de casos, articuladas a uma abordagem composicional de unidades nominais típicas das linguagens especializadas (estruturas do tipo nome+adjetivo e nome+preposição+nome). Nessa perspectiva, sintagmas deixam de ser tratados como blocos opacos e passam a ser analisados como predicados com valência própria (classificação, notificação, controle) que selecionam argumentos dotados de papéis temáticos específicos (objeto, agente, alvo etc.). O resultado é a substituição de uma relação associativa indiferenciada entre, por exemplo, “indexação” e “classificação”, por uma relação tipada do tipo

processo–objeto, mais informativa do ponto de vista semântico e mais adequada para reutilização em diferentes aplicações de organização do conhecimento.

No âmbito da linguística cognitiva, a Semântica de *Frames* proposta por Fillmore (1982) oferece um enquadramento particularmente fecundo para pensar a relação entre termos, conceitos e contextos de uso na OC. Em vez de conceber significados como listas de traços descontextualizados, Fillmore parte da ideia de que unidades lexicais evocam *frames* (que são estruturas esquematizadas de experiência ou "cenários" conceituais) dentro dos quais se organizam participantes, papéis, props e relações típicas. Assim, verbos como comprar, vender, pagar e cobrar são descritos não de forma isolada, mas como diferentes pontos de acesso a um mesmo *frame* de evento comercial, no qual se articulam comprador, vendedor, bens, dinheiro e condições de troca. A partir dessa perspectiva, as categorias semânticas relevantes deixam de ser apenas rótulos abstratos (agente, paciente, instrumento etc.) e passam a ser definidas relacionalmente, em função de sua posição em um *frame* específico.

Isso é particularmente relevante para a terminologia, na medida em que termos especializados também podem ser entendidos como gatilhos de *frames* de conhecimento especializado. No âmbito da Terminologia Baseada em *Frames*, Faber e León-Aráuz (2016, p. 2) partem justamente desse pressuposto: unidades terminológicas de domínio só são plenamente compreendidas quando ancoradas em um *frame* subjacente, que explicita os conceitos envolvidos, seus papéis funcionais e as relações que os interligam. Nessa abordagem, o significado terminológico não é um valor estático, mas um potencial de ativação de diferentes porções do *frame*, condicionadas por parâmetros de contexto (escopo local/global, dimensões sintáticas, semânticas e pragmáticas, perfis de usuário, finalidade comunicativa, entre outros). A definição terminológica passa, assim, a ser concebida como uma janela perspectivada sobre o *frame*, apta a prever extensões de sentido em termos compostos, colocações e construções frequentes no discurso especializado.

Faber e León-Aráuz (2016, p. 1-5) mostram que esse movimento "dos termos aos *frames*" pode ser operacionalizado por meio de procedimentos graduais de expansão de contexto: parte-se de unidades monoléxicas, passa-se à análise de termos complexos e sintagmas recorrentes, e destes às estruturas proposicionais que permitem identificar o *frame* de base e seus elementos. Nesse processo,

“*multi-word terms*” (termos pluriverbais) que denotam processos são modelados como predicados com valência própria, cujos argumentos correspondem a entidades, participantes ou parâmetros processuais especificados em termos de papéis de *frame*. O resultado é uma representação na qual a rede terminológica deixa de ser apenas um grafo de relações paradigmáticas (hiponímia, meronímia, equivalência) para incorporar, de modo sistemático, padrões sintagmáticos e relações estruturais derivadas de usos textuais do domínio. Tal proposta dialoga diretamente com as preocupações de Bräscher (2014) quanto à necessidade de restringir e tipificar relações associativas: a Semântica de *Frames* fornece justamente um repertório de estruturas e papéis que podem servir de base para refinar, em SOC, aquilo que hoje aparece sob o rótulo genérico de “associativa”.

Ao integrar a Semântica de *Frames* à discussão sobre terminologia e SOC, abre-se a possibilidade de conceber dicionários terminológicos e tesauros não apenas como listas de termos acompanhados de definições, mas como modelos “framificados” de domínios especializados, em que cada entrada terminológica se insere explicitamente em uma ou mais estruturas de *frame*. Isso implica, por um lado, explicitar para cada termo os elementos de *frame* que o circundam (entidades correlatas, processos em que participa, papéis que desempenha, causas, resultados, instrumentos, normas pertinentes etc.) e, por outro, permitir que as relações entre termos sejam descritas como relações entre elementos de *frames*, em vez de permanecerem como meras associações não tipadas. Em termos de modelagem, essa mudança aproxima as práticas da OC de desenvolvimentos contemporâneos em ontologias linguístico-conceituais e em recursos como FrameNet, nos quais a unidade básica de organização não é o termo isolado, mas o *frame* enquanto matriz de papéis e relações.

No contexto específico desta dissertação, a adoção de uma perspectiva orientada por *frames* se traduz na decisão de focalizar analiticamente um único termo do domínio da OC: indexação. Em vez de reconstruir a árvore de domínio completa do “Diccionario de Organización del Conocimiento” ou de mapear exaustivamente as relações entre todos os seus verbetes, opta-se por aprofundar, a partir de evidência textual e teórica, o *frame* de indexação como unidade nuclear. Isso supõe identificar: (a) quais participantes e entidades são mobilizados quando o termo é definido e utilizado (documentos, índices, termos, conceitos, sistemas, agentes indexadores, usuários, normas etc.); (b) quais processos e subprocessos

se articulam nesse *frame* (análise de assunto, seleção de termos, atribuição de descritores, validação, revisão); e (c) quais relações funcionais ligam indexação a outros conceitos da OC (por exemplo, indexação→recuperação da informação, indexação→vocabulários controlados, indexação→representação temática). Ao fazê-lo, busca-se demonstrar como a Semântica de *Frames* pode servir de eixo unificador entre terminologia, relações associativas em SOC e Análise de Domínio, oferecendo um modelo mais robusto para a descrição conceitual de um termo central para a área.

2 PROBLEMA E JUSTIFICATIVA

A Organização do Conhecimento (OC) é um campo interdisciplinar responsável por desenvolver metodologias e instrumentos para a representação e recuperação sistemática da informação. Nesse processo, a terminologia emerge como elemento central, ao prover os conceitos e os termos técnicos que sustentam as estruturas classificatórias e semânticas utilizadas em vocabulários controlados, taxonomias e tesauros. A OC depende, assim, da estabilidade conceitual oferecida por sistemas terminológicos robustos, capazes de articular as múltiplas dimensões do conhecimento científico e técnico.

Entre os esforços voltados à sistematização terminológica da OC, destaca-se o “Diccionario de Organización del Conocimiento”, publicado em 2015 sob a coordenação de Mario Barité. A obra reúne um acervo valioso de definições e articulações conceituais relativas às principais categorias e processos da área, tendo se consolidado como referência teórica e prática para pesquisadores e profissionais da Ciência da Informação (CI) e da Biblioteconomia. Além disso, uma árvore de domínio foi construída e publicada a partir desse dicionário, permitindo visualizar as principais ramificações temáticas do campo, atualmente em processo de atualização por Barité, Laipelt, Petroccelli e Gabriel Jr. (2025), no âmbito do projeto “Rumo a um Tesauro-Dicionário da Organização do Conhecimento”.

Contudo, embora esse mapeamento estrutural represente um avanço significativo, a conversão efetiva do “Diccionario” em um tesauro-dicionário depende não apenas do refinamento hierárquico ou da normalização das relações de equivalência, mas também da explicitação e tipificação das relações associativas entre os conceitos do domínio. As relações hierárquicas e de equivalência são, nesta pesquisa, tomadas como contexto descritivo; o foco analítico recai sobre as relações associativas, cuja ausência de sistematização compromete a interoperabilidade com outros sistemas terminológicos e limita a aplicação do vocabulário em ambientes de indexação e recuperação automatizada da informação.

No âmbito da CI, o termo indexação designa o processo de análise e representação temática de documentos com vistas à sua recuperação posterior. A indexação envolve a identificação dos assuntos ou características relevantes de um documento e sua tradução para termos de uma linguagem de indexação, de modo a

permitir o armazenamento, a busca e a recuperação eficientes da informação. Lancaster (2004) trata a indexação como um componente decisivo dos sistemas de recuperação da informação, ao passo que Soergel (1974) enfatiza o papel das linguagens documentárias e dos tesauros na garantia de consistência terminológica e navegabilidade conceitual.

A centralidade da indexação para a OC decorre do fato de que ela articula, de maneira exemplar, leitura documentária, análise conceitual, seleção terminológica e mediação entre documentos e usuários. Fujita (2003) ressalta que a identificação de conceitos no processo de análise de assunto exige operações cognitivas específicas, nas quais o indexador interpreta o conteúdo documental e o reconstrói em forma de representação temática. Desse modo, a indexação não pode ser reduzida à mera atribuição de palavras-chave, mas deve ser compreendida como um processo complexo de representação, no qual se cruzam conceitos, instrumentos, normas, métodos, finalidades e produtos da organização da informação.

É precisamente essa densidade conceitual que justifica a escolha de indexação como termo focal desta dissertação. Em vez de distribuir o esforço analítico por todo o conjunto de verbetes do “Diccionario de Organización del Conocimiento”, opta-se por concentrar a investigação em um processo nuclear do domínio, capaz de condensar múltiplas relações associativas relevantes para a OC. Quando vamos ao “Diccionario” verificar o que significa indexação lemos o seguinte:

> A operação mais significativa da análise documental, que se refere ao processo de descrição e representação do conteúdo temático de um documento, por meio de um número limitado de termos tomados diretamente do texto dos documentos (palavras-chave) ou de vocabulários controlados (descritores ou cabeçalhos de assunto). Esses termos devem ser aqueles que mais adequadamente caracterizem o documento, de modo a facilitar sua futura recuperação temática.[2]

Parte-se do pressuposto de que o exame aprofundado das associações em torno de “indexação” permitirá observar, em escala controlada, como se articulam relações entre processo, objeto, instrumento, norma, agente, finalidade, produto e

[2] “La operación más significativa del análisis documental que refiere al proceso de descripción y representación del contenido temático de un documento, mediante un número limitado de términos tomados directamente del texto de los documentos (palabras clave) o de vocabularios(…)”

fundamentação teórica, oferecendo um modelo mais preciso para a futura construção do tesauro-dicionário.

Essa delimitação também responde às transformações contemporâneas dos Sistemas de Organização do Conhecimento (SOC). Como argumenta Jian Qin (2020), a convergência entre paradigmas de OC e Representação do Conhecimento em Inteligência Artificial (IA) tem deslocado o interesse de vocabulários apenas normalizadores para estruturas conceituais mais formais, capazes de sustentar relações explicitadas, interoperabilidade semântica e, em certos casos, inferência. Nessa perspectiva, o refinamento das relações associativas em torno de indexação não representa apenas um ganho descritivo para o "Diccionario", mas também uma aproximação com modelos contemporâneos de representação do conhecimento, em que ontologias, *linked data* e recursos de IA operam sobre conceitos e relações de modo mais estruturado.

A crescente demanda por SOC interoperáveis, em consonância com padrões como a ISO 25964, torna urgente o desenvolvimento de vocabulários que não apenas definam termos, mas também estabeleçam redes semânticas entre eles. Essa necessidade se torna ainda mais evidente diante da intensificação das trocas científicas internacionais, do avanço das tecnologias de informação e da multiplicação de domínios híbridos, que exigem uma terminologia cada vez mais articulada e precisa. Como afirma Cabré (2013)[3], a expansão do conhecimento científico implica o surgimento contínuo de novos termos e sentidos, exigindo que os instrumentos terminológicos estejam atentos não apenas à atualização das definições, mas também à reestruturação de suas relações internas.

A riqueza conceitual pode ser amplificada mediante um conjunto explícito de relações associativas, estruturadas segundo padrões normativos e sustentadas por evidência textual. Assim, o problema que orienta esta dissertação pode ser formulado nos seguintes termos: como identificar, tipificar e sistematizar as relações associativas que envolvem o conceito de indexação em conformidade com os princípios da OC, da terminologia aplicada e das atuais demandas de representação semântica?

---

[3] KRIEGER, M. da G.; SANTIAGO, M. S.; CABRÉ, M. T. Terminologia em foco: uma entrevista comentada com Maria Teresa Cabré. **Calidoscópio**, *[S. l.]*, v. 11, n. 3, p. 328–332, 2013. Disponível em: https://revistas.unisinos.br/index.php/calidoscopio/article/view/cld.2013.113.11. Acesso em: 22 abr. 2025.

A relevância desta investigação decorre de sua inserção no processo em curso de transformação do dicionário de Barité em um tesauro-dicionário, ao mesmo tempo em que responde a uma necessidade metodológica própria da área. Para enfrentar esse problema, serão examinados artigos científicos sobre indexação recuperados em uma base especializada, a *Web of Science*. O recorte definitivo do *corpus* trabalhará com um conjunto delimitado de textos selecionados totalizando 56 artigos, conforme critérios de pertinência temática, recorrência conceitual e potencial relacional a serem definidos no capítulo metodológico.

A partir desse *corpus*, o modelo proposto contempla: a extração de evidências textuais em títulos, resumos, palavras-chave, definições e trechos analíticos; a identificação de coocorrências conceitualmente relevantes em torno de indexação; e a codificação dessas associações segundo uma tipologia funcional de relações, como <agentOf>, <instrumentFor>, <isPropertyOf>, <productOf>, <isConditionFor>. A justificativa crítica dessas relações será conduzida à luz da Análise de Domínio de Hjørland, enquanto as diretrizes da ISO 25964 serão mobilizadas no que concerne especificamente às relações associativas e à sua futura implementação em um tesauro-dicionário.

A sistematização dessas relações associativas beneficiará a organização interna do vocabulário e sua integração a ambientes de busca, bases de dados e sistemas informacionais complexos, favorecendo a navegação orientada do usuário, reduzindo ambiguidades na indexação e ampliando a precisão na recuperação. Como resultado, esta pesquisa pretende oferecer uma contribuição teórico-metodológica focal à OC e à terminologia aplicada, ao desenvolver um modelo de mapeamento de relações associativas centrado no termo indexação, ancorado na Análise de Domínio, informado pelos princípios pertinentes da ISO 25964 e sensível ao diálogo contemporâneo entre OC e IA.

Por fim, a etapa de identificação, organização e exploração inicial dessas associações contará com o suporte do Notebook LM, aqui mobilizado como agente de IA para auxílio analítico na leitura, comparação e estruturação de evidências textuais. A pertinência desse apoio reside no fato de que ferramentas baseadas em IA são especialmente eficazes na síntese de grandes volumes de literatura, na identificação de padrões relacionais recorrentes e na organização preliminar de *frames* conceituais, o que as torna particularmente adequadas para uma pesquisa

que pretende tipificar relações associativas em torno de um conceito-processo central como indexação.

2.1 Objetivo geral

Analisar e sistematizar, em perspectiva relacional e orientada por *frames*, as relações associativas que envolvem o conceito de indexação no domínio da OC, com base na literatura especializada, visando à proposição de um modelo semântico aplicável por uma IA e também aplicável à futura construção de um tesauro-dicionário interoperável.

2.2 Objetivos específicos

- Delimitar o *corpus* da pesquisa a partir do verbete "*indexing*" e de textos especializados recuperados na base de dados *Web of Science*, definindo critérios de inclusão e exclusão para a análise relacional.

- Identificar, nos textos do *corpus*, evidências semânticas e contextuais que indiquem relações associativas entre indexação.

- Classificar e codificar as relações identificadas em torno de indexação segundo uma tipologia funcional de relações associativas, composta, entre outras, por makesUseOf (dependência instrumental), actsUpon (processo→objeto), isControlledBy (dependência normativa), hasGoalOrProcess (finalidade comum), agentOf (agente→processo); isPropertyOf (propriedade→entidade/processo); measures (indicador→aspecto da indexação).

- Interpretar criticamente essas relações à luz da Análise de Domínio e das discussões contemporâneas sobre relações paradigmáticas e sintagmáticas em Sistemas de Organização do Conhecimento (SOC), verificando sua pertinência conceitual, sua coerência terminológica e seu potencial de reutilização em estruturas mais formais de representação.

- Elaborar quadros analíticos do termo indexação, contendo o rótulo relacional, o conceito relacionado, a evidência textual correspondente e a justificativa crítica de cada associação, com vistas à futura incorporação dessas relações em um tesauro-dicionário da Organização do Conhecimento.

3 REFERENCIAL TEÓRICO

3.1 Organização do Conhecimento

A Organização do Conhecimento (OC) constitui um campo teórico e aplicado voltado ao estudo dos princípios, métodos e instrumentos por meio dos quais conceitos, categorias, termos e relações são estruturados com vistas à representação, mediação, recuperação e uso da informação em contextos específicos (Hjørland, 2008, 2016). Em sentido estrito, a OC envolve atividades como descrição documental, indexação, classificação, catalogação e construção de linguagens documentárias realizadas em bibliotecas, arquivos, bases bibliográficas e outras instituições de memória. Em sentido amplo, refere-se também ao estudo de como o conhecimento é socialmente organizado, estabilizado e institucionalizado em disciplinas, domínios, práticas discursivas e sistemas de representação. Essa duplicidade é fundamental porque mostra que a OC não se reduz a procedimentos técnicos de ordenação documental, mas inclui uma reflexão epistemológica sobre as bases conceituais e sociais que sustentam os sistemas de organização (Hjørland, 2008).

A delimitação conceitual do campo também passa pela disputa terminológica em torno de expressões como “*knowledge organization*”, “*information organization*”, “*organization of information*” e “*information architecture*”. Hjørland (2012) demonstra, por meio de análise bibliométrica e contextual, que “*knowledge organization*” se consolidou como o rótulo mais característico de uma tradição específica da Ciência da Informação (CI), vinculada à *International Society for Knowledge Organization* (ISKO), à revista *Knowledge Organization* e a uma linhagem teórica marcada por autores como Dahlberg, Beghtol, Ranganathan, Soergel e o próprio Hjørland. Já termos como “*information organization*” e “*organization of information*” aparecem em usos mais heterogêneos, por vezes próximos, mas nem sempre equivalentes, o que reforça a conveniência teórica de adotar a OC como designação central desta pesquisa. Essa escolha não é meramente lexical, mas epistemológica, pois alinha o trabalho a uma tradição que privilegia a relação entre organização conceitual, teoria do conhecimento, Análise de Domínio (AD) e representação documental.

Sob esse enquadramento, a OC pode ser compreendida como um campo que opera simultaneamente em dois níveis. De um lado, ela constrói artefatos e

linguagens para representar o conteúdo de documentos e coleções, de outro, ela interroga os critérios, pressupostos e compromissos epistemológicos que orientam tais construções. Hjørland (2008; 2016) insiste que toda técnica de organização está ancorada em pressupostos teóricos e metateóricos, mesmo quando eles não são tematizados, e que as escolhas relativas a conceitos, classes e relações dependem de teorias sobre significado, documentos, usuários, domínios e práticas de uso. Nessa perspectiva, a OC deve ser compreendida não como uma simples tarefa de espelhamento da realidade, mas como uma atividade interpretativa, situada e socialmente condicionada.

Essa dimensão interpretativa é aprofundada por Jens-Erik Mai (2011) ao discutir a modernidade da classificação. Segundo o autor, boa parte da tradição classificatória em CI foi construída sobre uma ontologia realista e, segundo a qual os objetos possuem essências estáveis e a função dos sistemas classificatórios consiste em descobrir e representar essas essências de maneira objetiva, neutra e universal. Mai (2011, p. 711) contesta essa herança ao mostrar que a classificação envolve sempre flexibilidade interpretativa, pluralidade de perspectivas e contextos sociais de atribuição de sentido, de modo que não se pode sustentar seriamente a ideia de um "único melhor sistema" capaz de servir igualmente a todos os propósitos. Em lugar de uma "*classification-as-ontology*", baseada na pretensão de representar o mundo "como ele é", o autor propõe uma "*classification-as-epistemology*", em que os sistemas são entendidos como construções interpretativas que expressam modos historicamente situados de conhecer e organizar a experiência.

A crítica de Mai (2011) tem consequências diretas para a presente pesquisa, porque desloca o foco da busca por relações "naturais" entre conceitos para o exame de como essas relações são produzidas, legitimadas e operacionalizadas no interior de um domínio. Isso significa que, ao estudar conceitos da OC e suas conexões em um recurso terminológico, não se trata apenas de registrar uma estrutura semântica preexistente, mas de explicitar os critérios epistemológicos e funcionais que tornam certas relações mais pertinentes, mais úteis e mais coerentes do que outras. Tal deslocamento é particularmente importante quando o objeto de análise envolve relações associativas, historicamente menos estabilizadas e mais dependentes de interpretação contextual do que as relações de equivalência e de hierarquia.

No âmbito dos produtos mais característicos da OC, destacam-se os Sistemas de Organização do Conhecimento (SOC), entendidos como ferramentas semânticas utilizadas para representar conceitos e seus relacionamentos em domínios específicos. Svenonius (2000) e Zeng (2008) são particularmente importantes para compreender esses sistemas em sua dimensão funcional e estrutural. Svenonius (2000) mostra que tais sistemas devem ser entendidos como linguagens estruturadas, regidas por objetivos comunicacionais e cognitivos, ao passo que Zeng (2008) os apresenta como uma família ampla de estruturas que vai de listas simples a ontologias complexas.

Svenonius (2000) atribui às linguagens bibliográficas e documentárias cinco funções principais: encontrar (*find*), reunir ou identificar (*identify/gather*), selecionar (*select*), obter (*obtain*) e navegar (*navigate*). Esses objetivos mostram que um sistema de organização não serve apenas para rotular documentos, mas para criar condições de acesso intelectual, distinção, agrupamento e circulação semântica entre itens, assuntos e classes. A função de navegação, em particular, é central para esta dissertação, pois pressupõe a existência de relações suficientemente claras e inteligíveis entre conceitos, permitindo ao usuário mover-se semanticamente de um termo a outro. Svenonius (2000) também distingue entre a linguagens das obras (*work languages*), linguagens dos documentos (*document languages*) e linguagens dos assuntos (*subject languages*), reforçando que os sistemas de organização são construções linguísticas orientadas por diferentes planos de representação e não meros inventários de palavras.

Essa compreensão se articula de modo produtivo com a definição de Zeng (2008), para quem o SOC é um termo-guarda-chuva que abrange listas, glossários, dicionários, arquivos de autoridade, taxonomias, classificações, tesauros, redes semânticas e ontologias. A autora mostra que esses sistemas podem ser agrupados segundo a complexidade de suas estruturas e segundo suas funções principais, que incluem eliminar ambiguidade, controlar sinônimos ou equivalentes, estabelecer relações semânticas explícitas e apresentar propriedades conceituais. Assim, um SOC não é apenas um repertório lexical controlado, mas um modelo da estrutura semântica subjacente de um domínio, funcionando como mapa conceitual para indexadores, usuários humanos e agentes computacionais.

Se Svenonius (2000) e Zeng (2008) fornecem a base funcional e estrutural dos SOC, Tennis (2012, p. 27) contribui para o aprofundamento de sua dimensão

epistemológica. Ao propor o conceito de "*forms of knowing*" (formas de conhecer), o autor chama atenção para o fato de que os sistemas de organização incorporam modos específicos de conhecer, e que suas estruturas são moldadas por escolhas históricas, culturais e teóricas frequentemente implícitas. Sua distinção entre significado e propósito é especialmente fecunda: um sistema pode não ser "correto" em sentido forte de correspondência com a verdade e ainda assim ser útil para os fins a que se destina. Em vez de medir os SOC apenas por um ideal de representação exata do real, Tennis (2012) propõe avaliá-los também por sua utilidade, coerência e capacidade de operar em contextos pragmáticos específicos.

Essa orientação interessa diretamente à presente pesquisa porque o problema aqui não consiste em alcançar uma "verdade definitiva" sobre os conceitos da OC, mas em construir um modelo relacional útil, justificável e aplicável à reorganização de um recurso terminológico da área. Ao tipificar relações associativas, a dissertação trabalha justamente na fronteira entre significado e propósito: precisa respeitar a coerência semântica do domínio, mas também produzir uma estrutura que seja operacionalmente valiosa para navegação, indexação, interoperabilidade e futura implementação em tesauro.

3.2 Relações semânticas e relações associativas nos SOC

No interior dos SOC, as relações semânticas constituem o eixo por meio do qual os conceitos deixam de existir como unidades isoladas e passam a compor redes de significado. Tradicionalmente, a literatura e as normas de construção de tesauros, como as ISO 2788/5964 e, mais recentemente, a ISO 25964-1, distinguem três grandes classes de relações: equivalência, hierarquia e associação. As relações de equivalência controlam a variação denominativa e reúnem sinônimos, variantes e quase-sinônimos; as relações hierárquicas organizam conceitos em vínculos de superordenação e subordinação; já as relações associativas reúnem conexões não hierárquicas e não equivalenciais entre conceitos semanticamente relacionados. Embora essa tipologia seja amplamente aceita, ela não resolve, por si só, o problema da heterogeneidade interna das relações associativas.

Holma (2005) observa que a semântica relacional em SOC depende de um sistema de referência, isto é, de uma posição epistemológica e de uma teoria de linguagem e semântica que influenciam diretamente a identificação e a definição

das relações. A autora demonstra que relações de hierarquia, sinonímia e associação não são simplesmente “dadas”, mas construídas a partir de concepções de significado, categorização e pertinência contextual. Além disso, ela distingue entre relações sintagmáticas (baseadas em coocorrência em contexto) e relações paradigmáticas (estabilizadas entre unidades de um mesmo campo semântico), indicando que a modelagem relacional de um sistema depende da forma como essas dimensões são interpretadas e articuladas. Essa contribuição é particularmente relevante porque reforça que a escolha de uma relação não pode ser puramente mecânica: requer fundamentação conceitual.

Peters e Weller (2008) aprofundam essa discussão ao argumentar que os métodos clássicos de representação do conhecimento operaram durante muito tempo com um repertório restrito e vago de relações, constituído basicamente por sinonímia, hierarquia e associações indiferenciadas. Para as autoras (2008, p. 100), o desenvolvimento de ontologias, *Web Semântica* e *folksonomias* recoloca a necessidade de reconsiderar a classificação das relações para uso prático, pois esses novos ambientes demandam maior expressividade relacional. Seu texto distingue claramente as relações paradigmáticas, fixas, acopladas aos conceitos e típicas de vocabulários controlados, das relações sintagmáticas, que emergem da coocorrência efetiva de termos em documentos, indexações e outros contextos de uso. Ambas são importantes para a recuperação da informação: as primeiras sustentam a estabilidade do sistema; as segundas oferecem indícios empíricos sobre associações relevantes que podem vir a ser conceitualmente estabilizadas.

Um dos pontos mais férteis do argumento de Peters e Weller (2008) é a observação de que as ontologias ampliam significativamente o repertório relacional disponível nos SOC. Enquanto tesauros tradicionais tendem a manter as associações em um nível genérico e reversível, ontologias possibilitam a decomposição dessas relações em vínculos semanticamente mais precisos. Essa expansão relacional não é apenas um detalhe técnico, mas um sinal de mudança paradigmática: o interesse desloca-se do controle vocabular mínimo para a representação mais rica da estrutura interna dos domínios. Para uma pesquisa que pretende contribuir com a transformação de um dicionário em tesauro-dicionário, essa discussão é decisiva, pois legitima o movimento de especificação das associações em vez de mantê-las como conexões vagas e genéricas.

É justamente nesse ponto que entram as contribuições de Green (2002) e de Bean e Green (2001), mencionadas por Peters e Weller (2008) ao discutir relações semanticamente generalizáveis em SOC. A ideia central é que as relações associativas não devem ser tratadas como um “resto” residual, mas podem ser subdivididas segundo padrões recorrentes e teoricamente justificáveis. Entre os exemplos citados no debate aparecem relações como disciplina-fenômeno, processo-instrumento, ocupação-agente, ação-produto, ação-paciente, causa-efeito, matéria-prima-produto, conceito-oposto e outras formas de conexão que reaparecem em domínios distintos. Essa perspectiva fornece o fundamento para passar do simples rótulo “termo relacionado” a uma tipologia funcional mais refinada, capaz de descrever com mais precisão a natureza das conexões entre conceitos.

A tipologia de papéis Qualia proposta por Maculan (2024) oferece um enquadramento particularmente fértil para compreender o funcionamento das relações associativas em sistemas de organização do conhecimento. Ao distinguir entre Qualia, Agentivo, Constitutivo, Télico e Formal, a autora mostra que não se trata apenas de nomear ligações entre conceitos, mas de explicitar que tipo de dependência está em jogo: quem age sobre quem, que partes compõem um todo, quais funções ou objetivos orientam determinado ente, e em que categoria mais ampla ele se insere. No caso das relações associativas causativas, o papel Agentivo evidencia a dimensão processual e dinâmica dos vínculos, aproximando a semântica dos tesauros de descrições típicas da teoria de *frames*.

As relações marcadas como de papel Constitutivo expressam, por sua vez, interdependências entre todo e partes, entre recursos e produtos ou entre entidades e seus atributos, dialogando com a tradição da meronímia e das relações partitivas. Já o papel Télico aparece associado a vínculos que ligam objetos, entidades ou processos às suas funções, atividades ou finalidades típicas, tornando visível que certas associações são, na verdade, relações de uso ou de objetivo (como ser prática para, processo para, usado em, utilizado como). Ao identificar o papel Qualia de cada relação, torna-se possível associar a ela um conjunto coerente de propriedades, o que, como argumenta Maculan (2024), auxilia diretamente na elaboração de definições mais precisas e na aproximação entre a semântica tradicional de tesauros e modelos mais formais de representação do conhecimento.

3.3 Semântica de *Frames*

A evolução dos estudos sobre a categorização e a representação do conhecimento evidenciou as limitações dos modelos formalistas clássicos. Tradicionalmente, as teorias da área apoiavam-se na análise componencial sêmica, que propunha uma lista de componentes semânticos binários necessários e suficientes para definir o significado das palavras e delimitar rigidamente as fronteiras. Paralelamente, o cognitivismo clássico sustentava a existência de um "conhecimento dicionarístico" essencial e puramente linguístico, segregado do conhecimento enciclopédico ou de mundo (Krebs; Laipelt, 2018).

Segundo Krebs e Laipelt (2018), a Linguística Cognitiva desafia esse paradigma ao defender a não-modularidade da mente, pressupondo que o significado linguístico não reflete uma relação direta e abstrata entre palavra e mundo, mas é integralmente mediado pelos processos da cognição humana e pelas experiências corporificadas e socioculturais do falante. Sob essa perspectiva, o conhecimento dicionarístico passa a ser compreendido como uma subparte do conhecimento enciclopédico. É nesse cenário de transição que emergem os *frames* como estruturas fundamentais de representação do conhecimento.

Proposta originalmente por Fillmore em 1975, a Semântica de *Frames* define o *frame* como uma estrutura esquemática de conhecimento estabelecida na memória permanente, fruto da consolidação da experiência diária e das interações humanas com o mundo. Formalmente, trata-se de um sistema de conceitos interrelacionados de tal forma que, para compreender qualquer um de seus componentes individuais, torna-se obrigatório apreender a estrutura como um todo. Quando um elemento desse sistema é ativado em um texto ou conversação, todos os demais papéis e entidades associados àquela cena conceitual tornam-se automaticamente disponíveis na cognição do receptor. Diferenciando-se de abordagens baseadas em condições de verdade, Fillmore propõe uma semântica da compreensão voltada para a riqueza contextual e a intencionalidade dos participantes em instituições sociais e culturais.

Em sentido amplo, o *frame* pode ser entendido como uma estrutura de conhecimento que representa uma situação, processo, evento ou cena conceitual recorrente, organizada em torno de elementos típicos, participantes, papéis, propriedades e relações. Diferentemente de uma simples lista de termos ou de uma

malha de relações genéricas, o *frame* procura explicitar a configuração interna de um conceito, mostrando quais entidades tendem a estar implicadas quando esse conceito é ativado em práticas discursivas, cognitivas ou documentárias. Essa característica torna os *frames* especialmente adequados para lidar com conceitos-processo, como indexação, cujo significado depende de operações, agentes, instrumentos, objetos, normas, finalidades e produtos.

Aprofundando a caracterização psicológica e estrutural dessas estruturas, Barsalou (1992) demonstra que, ao contrário das "listas planas" de características, os *frames* operam em múltiplos níveis de análise por meio de conjuntos de atributos e valores e conexões relacionais dinâmicas. Dessa maneira, o autor defende que os *frames* não são estruturas rígidas, mas configurações dinâmicas e flexíveis, de modo que diferentes subconjuntos de atributos e valores são ativados probabilisticamente a depender do exemplar concreto e do contexto circundante (Barsalou, 1992).

Vemos em Jian Qin (2020) que os *frames* aparecem como um dos paradigmas centrais da Representação do Conhecimento (RC) em IA, ao lado de regras de produção, redes semânticas e lógica de primeira ordem. Nessa tradição, *frames* são estruturas compostas por atributos ou *slots* que conectam entidades e propriedades, permitindo organizar conhecimento de maneira suficientemente expressiva para apoiar inferência, decisão e solução de problemas. Ainda que essa formulação venha do campo da Inteligência Artificial (IA), ela oferece uma ponte fértil com a OC, na medida em que permite transformar associações genéricas em papéis relacionais mais precisos e teoricamente controláveis. Assim, em vez de apenas dizer que dois conceitos “estão relacionados”, uma abordagem orientada por *frames* permite explicitar como estão relacionados.

Essa contribuição é decisiva para esta dissertação, porque o objeto aqui investigado não é um conceito estático, mas um conceito-processo central da OC. No caso da indexação, por exemplo, um *frame* permite reunir, em torno do núcleo conceitual, elementos como: quem realiza a ação; sobre o que ela incide; quais instrumentos ou linguagens mobiliza; por quais normas é regulada; que produtos gera; a que finalidade serve; e em que tradição teórica se ancora. Cada um desses elementos pode ser convertido em papéis relacionais relativamente estáveis, como, por exemplo, *actsUpon, makesUseOf, isControlledBy, resultsIn* e *isBasedOn*.

Os *frames* ajudam a mediar a passagem entre relações sintagmáticas e relações paradigmáticas. Uma abordagem baseada em *frames* permite tomar regularidades sintagmáticas observadas em *corpus* (por exemplo, a recorrrente associação entre indexação e termos como “linguagem documentária”, “documento”, “análise de assunto”, “recuperação da informação”, “política de indexação” e “usuário”) e reorganizá-las como papéis conceituais sistematizados. Essa passagem é coerente tanto com a AD quanto com o movimento contemporâneo de aproximação entre OC e RC.

Além disso, a noção de *frame* é particularmente compatível com uma visão não essencialista e não rigidamente classificatória da OC. Em vez de pressupor que o significado de um “*concept*” (no sentido de unidade conceitual representada em sistemas de OC) está inteiramente encerrado em uma definição isolada, a abordagem por *frames* reconhece que os conceitos ganham inteligibilidade por meio das estruturas de situação e das redes de relações em que se inserem. Isso dialoga fortemente com Mai (2011), ao afastar a ideia de que a representação conceitual possa ser completamente esgotada por categorias estáveis, neutras e universais. Também dialoga com Tennis (2012), ao permitir pensar sistemas não apenas em termos de correção abstrata, mas em termos de utilidade estrutural e adequação a propósitos de navegação, recuperação e interoperabilidade.

3.4 Organização do Conhecimento, Inteligência Artificial e convergência com a Representação do Conhecimento

As transformações recentes nos ambientes digitais tornaram ainda mais visível a necessidade de estruturas relacionais densas e interoperáveis. Jian Qin (2020) mostra que a OC, tal como desenvolvida na CI, e a Representação do Conhecimento (RC) (*Knowledge Representation*) em IA seguiram trajetórias historicamente distintas, mas vêm convergindo de modo cada vez mais intenso. Segundo a autora, a diferença clássica é que a OC foi orientada sobretudo à organização do universo conceitual e à normalização vocabular para fins de descrição e recuperação, enquanto a RC em IA foi concebida para expressar fatos, relações e condições em formatos formais sobre os quais se possa raciocinar. Ainda assim, ontologias, *linked data*, extração de conceitos e relações por métodos

computacionais e outros desenvolvimentos recentes aproximaram significativamente esses dois paradigmas.

Qin (2020) também retoma Hjørland (2008) para reforçar que a OC, em sentido estrito, envolve atividades como descrição documental, indexação e classificação em instituições de memória, e, em sentido amplo, diz respeito à forma como o conhecimento é socialmente organizado e refletido em SOC. A autora observa que, no contexto da explosão de dados digitais, os “velhos truques” da OC continuam necessários, mas já não bastam para responder à escala, à velocidade e à complexidade informacional contemporâneas. É nesse cenário que ontologias e modelos relacionados ganham relevo, por combinarem elementos tradicionais da OC (controle terminológico, classificação, relações conceituais) com recursos formais típicos da RC, como codificação explícita, declaratividade e potencial inferencial.

Essa convergência é especialmente importante porque mostra que o refinamento das relações associativas não interessa apenas à construção de tesauros mais elegantes, mas à possibilidade de tornar um recurso terminológico mais interoperável, mais computacionalmente tratável e mais alinhado às exigências da *Web Semântica* e de aplicações de IA. Quando um sistema passa a explicitar relações do tipo processo-instrumento, processo-objeto, ação-produto ou fundamentação teórica, ele se aproxima de formas de representação mais úteis tanto para navegação humana quanto para exploração automatizada. Em outras palavras, o debate contemporâneo não opõe OC e IA, antes, sugere que a OC pode se fortalecer ao incorporar, criticamente, modelos relacionais mais explícitos.

3.5 Ferramentas de IA generativa e desafios de confiabilidade

Ao longo da construção desta dissertação, diferentes ferramentas de IA generativa foram mobilizadas como apoio à leitura, à síntese e à organização do material teórico e empírico. Entre essas ferramentas, destacam-se sobretudo modelos de linguagem de uso geral, acessados em plataformas comerciais, bem como soluções voltadas à análise de conjuntos documentais fornecidos pela pesquisadora. Embora esses recursos tenham se mostrado úteis em tarefas pontuais de reescrita, tradução e geração de rascunhos, sua utilização evidenciou de forma recorrente o problema das chamadas “alucinações”, isto é, a produção de

respostas aparentemente plausíveis, mas desconectadas das fontes efetivamente fornecidas ou mesmo factualmente incorretas.

Em uma etapa inicial do trabalho, foram testados, em versões pagas, modelos como Gemini e ChatGPT, com o objetivo de apoiar a análise focada de artigos, capítulos de livros e documentos que compõem o *corpus* teórico-metodológico da pesquisa. A experiência prática mostrou que, apesar de sua capacidade de síntese linguística, esses sistemas apresentaram importantes limitações quando se tratava de ancorar rigorosamente as respostas nos arquivos enviados. Em diversos casos, as ferramentas extrapolavam o conteúdo disponível, inferindo argumentos não sustentados ou atribuindo a autores formulações que não constavam dos textos analisados. Essa tendência à projeção e à "costura criativa" dos dados tornava arriscado o uso direto desses resultados em um trabalho que exige rastreabilidade e fidelidade às fontes.

Diante dessas limitações, a pesquisadora passou a utilizar a plataforma Perplexity Pro, que combina diferentes modelos de base, incluindo Gemini, ChatGPT, Sonar 2 e Claude Sonnet, e oferece mecanismos mais robustos de vinculação entre respostas e documentos de referência. Essa mudança permitiu realizar análises mais finas e controladas, sobretudo no que diz respeito à leitura cruzada de artigos, à extração de trechos relevantes e à comparação de formulações conceituais. Ainda assim, observou-se que, para análises mais extensas e sofisticadas de *corpora*, a plataforma exige níveis de assinatura mais avançados (contas Max), o que evidencia a dimensão material e econômica envolvida no acesso a ferramentas de IA com maior capacidade de processamento e contexto.

Paralelamente, foi testado o Google NotebookLM, ferramenta gratuita que permite a criação de cadernos temáticos com múltiplas fontes anexadas e interação conversacional orientada por esse material. O plano gratuito mostrou-se bastante generoso, possibilitando a criação de até 100 cadernos, o que foi suficiente para os usos exploratórios realizados no âmbito desta pesquisa. Ainda que o Google ofereça um plano pago com limites expandidos, a versão gratuita atendeu adequadamente ao objetivo de apoiar a organização de leituras, a geração de resumos intermediários e a formulação de questões de pesquisa derivadas de conjuntos textuais específicos.

Essas experiências acumuladas reforçam dois pontos centrais para a discussão metodológica contemporânea. Em primeiro lugar, evidenciam que ferramentas de IA generativa podem ser valiosas como apoio instrumental à pesquisa, desde que utilizadas de modo crítico e reflexivo, com verificação sistemática das informações produzidas e cotejo constante com as fontes originais. Em segundo lugar, mostram que a presença de “alucinações” e de extrapolações não fundamentadas não é um detalhe marginal, mas um aspecto estrutural desses sistemas, que não podem ser tratados como substitutos da leitura analítica, da interpretação teórica e do julgamento metodológico do pesquisador. Em uma dissertação ancorada em análise conceitual e em modelagem de relações associativas, essa constatação reforça a necessidade de manter a IA no lugar de ferramenta auxiliar, articulada a uma postura epistemológica vigilante e a procedimentos de validação que garantam a confiabilidade dos resultados.

4 METODOLOGIA

Segundo Barité (2001), os profissionais da informação têm historicamente vivenciado uma tensão constante entre o uso da linguagem natural e o emprego de vocabulários controlados. Parte significativa do trabalho bibliotecário e documentalista tem se concentrado em responder às demandas dos usuários por representações fidedignas do conteúdo informacional, ao mesmo tempo em que enfrenta limitações linguísticas, epistemológicas e metodológicas. Essa realidade evidencia a complexidade da Organização do Conhecimento (OC) e a necessidade de abordagens mais refinadas para lidar com os desafios da representação terminológica em domínios especializados.

Nesse mesmo contexto, Barité (2001) observa que os estudos em OC tendem a se articular em torno de duas grandes vertentes teóricas: as macroestruturais e as microestruturais. As primeiras buscam compreender os sistemas conceituais a partir de perspectivas amplas, vinculadas a disciplinas de base, como a Linguística Aplicada (LA) e a Terminologia, que examinam o uso social da linguagem de especialidade. Essa abordagem confere à terminologia um caráter transversal, relevante não apenas para bibliotecários, mas também para tradutores, linguistas e documentalistas. As abordagens microestruturais, por sua vez, concentram-se na linguagem natural e em sua aplicação nas interações cotidianas dos usuários com os sistemas de informação, especialmente no processo de busca, recuperação e mediação do conhecimento.

É nesse cenário que se insere a presente pesquisa, cuja metodologia fundamenta-se na Análise de Domínio (AD), conforme proposta por Hjørland (2008; 2024), adotada aqui como eixo central para a compreensão das estruturas conceituais e das relações semânticas que configuram o campo da OC. A AD permite investigar a estrutura interna de um domínio, identificando os conceitos que o compõem, suas articulações e as lógicas de organização que lhes conferem coerência. Essa abordagem é particularmente eficaz para a análise de relações semânticas, pois parte do princípio de que os significados são socialmente construídos e compartilhados dentro de comunidades epistêmicas específicas, refletindo práticas, valores e tradições intelectuais próprias de cada campo.

A presente pesquisa adota uma abordagem qualitativa de caráter exploratório, ancorada nos aportes da CI, da OC, da Teoria de *Frames* e da Terminologia Baseada em *Frames*, visando descrever e modelar o *frame* "*indexing*" e suas relações associativas tipadas em literatura especializada da área. A análise concentra-se nas relações associativas do conceito selecionado, deixando de lado, nesta etapa, as relações hierárquicas e equivalentes (reconhecidas, porém, conforme a ISO 25964-1:2011), por compreender que os vínculos associativos expressam com maior precisão as interdependências funcionais e cognitivas entre os processos da OC. A descrição e interpretação dessas relações seguem o modelo proposto por Maculan (2015; 2024), que identifica um conjunto de tipos relacionais aplicáveis a domínios especializados (como fazUsoDe, éProcessoPara, éUsadoEm, temObjetivoOuProcesso, usaProcesso e éFeitoAPartirDe).

Essas relações são interpretadas à luz da Teoria dos *Frames* (Fillmore, 1982; Barsalou, 1992; Faber; León-Araúz, 2016), entendendo-se que cada conceito evoca um quadro de conhecimento (*frame*) que estrutura as entidades, papéis e processos típicos de uma situação cognitiva. Assim, "*indexing*" é analisado como *frames* interdependentes dentro do macroprocesso de representação temática da informação. O objetivo não é apenas descrever o termo isoladamente, mas analisar as conexões semânticas que expressam suas funções e interdependências.

Como fonte principal de dados empíricos, foram utilizados artigos científicos indexados na base *Web of Science*, por se tratar de um índice multidisciplinar amplamente reconhecido, com controles rigorosos de qualidade editorial e recursos avançados para busca por campos específicos e análise de citações. A extração e a análise do *frame* "*indexing*" foram apoiadas por um modelo de linguagem de grande porte, operacionalizado por meio da ferramenta NotebookLM, que permite o carregamento de conjuntos de documentos e a formulação de *prompts* analíticos fundamentados em referenciais teóricos, devolvendo respostas com citações aos textos de origem.

4.1 Construção do *corpus* 1 - Busca por décadas

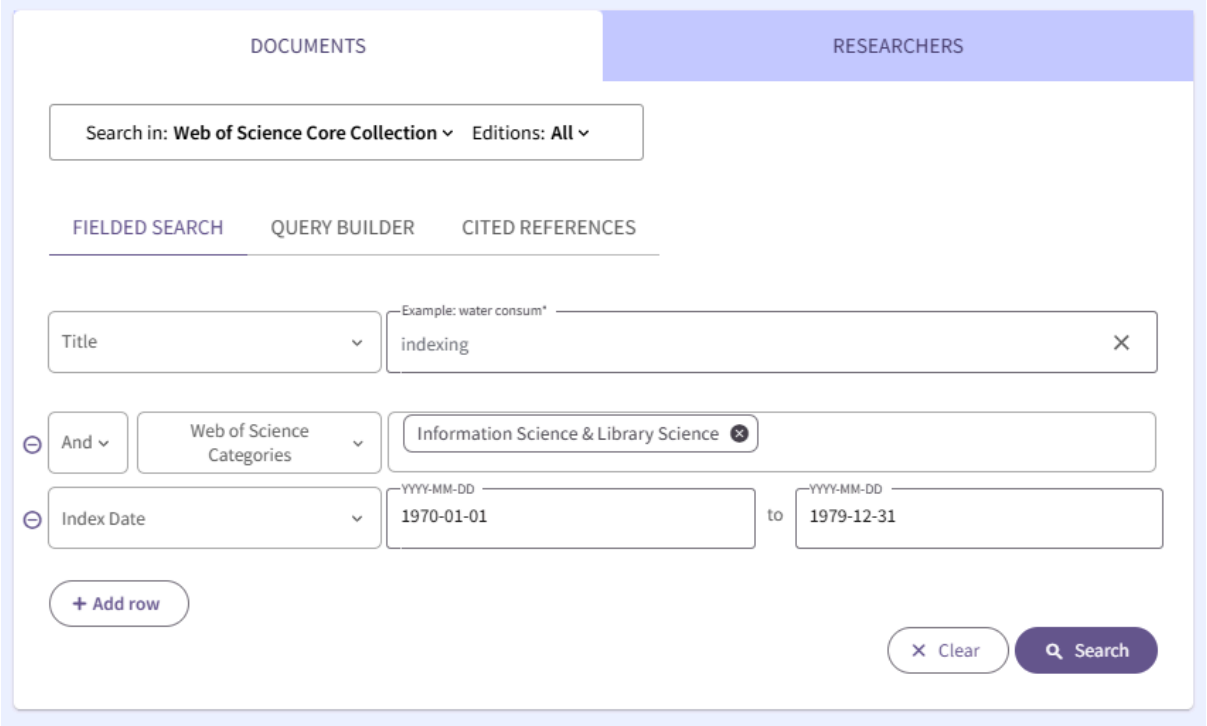
A primeira etapa consistiu na construção de um *corpus* histórico de artigos sobre indexação, distribuídos por décadas, utilizando a base de dados *Web of Science*. No campo de busca, foi selecionado o índice "*Title"*, de modo que o termo

de busca fosse recuperado especificamente no título dos artigos, restringindo o *corpus* a publicações cujo foco explícito incluísse a noção de "*indexing*". O termo de busca utilizado foi a palavra "*indexing*", escrita em inglês, sem truncamentos.

Os resultados foram refinados pela categoria "*Web of Science Categories"*, selecionando-se exclusivamente a categoria "*Information Science & Library Science"*, de forma a delimitar o *corpus* ao domínio disciplinar da CI e áreas afins de OC.

Foram realizadas buscas separadas para cada período decenal completo entre 1970 e 2025, correspondendo às décadas de 1970, 1980, 1990, 2000, 2010 e 2020. Para cada década, aplicou-se o filtro de "*Citations: highest*" correspondente, que coloca em primeiro lugar os artigos mais citados dentro do conjunto.

Imagem 1 - Tipo de seleção na busca por décadas

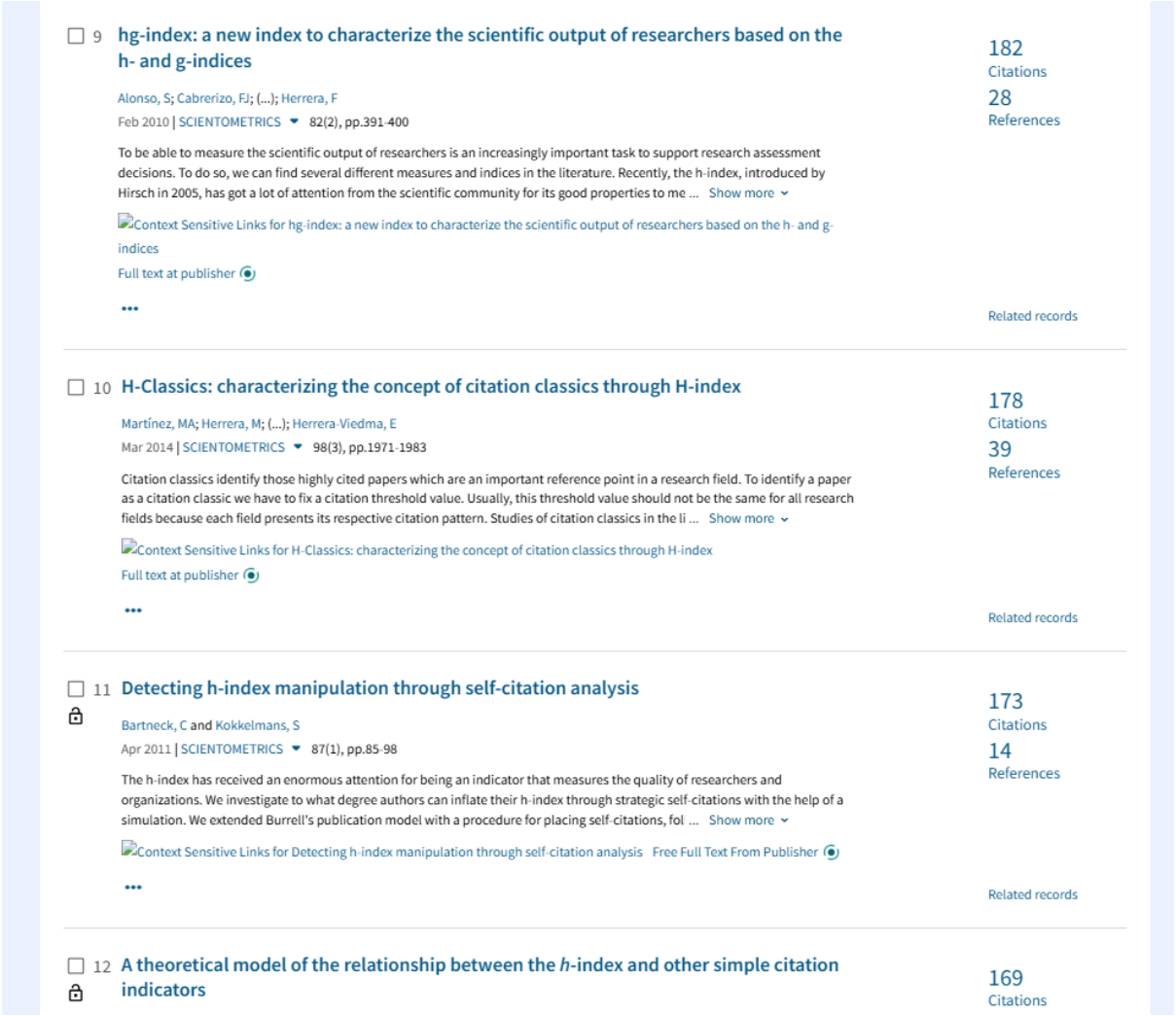


Fonte: autora

No recorte da década de 2010, observou-se que parte dos resultados recuperados pela busca por "*indexing*" no campo "*Title*" correspondia a estudos centrados no *h-index* e em indicadores bibliométricos correlatos, em que o termo "*index*" aparece de forma proeminente. Essa concentração indica uma limitação típica de buscas baseadas em termos literais no título, nas quais a correspondência

entre “*index*” e “*indexing*” nem sempre reflete, de modo preciso, o foco conceitual em processos de indexação, exigindo uma triagem manual e uma leitura analítica atenta para assegurar a pertinência dos artigos selecionados.

Imagem 2 - Artigos que não foram selecionados



Fonte: autora

Para o período a partir de 2020, considerando que o recorte temporal da pesquisa se estende até 2025, realizou-se uma busca específica para o intervalo de 2020 a 2025. Ao final dessa etapa, o *corpus* 1 totalizou 28 artigos, dos quais 5 de cada década entre 1970 e 2010 e 3 do período 2020–2025.

Por década:

Anos 70:

On the specification of term values in automatic indexing. Autores: Salton e Yang. Ano: 1973.

Index term weighting. Autor: Karen Sparck Jones. Ano: 1973.

Probabilistic models for automatic indexing. Autores: Abrahan Bookstein e Don Swanson. Ano: 1974.

A probabilistic approach to automatic keyword indexing. Autor: Stephen P. Harter. Ano: 1975.

On indexing, retrieval and the meaning of about. Autor: M.E. Maron S. Ano: 1977.

Anos 80:

Indexing Consistency, Quality And Efficiency. Autor: L. Rolling. Ano: 1981.

FASIT: A fully automatic syntactically based indexing system. Autores: Martin Dillon e Ann S. Gray. Ano: 1983.

Indexing Consistency in MEDLINE. Autores: Mark E. Funk, Carolyn Anne Reid e Leon S. McGoogan. Ano: 1983.

Knowledge Based Indexing of the Medical Literature: The Indexing Aid Project. Autor: Susanne M. Humphrey e Nancy E. Miller. Ano: 1987.

The Effectiveness of a Nonsyntactic Approach to Automatic Phrase Indexing for Document Retrieval. Autor: Joel L. Fagan. Ano: 1989.

Anos 90:

Indexing by Latent Semantic Analysis. Autores: Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer e Richard Harshman. Ano: 1990.

KeyWords Plus algorithmic derivative indexing. Autores: Eugene Garfield e Irving H. Sher. Ano: 1993.

Some Issues in the Indexing of Images. Autor: Sara Shatford Layne. Ano: 1994.

Using Latent Semantic Indexing for Literature Based Discovery. Autores: Michael D. Gordon e Susan Dumais. Ano: 1998.

Indexing and Access for Digital Libraries and the Internet: Human, Database, and Domain Factors. Autores: Marcia J. Bates. Ano: 1998.

Anos 2000:

Coauthorship Patterns and Trends in the Sciences (1980-1998): A Bibliometric Study with Implications for Database Indexing and Search Strategies. Autor: Wolfgang Glanze. Ano: 2002.

A framework for understanding Latent Semantic Indexing (LSI) performance. Autores: April Kontostathis e William M. Pottenger. Ano: 2005.

An extension of the Hirsch index: Indexing scientific topics and compounds. Autor: Michael G. Banks. Ano: 2006.

Thesaurus Based Automatic Keyphrase Indexing. Autores: Olena Medelyan e Ian H. Witten. Ano: 2006.

Domain-Independent Automatic Keyphrase Indexing with Small Training Sets. Autores: Olena Medelyan e Ian H. Witten. Ano: 2008.

Anos 2010:

A comparative study of Flickr Tags and IndexTerms in a General Image Collection. Autor: Abebe Rorissa. Ano: 2010.

Invisible institutional repositories Addressing the low indexing ratios of IRs in Google Scholar. Autores: Kenning Arlitsch e Patrick S. O'Brien. Ano: 2011.

MapReduce indexing strategies: Studying scalability and efficiency. Autores: Richard McCreadie, Craig Macdonald e Iadh Ounis. Ano: 2011 / 2012.

Coverage, field specialisation and the impact of scientific publishers indexed in the Book Citation Index. Autores: Daniel Torres-Salinas, Nicolas Robinson-Garcia, Juan Miguel Campanario e Emilio Delgado López-Cózar. Ano: 2013.

Errors in DOI indexing by bibliometric databases. Autores: Fiorenzo Franceschini, Domenico Maisano e Luca Mastrogiacomo. Ano: 2014.

Anos 2020-2025:

Subject indexing in humanities: a comparison between a local university repository and an international bibliographic service. Autores: Koraljka Golub, Jukka Tyrkko, Joacim Hansson e Ida Ahlstrom. Ano: 2020.

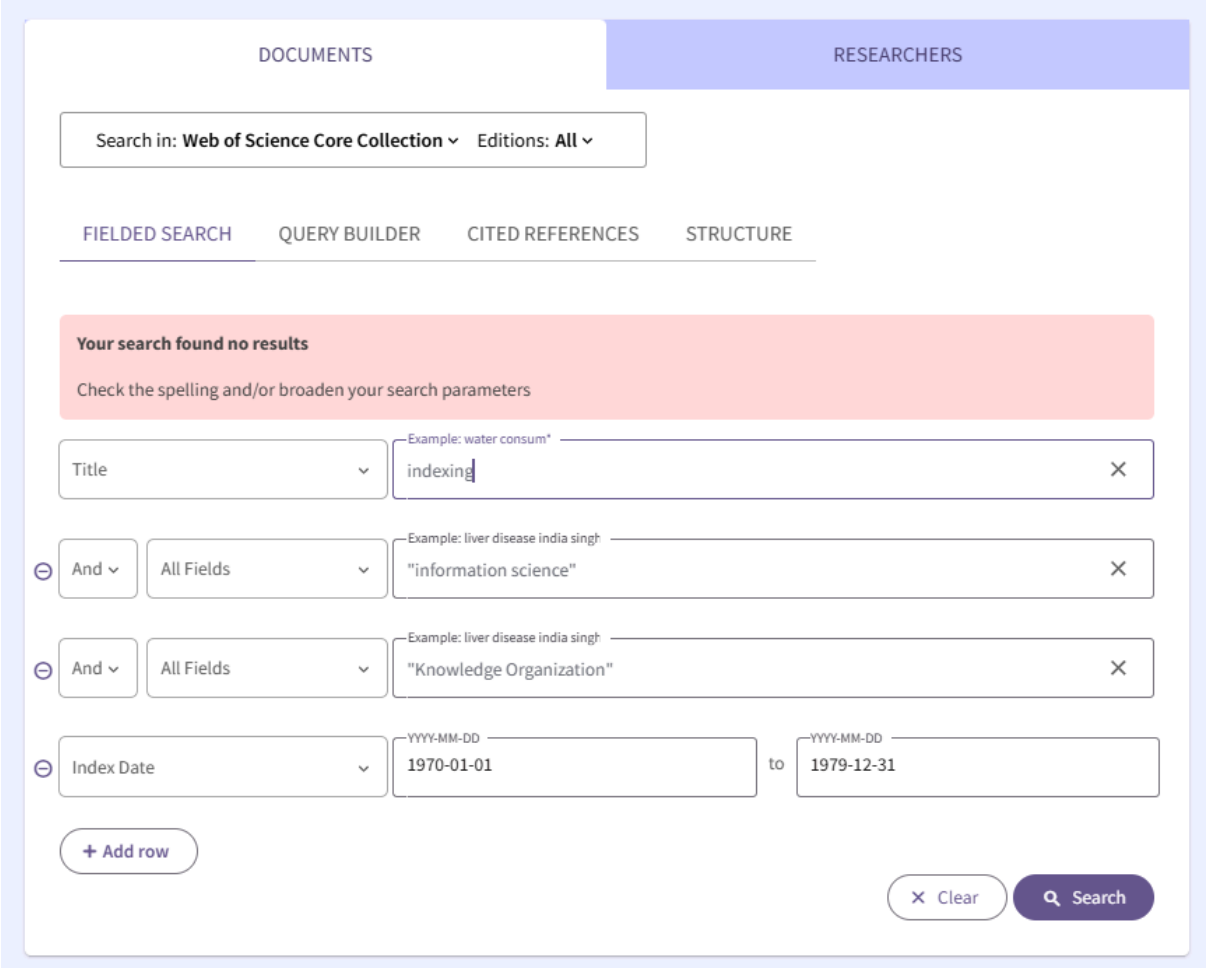
Annif and Finto AI: Developing and Implementing Automated Subject Indexing. Autores: Suominen, Osma; Inkinen, Juho; Lehtinen, Mona. Ano: 2021.

Automated Subject Indexing: An Overview. Autor: Koraljka Golub. Ano: 2021.

4.2 Construção do *corpus* 2 - Busca temática ampliada

Inicialmente, buscou-se aplicar ao *corpus* 2 uma estratégia análoga à utilizada no *corpus* 1, combinando o termo "*indexing*" no campo "*Title*" com os termos "*information science*" e "*knowledge organization*" em "*All Fields*" e recortes temporais por décadas. No entanto, essa configuração resultou, para várias décadas, em consultas sem retorno de registros, com a mensagem "Sua pesquisa não encontrou resultados. Verifique a grafia e/ou amplie seus parâmetros de busca", indicando que a combinação entre os campos e os filtros temporais era excessivamente restritiva para os objetivos desta etapa.

Imagem 3 - Falha na recuperação ao combinar recorte por décadas

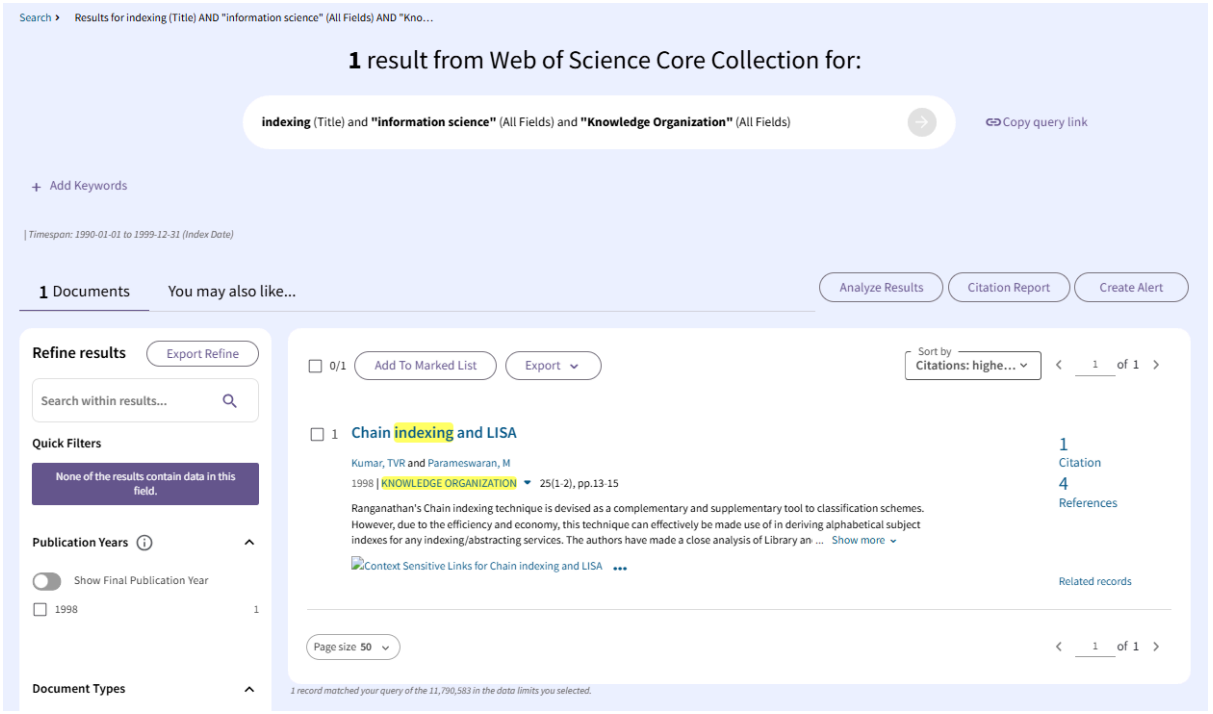


Fonte: autora

Quando se tentou manter o termo "*indexing*" apenas no campo "*Title",* associado a "*information science*" e "*knowledge organization*", mas ainda com

recortes por décadas, a recuperação foi extremamente reduzida, chegando, em certos períodos, a apenas um artigo identificado na década de 1990.

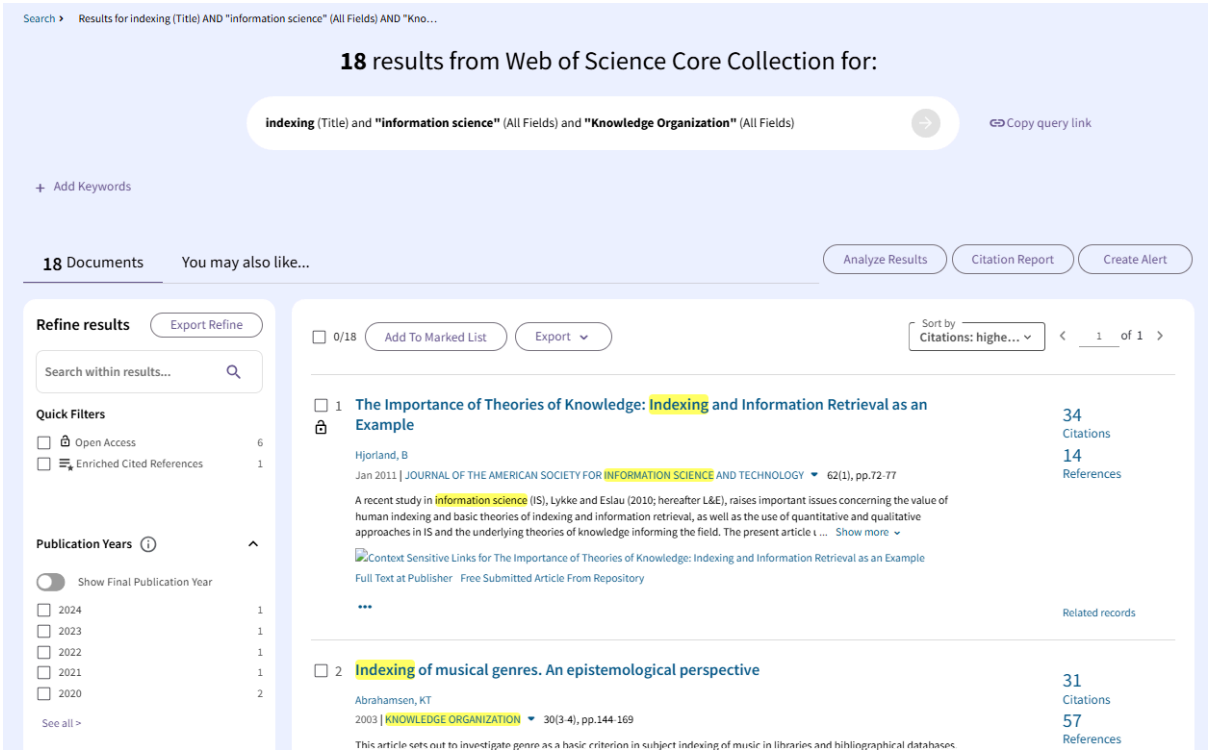
Imagem 4 - Recuperação limitada nos anos 90



Fonte: autora

Mesmo ao suprimir o filtro por décadas e buscar apenas “*indexing*" em “*Title*”, a base retornou um conjunto total de 18 registros, quantitativo insuficiente para compor um segundo *corpus* de 28 artigos que dialogasse de forma consistente com o *corpus* histórico já estabelecido.

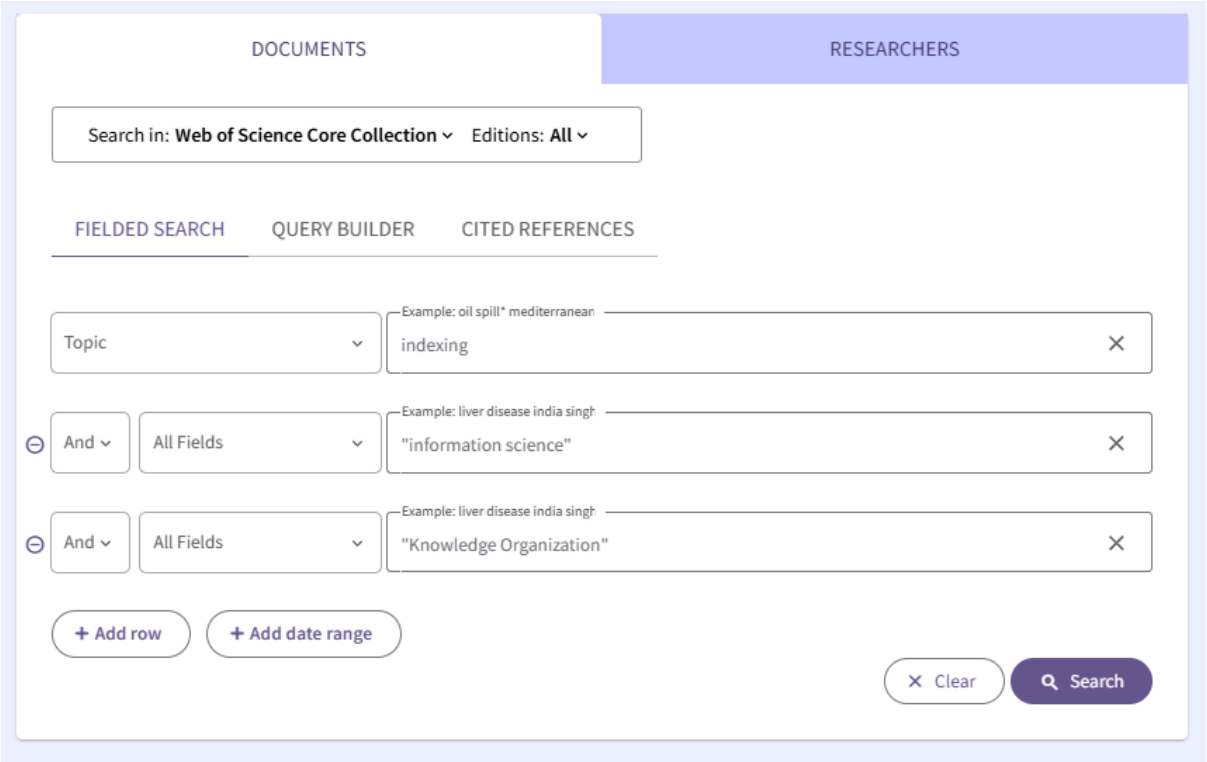
Imagem 5 - Conjunto reduzido de 18 artigos



Fonte: autora

Diante desses resultados limitados, optou-se por deslocar o termo "*indexing*" do campo "*Title*" para o campo "*Topic*", que abrange título, resumo, palavras-chave e palavras-chave dos autores, mantendo-se os termos de contexto "*information science*" e "*knowledge organization*" em "*All Fields*".

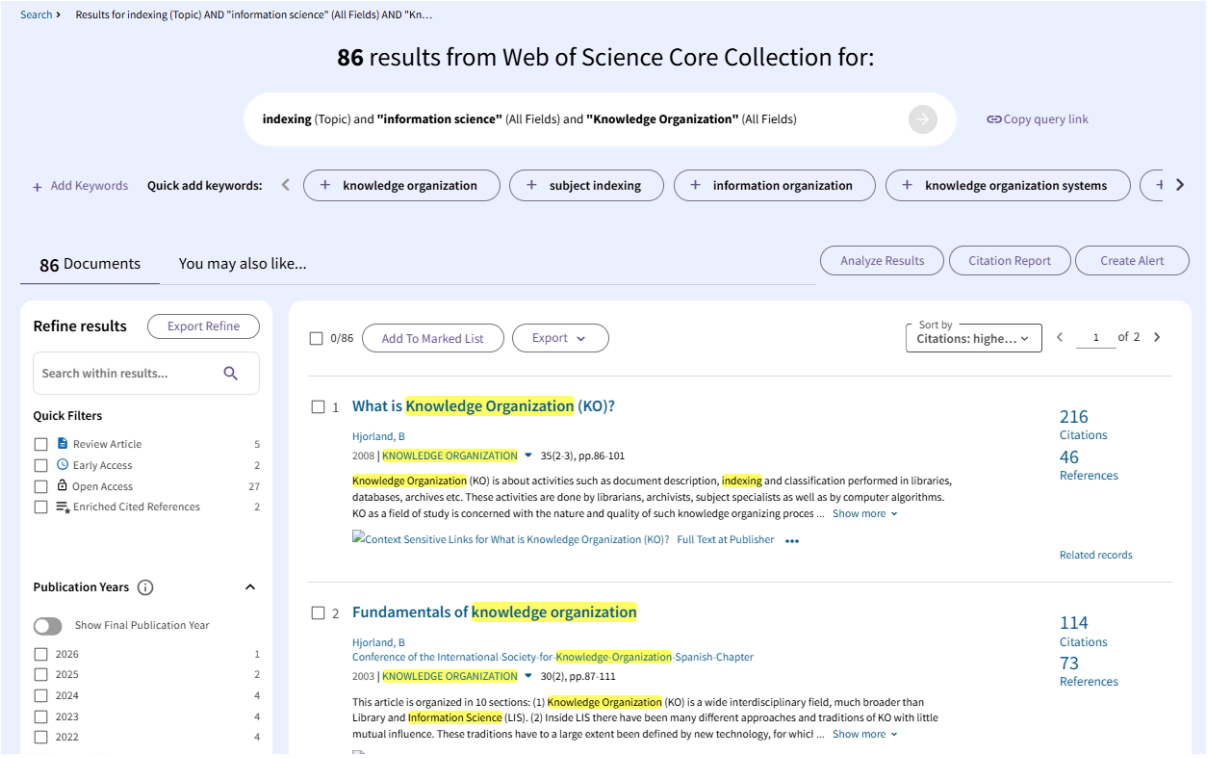
Imagem 6 - Busca ampliada



Fonte: autora

A estratégia sem recorte temporal, por outro lado, possibilitou reunir um conjunto mais robusto de 86 resultados, a partir do qual foi possível selecionar 28 artigos conceitualmente pertinentes ao problema da indexação no contexto da OC.

Imagem 7 - Resultados da busca temática ampliada



Fonte: autora

Essa simetria metodológica entre os dois *corpora*, tanto na forma de organização dos artigos em blocos quanto no uso do mesmo modelo de linguagem e do mesmo *prompt* analítico, garante condições de comparabilidade entre os resultados, permitindo integrar as extrações de *frames* e relações associativas em uma leitura conjunta da evolução do conceito de "*indexing*" no campo da OC.

Busca temática ampliada:

What is Knowledge Organization (KO)?. Autor: Birger Hjørland. Ano: 2008.

Fundamentals of Knowledge Organization. Autor: Birger Hjørland. Ano: 2003.

Reviews of concepts in knowledge organization. Subject (of Documents). Autor: Birger Hjørland. Ano: 2016

From knowledge organization to knowledge representation. Autores: Fausto Giunchiglia, Biswanath Dutta e Vincenzo Maltese. Ano: 2013.

The Importance of Theories of Knowledge: Indexing and Information Retrieval as an Example. Autor: Birger Hjørland. Ano: 2011.

Indexing of Musical Genres. An Epistemological Perspective Knut. Autor: Tore Abrahamsen. Ano: 2003.

Folksonomies and the new order. Autor: Jens-Erik Mai. Ano: 2011.

A Knowledge Based Approach to Information Extraction for Semantic Interoperability in the Archaeology Domain. Autores: Andreas Vlachidis e Douglas Tudhope. Ano: 2025.

Basic-level categories: A review. Autor: Lala Hajibayova. Ano: 2013.

Indexing: Concepts and Theory. Autor: Birger Hjørland. Ano: 2018.

Comparing social tags with subject headings on annotating books: A study comparing the information science domain in English and Chinese. Autores: Dan Wu, Daqing He, Jin Qiu, Ruonan Lin e Yang Liu. Ano: 2012.

Member Activities and Quality of Tags in a Collection of Historical Photographs in Flick. Autores: Besiki Stvilia e Corinne Jörgensen. Ano: 2010.

Knowledge organization trends in library and information studies: a preliminary comparison of the pre- and post-web eras. Autores: Kristie Saumure and Ali Shiri. Ano: 2008.

Provenance as a knowledge organization principle. Autores: Natalia Tognoli e José Augusto Chaves Guimarães. Ano: 2019.

This is Not Domain Analysis. Autor: Hanne Albrechtsen. Ano: 2015.

Extending the Simple Knowledge Organization System for Concept Management in Vocabulary Development Applications . Autores: Joseph T. Tennis e Stuart A. Sutton. Ano: 2008.

An evaluation of enhancing social tagging with a knowledge organization system. Autores: Brian Matthews, Catherine Jones, Bartłomiej Puzon, Jim Moon, Douglas Tudhope e Marianne Lykke Nielsen. Ano: 2010.

Terminology Registries for Knowledge Organization Systems: Functionality, Use, and Attributes. Koraljka Golub, Douglas Tudhope, Marcia Lei Zeng e Maja Žumer. Ano: 2014.

Unleashing the Power of Data Through Organization: Structure and Connections for Meaning, Learning and Discovery. Autor: Dagobert Soergel. Ano: 2015.

Political versus apolitical epistemologies in knowledge organization. Autor: Birger Hjørland. Ano: 2020.

Exploitation of folksonomies in subject analysis. Autores: Constantia Kakali e Christos Papatheodorou. Ano: 2010.

The Elusive Tale: Leveraging the Study of Information Seeking and Knowledge Organization to Improve Access to and Discovery of Folktales. Autores: Kathryn A. La Barre e Carol L. Tilley. Ano: 2012.

KO for feminism and feminist research: a discourse oriented study of systematic. Autor: J. Samuelsson. Ano: 2010.

Written Knowledge: A Literary Perspective on Indexing Theory. Autor: Jack Andersen. Ano: 2000

The indexing policy in the practices of brazilian institutional repositories a diagnostic study from. Autores: Mariângela Spotti Lopes Fujita, Jessica Cristina Panuto (e colaboração de Roberta Cristina Dal'Evedove Tartarotti). Ano: 2023.

The Empirical Approach in the Evaluation of Information systems. Autor: Robert Fugmann. Ano: 1999.

The Representation of Ethics and Knowledge Organization. Autores: Daniel Martínez-Ávila, José Augusto Chaves Guimarães, Fabio Assis Pinho e Melodie J. Fox. Ano: 2015

Male homosexuality in brazilian indexing languages. Autores: F. A. Pinho e J. A. Ch. Guimarães. Ano: 2012.

4.3 Extração do *frame* "*indexing*" com apoio do NotebookLM

Esta etapa metodológica consistiu na análise dos artigos do *corpus* 1 por meio da ferramenta NotebookLM, com vistas à identificação do *frame* "*indexing*" e de suas relações associativas tipadas, em diálogo com a Teoria de *Frames* e a Terminologia Baseada em *Frames*.

Os artigos selecionados foram obtidos em formato digital (PDF), organizados e carregados em um caderno (*notebook*) único dentro da ferramenta NotebookLM.

Devido a limitações operacionais o *prompt* analítico foi aperfeiçoado progressivamente, a partir da leitura prévia de um subconjunto dos artigos, até chegar a uma formulação estável adequada aos objetivos teórico-metodológicos da pesquisa. Para isso foi feita uma análise da CI, da OC, da Teoria de *Frames*, da Terminologia Baseada em *Frames* e das relações associativas tipadas. Os arquivos foram:

Frames, concepts, and conceptual fields. Lawrence W. Barsalou.

Specialized Knowledge Representation and the Parameterization of Context. Pamela Faber e Pilar León-Aráuz.

Aplicação de reengenharia de tesauro: modelagem do THESAGRO. Benildes C. M. S. Maculan, Gercina A. B. O. Lima, Ivo Pierozzi Jr., Leandro H. M. Oliveira.

Reengineering Thesauri for New Applications: the AGROVOC Example. Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer e Stephen Katz.

Para cada artigo, NotebookLM foi instruído a identificar: os elementos que compõem o *frame* "*indexing*" (agentes, objetos, instrumentos, critérios, resultados, objetivos etc.); as relações associativas relevantes entre "*indexing*" e outros conceitos, explicitando os tipos de relação, em analogia às relações tipadas propostas por Maculan (2015; 2024).

Além disso, especificamos que os rótulos das relações associativas tipadas fossem gerados em inglês, de modo a alinhar a nomenclatura relacional ao idioma predominante da literatura analisada e facilitar futuras aplicações em ambientes de OC e *Web Semântica*.

O *prompt* final utilizado orientou a ferramenta da seguinte forma:

"Com base em Ciência da Informação, Organização do Conhecimento, Teoria de *Frames* e Terminologia Baseada em *Frames*, leia os artigos em anexo. Partindo da premissa de que conceitos evocam quadros de conhecimento, extraia o *frame* "*INDEXING*" de cada um dos artigos, suas relações associativas tipadas, os termos relacionados e um excerto do texto."

Essa instrução incorpora os pressupostos de que cada conceito evoca um *frame* e de que as relações associativas tipadas podem ser descritas como relações cognitivas entre entidades, processos, instrumentos, agentes e objetivos, em consonância com a literatura de OC e Terminologia Baseada em *Frames*.

As respostas geradas pelo modelo foram revisadas criticamente pela pesquisadora, com apoio do referencial teórico, de modo a validar, ajustar ou refutar interpretações automáticas e garantir coerência terminológica e epistemológica. Ao final, os resultados obtidos com os dois *corpora* foram comparados e integrados em um conjunto único de *frames* e relações associativas, permitindo uma visão mais ampla e historicamente situada de como "*indexing*" é conceituado e relacionado a outros processos de OC ao longo do tempo.

O mesmo procedimento foi aplicado aos artigos do *corpus* 2, igualmente carregados em um caderno do NotebookLM e submetidos ao *prompt* analítico descrito nesta seção. Em um primeiro teste, todos os 28 artigos do *corpus* 2 foram enviados em um único bloco, o que resultou em respostas mais gerais e menos detalhadas quanto às relações associativas tipadas. Em seguida, buscando manter a simetria metodológica com o *corpus* 1, os artigos do *corpus* 2 foram reagrupados em conjuntos de cinco textos por vez (e três ao final) e novamente analisados. Nessa configuração em blocos menores, o modelo passou a explicitar com maior precisão as relações associativas tipadas e os elementos do *frame* "*indexing*", produzindo resultados mais comparáveis aos obtidos no *corpus* 1.

Essa simetria metodológica contribui para que a extração de *frames* e relações associativas nos dois *corpora* seja conduzida com o mesmo modelo de linguagem, o mesmo *prompt* analítico e uma organização semelhante dos conjuntos de artigos, permitindo a comparação integrada dos resultados.

5 RESULTADOS

Para estruturar esta análise, a Inteligência Artificial (IA) utilizou a Teoria de *Frames* e a Terminologia Baseada em *Frames*, que tratam o significado terminológico como uma expansão progressiva do contexto em estruturas de conhecimento cognitivo. Além disso, aplicou a modelagem de relações associativas tipadas para trazer uma semântica formal que ultrapassa as simples relações hierárquicas e associativas genéricas dos tesauros tradicionais.

A seguir, apresentam-se os resultados obtidos na primeira etapa de análise do *corpus*, estruturada por décadas.

Quadro 1 - Anos 70

Artigo	Frame	Relação associativa tipada	Termos relacionados	Excerto do Texto
"A probabilistic approach to automatic keyword indexing" (Harter)	**Automatic Keyword Indexing** (Foco estatístico)	statistical rule <instrumentFor> automatic keyword indexing specialty word <targetOf> keyword indexing	specialty words, keywords, index terms, statistical rules, 2-Poisson model	"The problem studied in this research is that of developing a set of formal statistical rules for the purpose of identifying the keywords of a document—words likely to be useful as index terms for that document. [...] specialty words are not so distributed."
"Index term weighting" (Sparck Jones)	**Index Term Weighting** (Foco na atribuição de pesos)	statistical weighting <instrumentFor> index term weighting document term <hasProperty> weight	term weighting, document terms, request terms, statistical information, intrinsic weights	"Weighting document or request terms so that some are defined as more important than others is an appealing idea. [...] Statistical information is indeed more readily obtained automatically than manually."

"On indexing, retrieval and the meaning of about" (Maron)	**Subject Indexing / Aboutness** (Foco cognitivo e comportamental)	aboutness <isConditionFor> subject indexing document <patientOf> indexing procedure	aboutness, subject content, indexing procedure, document retrieval, satisfaction	"The key concept of such a theory of indexing would appear to be the concept about, because the key decision in the indexing procedure concerns whether a document is about some subject."
"On the Specification of Term Values in Automatic Indexing" (Salton & Yang)	**Automatic Indexing Evaluation** (Foco em métricas do sistema)	indexing exhaustivity <isPropertyOf> automatic indexing term specificity <isPropertyOf> automatic indexing	indexing exhaustivity, term specificity, term values, automatic indexing, retrieval	"Two fundamental notions in the theory of automatic indexing are known respectively as indexing exhaustivity and term specificity. Indexing exhaustivity refers to the accuracy and depth with which the various topic areas germane to a given document are reflected in the set of index terms assigned..."
"Probabilistic models for automatic indexing" (Bookstein & Swanson)	**Probabilistic Indexing / Word Clustering** (Foco na modelagem matemática de ocorrências)	probabilistic model <instrumentFor> automatic indexing word clustering <isConditionFor> indexing	clustering, word occurrences, probabilistic model, content-bearing words, index terms, Poisson process	"Thus the property of clustering to a degree exceeding whatever could be accounted for by a random distribution of word occurrences must be intrinsic to the concept of indexing".

Com esse quadro completo da década de 1970, podemos visualizar claramente como o frame de indexing oscila entre um procedimento de agrupamento estatístico/probabilístico (Harter, Bookstein & Swanson), ponderação algorítmica (Sparck Jones, Salton) e um julgamento cognitivo de pertinência (Maron).

Quadro 2 - Anos 80

Artigo	Frame	Relação associativa tipada	Termos Relacionados	Excerto do Texto
"FASIT: A fully automatic syntactically based indexing system" (Dillon & Gray, 1983)	**Syntactically Based Automatic Indexing** (Foco na extração de termos representativos baseada em padrões sintáticos)	syntactic pattern <instrumentFor> automatic indexing content bearing textual unit <targetOf> indexing	content bearing textual units, syntactic patterns, phrase indexing, canonical forms, stem indexing	"The aim of automatic indexing is to achieve a compact representation of a document suitable for retrieval. FASIT [...] identifies content bearing textual units without a full parse, and, without using semantic criteria, groups these units into quasisynonymous sets."
"Indexing consistency in MEDLINE" (Funk & Reid, 1983)	**Manual Indexing Consistency** (Foco no julgamento humano e uso de vocabulário controlado)	indexer <agentOf> manual indexing subject heading <instrumentFor> manual indexing consistency <isPropertyOf> indexing	manual indexing, interindexer consistency, subject headings, MeSH, central concepts, human indexer	"Manual indexing, the assignment of subject headings to a document, probably is the most critical aspect in any system's performance."

"Indexing consistency, quality and efficiency" (Rolling, 1981)	**Indexing Quality and Effectiveness Evaluation** (Foco nas métricas de qualidade baseadas na concordância humana)	indexing quality <isPropertyOf> indexing inter-indexer consistency <measures> indexing qualityinformation content <patientOf> indexing	indexing quality, indexing effectiveness, inter-indexer consistency, information content, term weights	"Indexing quality determines whether the information content of an indexed document is accurately represented. [...] indexing consistency is essentially a measure of the similarity of reaction of different human beings processing the same information."
"Knowledge-Based Indexing of the Medical Literature" (Humphrey & Miller, 1987)	**Interactive Knowledge-Based Indexing** (Foco no uso de *frames* computacionais como inteligência artificial para assistir o indexador)	knowledge base <instrumentFor> interactive indexing human indexer <agentOf> interactive indexing frame <instrumentFor> knowledge representation	knowledge-based systems, computer-assisted indexing, frames, knowledge representation, slot values	"The general purpose of the project is to develop and test interactive knowledge-based systems for computer-assisted indexing of the medical literature currently indexed in the MEDLINE database."
"The Effectiveness of a Nonsyntactic Approach to Automatic Phrase Indexing" (Fagan, 1989)	**Automatic Phrase Indexing** (Foco na identificação de sintagmas e descritores complexos via métodos não-sintáticos)	nonsyntactic procedure <instrumentFor> automatic phrase indexing complex descriptor <productOf> phrase indexing phrase descriptor <targetOf> indexing	complex descriptors, automatic phrase indexing, nonsyntactic approach, phrase descriptors, term relationships	"The general objective of the present study is to further investigate the potential benefits of incorporating informa-tion about term relationships into representations of document content, and to develop effective methods of identifying phrase descriptors that are accurate indicators of document content."

Neste período, você pode notar que o conceito de indexação se bifurca fortemente na literatura. Uma vertente volta a colocar o **fator cognitivo humano** (<agentOf>) no centro da análise de sistemas de informação, focando na **consistência** e **qualidade** de indexadores reais (como os da base MEDLINE). Ao mesmo tempo, a vertente da **automação** passa a explorar atributos linguísticos mais profundos, substituindo o simples isolamento estatístico de palavras por **padrões sintáticos** (*syntactic patterns*) e pela busca por **descritores complexos** ou frases inteiras (*phrase descriptors*), culminando em sistemas interativos de assistência que já empregavam a própria **teoria de frames** para representação do conhecimento médico em IA.

Quadro 3 - Anos 90

Artigo	Frame	Relação associativa tipada	Termos Relacionados	Excerto do Texto
"Indexing and access for digital libraries and the Internet..." (Bates, 1998)	**Digital Environment Subject Indexing** (Foco no desafio cognitivo do indexador humano em antecipar as lacunas de conhecimento do usuário em bibliotecas digitais)	human indexer <agentOf> subject indexing information gap <targetOf> indexer anticipation	subject indexing, human factors, digital resources, user, information gap, indexer	"The user, in effect, describes the fringes of a gap in knowledge, and can only guess what the “filler” for the gap would look like [...] The indexer, on the other hand, has the record in hand [...] the challenge for the indexer is to try to anticipate what terms people with information gaps of various descriptions might search for...".

"Indexing by latent semantic analysis" (Deerwester et al., 1990)	**Latent Semantic Indexing** (Foco na superação de problemas linguísticos como sinonímia e polissemia através de relações estatísticas de coocorrência latente)	singular value decomposition <instrumentFor> latent semantic indexing conceptual content <targetOf> indexing	latent semantic indexing, conceptual content, synonymy, polysemy, singular-value decomposition, artificial concepts	"We describe here a new approach to automatic indexing and retrieval. It is designed to overcome a fundamental problem that plagues existing retrieval techniques that try to match words of queries with words of documents. The problem is that users want to retrieve on the basis of conceptual content...".
"KeyWords Plus algorithmic derivative" (Garfield, 1993)	**Derivative Algorithmic Indexing** (Foco na geração de termos de indexação a partir de relações de citação, especificamente os títulos das obras citadas)	cited article title <instrumentFor> derivative indexing relevant indexing term <productOf> algorithmic indexing	derivative indexing, cited articles, indexing terms, KeyWords Plus, algorithmic derivative	"These unpublished experiments demonstrated that “derivative” indexing, i.e., using the titles of articles cited in a source text, would produce relevant indexing terms. In fact, the procedure anticipated most of the indexing terms assigned by experienced medical indexers".

"Some issues in the indexing of images" (Shatford Layne, 1994)	**Image Indexing** (Foco na decisão intelectual sobre quais atributos visuais, genéricos ou específicos, devem ser indexados ou agrupados)	image attribute <targetOf> image indexing image grouping <productOf> image indexing	image indexing, attributes, generic identity, specific identity, groupings, represented work	"...it is necessary to decide which attributes need to be indexed, which can be simply noted in conjunction with images, and which may be left for the searcher for the images to perceive. That is, it is necessary to determine which attributes are needed to provide useful groupings of images...".
"Using latent semantic indexing for literature based discovery" (Gordon & Dumais, 1998)	**Literature-Based Discovery Indexing** (Foco no uso da indexação semântica latente para descobrir conexões ocultas entre literaturas médicas isoladas)	latent semantic indexing <instrumentFor> literature-based discovery hidden relationship <targetOf> latent indexing	literature-based discovery, latent semantic indexing, intermediate literatures, term co-occurrences, hidden relationships	"We explore latent semantic indexing's effectiveness on two discovery processes: uncovering "nearby" relationships that are necessary to initiate the literature based discovery process; and discovering more distant relationships that may genuinely generate new discovery hypotheses".

A década de 90 expande enormemente o frame de *indexing*.

1. **O Fator Latente:** Com a criação da *Latent Semantic Indexing (LSI)*, a indexação deixa de tratar os termos textuais de forma literal e superficial. O frame evoca a necessidade de extrair o "conteúdo conceitual" (<targetOf> indexing) calculando matematicamente as estruturas ocultas da linguagem (Deerwester et al., 1990; Gordon & Dumais, 1998).
2. **A Fronteira Não-Textual:** O texto de Shatford Layne (1994) transfere o frame de indexação para o ambiente imagético, onde a complexidade reside em determinar a diferença

ontológica entre o que a imagem *é* (uma gravura) e o que ela *representa* (uma ponte genérica ou a Ponte do Brooklyn) (image attribute <targetOf> image indexing).

3. **Novos Paradigmas Cognitivos e Bibliométricos:** Bates (1998) traz de volta a figura cognitiva do indexador (<agentOf>), mas agora modelando a indexação como uma tentativa quase impossível de "adivinhar a lacuna de informação do usuário". Em paralelo, Garfield (1993) automatiza a indexação aproveitando o comportamento de citação dos próprios cientistas (indexação derivativa).

Quadro 4 - Anos 2000

Artigo	Frame	Relação associativa tipada	Termos Relacionados	Excerto do Texto
"A framework for understanding Latent Semantic Indexing (LSI) performance" (Kontostathis & Pottenger, 2006)	**Latent Semantic Indexing** (Foco na matemática computacional para extrair relações semânticas latentes baseadas na coocorrência de termos)	singular value decomposition <instrumentFor> latent semantic indexing term co-occurrence <isConditionFor> latent semantic indexing latent semantic <targetOf> indexing	latent semantics, term co-occurrence, Singular Value Decomposition (SVD), dimensionality reduction, term-to-term matrix	"LSI is a vector space approach for modeling documents, and many have claimed that the technique brings out the 'latent' semantics in a collection of documents. [...] we have discovered the basis for the claim that is frequently made for LSI: LSI emphasizes underlying semantic distinctions (latent semantics) while reducing noise in the data."

"An extension of the Hirsch index: Indexing scientific topics and compounds" (Banks, 2006)	**Bibliometric Topic Indexing** (Foco no uso de indicadores bibliométricos de impacto acadêmico, o *h-index*, para medir, classificar e indexar o "calor" de um tópico)	h-b index <instrumentFor> topic indexing scientific topic <targetOf> indexing compound <targetOf> indexing	Hirsch index, h-b index, scientific topics, compounds, hot topic, bibliometric indicator	"In this paper I give an interesting twist on the h index and apply it further to the case of interesting topics and compounds. [...] A simple and quick, but by no means full answer could be easily obtained by comparing the h index for topics or compounds, which will from now on be called the 'h-b index'".
"Coauthorship patterns and trends in the sciences... implications for database indexing" (Glänzel, 2002)	**Database Indexing / Search Strategies** (Foco na aplicação de estudos bibliométricos de coautoria para determinar quais documentos devem servir como "nós principais" na indexação de bases de dados)	multiauthored publication <targetOf> database indexing core document <instrumentFor> search strategy citation impact <isPropertyOf> indexed publication	database indexing, search strategies, core documents, multiauthored publications, bibliographic coupling	"From the viewpoint of library and database management, the following implications should be mentioned. [...] Multiauthored, and first of all internationally coauthored publications, proved to hold key positions within the framework of scientific communication; their citation impact is assumed to exceed standard reception. [...] frequently cited and strongly interrelated in terms of theme, can thus serve as core documents in search strategies".

"Domain-Independent Automatic Keyphrase Indexing with Small Training Sets" (Medelyan & Witten, 2008)	**Controlled Automatic Keyphrase Indexing** (Foco no método híbrido que cruza a extração livre com a atribuição terminológica via aprendizado de máquina e tesauro)	domain-specific thesaurus <instrumentFor> keyphrase indexing machine learning model <instrumentFor> automatic indexing well-formed phrase <productOf> indexing	keyphrase indexing, term assignment, keyphrase extraction, controlled vocabulary, machine learning, KEA++	"We have built a system for keyphrase indexing, an intermediate approach between keyphrase extraction and term assignment. It combines the advantages of both while avoiding their shortcomings. It uses a domain-specific thesaurus as the controlled vocabulary".
"Thesaurus Based Automatic Keyphrase Indexing" (Medelyan & Witten, 2006)	**Thesaurus-Based Automatic Keyphrase Indexing** (Foco na gênese do sistema KEA++, detalhando o uso de pseudo frases e relações semânticas do tesauro)	machine learning <instrumentFor> keyphrase indexing pseudo phrase <patientOf> candidate identification semantic information <instrumentFor> filtering	keyphrase indexing, term assignment, KEA++, thesaurus, controlled vocabulary, semantic information, candidate identification	"This paper describes keyphrase indexing, an intermediate approach between keyphrase extraction and term assignment that combines the advantages of both and avoids their shortcomings. The new keyphrase indexing algorithm, called KEA++ [...] is based on machine learning and works in two main stages: candidate identification... and filtering".

É interessante notar a **evolução terminológica do próprio conceito** representada nos dois artigos de Medelyan e Witten (2006, 2008). Eles cunham na literatura a separação teórica exata de três frames distintos:

1. *Keyphrase extraction* (extração estatística livre);
2. *Term assignment* (atribuição manual/rígida via especialistas); e
3. O novo **Keyphrase Indexing**, que funde ambos usando tesauros como instrumento (<instrumentFor>) para mapear n-gramas textuais.

Além disso, vemos o frame de indexação expandindo suas fronteiras epistemológicas: o artigo de Banks (2006) demonstra que "indexar" (topic indexing) também passou a significar a medição bibliométrica de relevância usando lógicas de citação (o *h-index*).

Quadro 5 - Anos 2010

Artigo	Frame	Relação associativa tipada	Termos Relacionados	Excerto do Texto
"A comparative study of Flickr tags and index terms..." (Rorissa, 2010)	**Social Image Indexing / Collaborative Tagging** (Foco na transição do indexador profissional para o usuário leigo por meio de etiquetas geradas em ambientes colaborativos da Web 2.0)	social tagging <instrumentFor> image indexing user <agentOf> social tagging image attribute <targetOf> indexing	social tagging, collaborative tagging, folksonomies, index terms, user-generated tags, controlled vocabularies, image attributes	"For the purpose of the current work, image indexing is defined as the representation of images through textual descriptions of their contents, what they are of or about [...], rather than their isness".
"Coverage, field specialisation and the impact of scientific publishers..." (Torres-Salinas et al., 2014)	**Bibliographic Database Indexing** (Foco na indexação de monografias/livros em bases de dados para fins de avaliação de impacto acadêmico e bibliometria)	monograph <targetOf> database indexing scientific publisher <agentOf> monograph publishing Book Citation Index <instrumentFor> research evaluation	Book Citation Index, monographs, citation indexes, scientific publishers, citation counts, bibliometric data	"The launch in October 2011 of Thomson Reuters' newest product for scientific literature – the Book Citation Index (hereafter BKCI) – may well be a historic milestone in the field of scientometrics or just another failure in the race to develop tools capable of evaluating monographs".

"Errors in doi indexing by bibliometric databases" (Franceschini et al., 2015)	**Bibliometric / DOI Indexing Accuracy** (Foco na precisão, mapeamento de banco de dados e no papel crítico de identificadores únicos para evitar problemas de desambiguação e falsas métricas)	DOI <instrumentFor> paper disambiguation multiple DOI assignment <isPropertyOf> database error scientific paper <targetOf> DOI indexing	DOI indexing, bibliometric databases, database mapping errors, multiple DOI assignment, Scopus, disambiguation	"In bibliometrics, DOIs are used for univocally identifying scientific papers. The aim of this short communication is to raise the reader's awareness of bibliometric database errors in DOI indexing, in particular, the incorrect assignment of a single DOI to multiple papers".
"Invisible institutional repositories: Addressing the low indexing ratios..." (Arlitsch & O'Brien, 2012)	**Search Engine Repository Indexing (SEO)** (Foco na relação entre a adequação de esquemas de metadados utilizados pelas bibliotecas e o sucesso da coleta de dados por algoritmos de motores de busca)	metadata schema <instrumentFor> search engine indexing Google Scholar <agentOf> repository indexing institutional repository <targetOf> indexing	Google Scholar, institutional repositories, search engine optimization (SEO), metadata schemas, Dublin Core, crawlers, indexing ratio	"Google Scholar has difficulty indexing the contents of institutional repositories, and the authors hypothesize the reason is that most repositories use Dublin Core, which cannot express bibliographic citation information adequately for academic papers".

"MapReduce indexing strategies: Studying scalability and efficiency" (McCreadie et al., 2012)	**Distributed Large-Scale Indexing** (Foco na computação distribuída e no paradigma de MapReduce para criar índices invertidos eficientes sobre coleções de dados de tamanho terabyte)	MapReduce <instrumentFor> large-scale indexing posting list <productOf> inverted indexing computational scalability <isPropertyOf> distributed indexing	MapReduce, inverted index, posting lists, terabyte-scale corpora, distributed processing, single-pass indexing, map tasks, reduce tasks	"In Information Retrieval (IR), the efficient indexing of terabyte-scale and larger corpora is still a difficult problem. MapReduce has been proposed as a framework for distributing data-intensive operations across multiple processing machines".

Neste período, o frame de "INDEXING" mostra uma fragmentação radical em resposta às demandas da era da informação:

1. **Indexação Computacional Escalável:** Com McCreadie et al. (2012), o limite humano é completamente ultrapassado; o *frame* evoca MapReduce e processamento distribuído como instrumentos (<instrumentFor>) essenciais para lidar com a web em escala de terabytes.
2. **SEO e Visibilidade em Máquinas:** Arlitsch & O'Brien (2012) mostram que a indexação agora também significa otimizar repositórios (<targetOf> indexing) para que agentes automatizados (Google Scholar - <agentOf>) consigam ler a representação, revelando que formatos genéricos como o *Dublin Core* falham nesse nível de granularidade bibliométrica.
3. **A "Multidão" como Indexador:** Rorissa (2010) insere as *folksonomias* e as plataformas da Web 2.0 no radar da Organização do Conhecimento, reintroduzindo o fator humano, mas não o profissional, e sim os usuários comuns atuando via *Social Tagging* (user <agentOf> social tagging).
4. **Precisão Bibliométrica:** Franceschini et al. (2015) e Torres-Salinas et al. (2014) atrelam a indexação à identificação inequívoca (DOI) e às métricas de citação, onde a falha na indexação afeta diretamente a avaliação institucional e de impacto de monografias e papers.

Quadro 6 - Anos 2020 até 2025

Artigo	Frame	Relação associativa tipada	Termos Relacionados	Excerto do Texto

"Automated Subject Indexing: An Overview" (Golub, 2021)	**Machine Learning Automated Indexing** (Foco na substituição ou auxílio do julgamento humano por algoritmos de NLP e aprendizado de máquina aplicados a Sistemas de Organização do Conhecimento)	machine learning <instrumentFor> automated subject indexing human intellectual process <isReplacedBy> automated technique controlled vocabulary <targetOf> automated indexing	automated subject indexing, machine learning, natural language processing (NLP), knowledge organization systems (KOS), document representation, tf-idf	"Automated subject indexing is then machine-based subject indexing where human intellectual processes involved in those three steps are replaced by, for example, techniques based on computational linguistics and statistics".
"Subject indexing in humanities... " (Golub et al., 2020)	**Humanities Subject Indexing** (Foco na falha dos sistemas automáticos e comerciais genéricos em capturar as necessidades facetadas, de alta especificidade e granularidade exigidas pelos pesquisadores de humanidades)	high granularity <isPropertyOf> humanities indexing faceted vocabulary <instrumentFor> humanities indexing humanities scholar <agentOf> subject searching	subject indexing, humanities, controlled vocabularies, facets, high granularity, precision, recall, interdisciplinary	"To counter high recall and low precision, specific subject indexing should be implemented, involving (1) indexing policies that promote a high level of specificity and (2) indexing languages that are deep and detailed for any given topic...".

"View of Annif and Finto AI: Developing and Implementing..." (Suominen et al., 2022)	**Ensemble Automated Subject Indexing** (Foco no uso de ferramentas de código aberto modulares que combinam abordagens lexicais, associativas e redes neurais para sugerir assuntos em ambientes de produção)	ensemble algorithm <instrumentFor> automated subject indexing text classification <isConditionFor> automated indexing indexing quality <patientOf> evaluation	automated subject indexing, Annif, Finto AI, text classification algorithms, lexical approaches, associative approaches, ensemble backends, quality evaluation	"Annif is multilingual, independent of the indexing vocabulary, and modular. It integrates many text classification algorithms, including Maui, fastText, Omikuji, and a neural network model based on TensorFlow. Best results can often be obtained by combining several algorithms".

Nesta década atual, o conceito de indexação atinge uma maturidade híbrida:

1. **O Triunfo dos "Ensembles" (Conjuntos Algorítmicos):** Com ferramentas como o *Annif* (Suominen et al., 2022), o *frame* deixa de buscar "a regra de ouro" estatística ou sintática única das décadas passadas. A indexação evoca agora uma rede modular (ensemble algorithm <instrumentFor>), combinando redes neurais artificiais, processamento de linguagem natural e abordagens lexicais para simular a classificação de um texto.
2. **A Resistência Semântica das Humanidades:** Golub et al. (2020) mostram que a automação e as métricas comerciais (*Big Data* bibliométrico) falham estrondosamente nas Humanidades. O *frame* aqui retrocede à sua função epistemológica: a indexação precisa lidar com a ambiguidade, as metáforas, e as identidades geográficas/cronológicas, revelando que a **granularidade** (high granularity <isPropertyOf> humanities indexing) é uma demanda humana que as máquinas ainda lutam para processar.
3. **Modelagem Baseada em KOS:** Em Golub (2021), a automação deixa de ser apenas "extração de palavras-chave" para finalmente ancorar-se em Sistemas de Organização do Conhecimento (KOS) preexistentes. A máquina agora mapeia representações latentes do texto diretamente para vocabulários controlados intelectuais.

As quadros 1, 2, 3, 4, 5 e 6 apresentam os resultados da primeira etapa de construção do *corpus*, baseada na seleção de artigos por décadas. Em seguida, são apresentados os resultados da busca temática ampliada. Cabe reiterar que o *corpus* 2 foi igualmente subdividido em blocos menores de artigos para análise no NotebookLM, replicando a estratégia adotada no *corpus* 1, de modo a garantir

consistência na extração e comparação das relações associativas tipadas. Os quadros, a partir do 7, apresentam os artigos do *corpus* 2 ordenados pelo número de citações na *Web of Science*, iniciando pelo artigo mais citado e seguindo, em ordem decrescente, até os menos citados.

Quadro 7 - Artigos mais citados do *corpus* 2 e seus *frames* de indexação

Artigo	Frame	Relações Associativas Tipadas	Termos Relacionados	Excerto do Texto
"Fundamentals of Knowledge Organization" (Hjørland, 2003)	**Teleologia e Epistemologia** (A indexação vista como um processo pragmático voltado para identificar o potencial informativo e epistemológico de um documento para auxiliar atividades humanas, como a citação futura).	indexing <identifies> epistemological potentialities indexing <helps> future authors	Epistemological (or informative) potentialities, citable attributes, relevance connections, information retrieval, teleological nature.	*"Indexing is a process in the service of helping future authors to identify documents to be cited. This is done by identifying exactly those attributes that make it citable. The connections between cited and citing papers are relevance connections, and so also are (or should be) the connections between indexing terms and the documents indexed."*

"The Importance of Theories of Knowledge: Indexing and Information Retrieval as an Example" (Hjørland, 2011)*	**Teorias Epistemológicas da Indexação** (A indexação não como um ato neutro, mas como um reflexo de diferentes teorias do conhecimento — pragmáticas, críticas, cognitivas — que avaliam a contribuição de um documento).	indexing <evaluates> document contribution to knowledge request-oriented indexing <supports> given goals	Human-derived indexing, automated indexing, pragmatic and critical theories, request-oriented indexing, informative potentials.	*"The core of indexing is, as stated by Rowley and Farrow (2000), to evaluate a paper's contribution to knowledge and index it accordingly; or, with the words of Hjørland (1997), to index its informative potentials. What is known as 'request oriented indexing' is very much in accordance with pragmatic and critical views of indexing."*
"From Knowledge Organization to Knowledge Representation" (Giunchiglia, Dutta, Maltese, 2013)	**Abordagem Tradicional Centrada no Documento** (A indexação como um processo prático da Organização do Conhecimento tradicional em que documentos são descritos puramente por suas propriedades e metadados via vocabulários controlados).	indexing <uses> document properties indexing <reliesOn> controlled vocabularies	Document properties, subject, classification, controlled vocabularies, Online Public Access Catalogue (OPAC).	*"...documents are typically classified by subject and indexed by document properties such as title, author as well as subject. Indexing by title and author are straightforward as they are directly taken from the document. Indexing by subject is far more complicated as it requires an analysis of the document content..."*

"Reviews of Concepts in Knowledge Organization: Subject (of Documents)" (Hjørland, 2017)**	**Análise de Assunto e Representação** (A indexação como uma atividade dividida metodologicamente em duas etapas interdependentes: a análise conceitual do documento e a tradução dessa análise em rótulos/termos).	indexing <consistsOf> conceptual analysis indexing <consistsOf> translation stage	Conceptual analysis, translation stage, subject labels, subject representation, request-oriented indexing.	*"As stated by Lancaster (2003, 21), it is important to distinguish the conceptual analysis and the translation stages in indexing and classification. In conceptual analysis, subjects are attributed to documents, and in the translation stage subject labels are assigned to documents."*
"What is Knowledge Organization (KO)?" (Hjørland, 2008)	**Processos de Organização do Conhecimento (KOP)** (A indexação categorizada formalmente como um processo prático executado em instituições de memória, envolvendo trabalho semântico e interpretativo de atribuição de significado).	indexing <isA> knowledge organizing process indexing <involves> semantic labor	Knowledge organizing processes (KOP), memory institutions, document description, semantic labor, syntactic labor.	*"In the narrow meaning Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, bibliographical databases, archives and other kinds of 'memory institutions' [...] KO as a field of study is concerned with the nature and quality of such knowledge organizing processes (KOP)..."*

Quadro 8 - Indexação semântica, folksonomias e epistemologias da indexação

Artigo	Frame	Relações Associativas Tipadas	Termos Relacionados	Excerto do Texto
"A Knowledge Based Approach to Information Extraction..." (Vlachidis & Tudhope, s.d.)	**Extração de Informação e Interoperabilidade Semântica** (A indexação semântica automatizada vista como um processo de atribuição de anotações e metadados estruturados por meio de ontologias para permitir a recuperação avançada de relatórios arqueológicos).	semantic indexing <generates> semantic metadata semantic indexing <supports> semantic queries information extraction <usedFor> semantic indexing	Semantic annotation, information extraction, semantic metadata, CIDOC CRM, ontology, rule-based approach.	*"This semantic indexing is targeted at supporting complex and semantically defined queries that facilitate information retrieval from archaeological grey literature reports and cross searching over reports and datasets."*

"Basic-level categories: A review" (Hajibayova, 2013)	**Categorização e Indexação Centrada no Usuário** (A indexação abordada a partir da teoria de categorias de nível básico, utilizada para representar o conteúdo conceitual de recursos — especialmente imagens — de forma otimizada para o usuário comum).	indexing <uses> basic-level terms basic-level categories <supports> image indexing user-centered indexing <reliesOn> crosswalks	Basic-level categories, image indexing, conceptual content, knowledge organization, information representation.	*"...the universality of basic-level categories in systems of knowledge organization and representation would seem to indicate their universality in indexing as well..." "...how basic-level terms are used to represent the conceptual content of resources and, especially, images..."*
"Folksonomies and the New Order: Authority in the Digital Disorder" (Mai, 2011)	**Indexação Democrática e Folksonomias** (A indexação concebida como um processo colaborativo, de baixo para cima, no qual a autoridade advém do consenso, interpretação coletiva e atribuição de rótulos (tags) pelos próprios usuários).	democratic indexing <determines> authority users <perform> democratic indexing social tagging <isA> democratic indexing	Democratic indexing, folksonomies, social tagging, constructive interpretation, users, authority, autopoietic warrant.	*"Tagging and folksonomies are conceptually related to democratic indexing." "The democratic approach to indexing, therefore 'determines the authority from the agreement of its users: its warrant comes from the constructive interpretation of its users'."*

"Indexing of Musical Genres. An Epistemological Perspective" (Abrahamsen, 2003)	**Análise de Assunto e Paradigmas Epistemológicos** (A indexação de gêneros musicais em bibliotecas estruturada como uma análise de assunto que flutua entre a intuição auditiva baseada na experiência e a leitura de fontes escritas, dependendo do paradigma adotado).	genre indexing <requires> subject analysis listening indexing <basedOn> musical experience reading indexing <reliesOn> written sources	Genre indexing, listening indexing, reading indexing, subject analysis, classical music, popular music, epistemology.	*"When a librarian or an information specialist is indexing music by genres they will not normally perform a thorough analysis of the music at hand. [...] The goal is to make a subject analysis rather than a regular music analysis."*
"Indexing: Concepts and Theory" (Hjørland, 2018)	**Epistemologia Social e Teoria da Indexação** (A indexação concebida não como um ato neutro, mas como um processo pragmático e altamente carregado de teorias epistemológicas, visando avaliar a contribuição do documento para o conhecimento a fim de atender a demandas específicas).	indexing <isA> theory-laden process indexer <evaluates> document contribution to knowledge index <provides> access to source documents	Subject content, target document, indexing theory, epistemological theories, request-oriented indexing, document-oriented indexing.	*"An index is a kind of target document, which has the function of providing access to information in or about some source documents..." "The core of indexing is, as stated by Rowley & Farrow (2000, 99) to evaluate a paper's contribution to knowledge and index it accordingly..."*

Quadro 9 - Indexação manual, automática e "*social tagging*" em contextos documentais distintos

Artigo	Frame	Relações Associativas Tipadas	Termos Relacionados	Excerto do Texto
"Comparing social tags with subject headings on annotating books" (Wu et al., 2013)	**Controle Terminológico vs. Extração Automática** (A indexação acionada como um espectro de processos de representação de conteúdo, variando da indexação manual controlada por especialistas à indexação automática por algoritmos e, alternativamente, ao *social tagging* por usuários).	manual indexing <uses> controlled vocabulary automatic indexing <extracts> keywords social tagging <isAlternativeTo> manual indexing	Manual indexing, automatic indexing, controlled vocabulary, social tagging, subject terms, keywords, content-based access.	*"To facilitate content-based access, index terms are identified through either manual assignment (thus, manual indexing) or automatic extraction (automatic indexing). Traditionally, manual indexing is conducted by highly trained information professionals, and controlled vocabulary-based subject terms [...] are used 'to ensure effective indexing and to maintain the overall efficacy'."*

"Member Activities and Quality of Tags in a Collection of Historical Photographs in Flickr" (Stvilia & Jörgensen, 2010)	**Indexação de Imagens e Folksonomia** (A indexação contextualizada no cenário de imagens digitais, onde a prática tradicional baseada em vocabulários controlados é contrastada com a geração de metadados por não-especialistas orientada às necessidades do usuário final).	image indexing <uses> controlled vocabularies nonexpert-generated tags <usedFor> image indexing indexing <basedOn> end-user information needs	Image indexing, nonexpert-generated tags, social tagging, controlled vocabularies, information retrieval, search queries.	*"Researchers have emphasized the importance of indexing documents based on end-user information needs and search queries [...]. Also, in a recent study, the traditional Precision and Recall-based evaluation framework from the field of information retrieval was used to evaluate the effectiveness of nonexpert-generated tags in image indexing and retrieval..."*

"Provenance as a Knowledge Organization Principle" (Tognoli & Guimarães, 2019)	**Método de Recuperação Arquivística** (A indexação no domínio arquivístico vista não como uma simples atribuição de assuntos isolados, mas como um processo dependente da proveniência, da história administrativa e do princípio de organização estrutural do arquivo).	provenance method <usedFor> content indexing subject retrieval <proceedsByLinking> indexing with provenance	Subject retrieval, provenance method, content indexing, context, administrative histories, organizing principles.	*"Subject retrieval in the Provenance Method proceeds by linking subject queries with provenance information contained in administrative histories or biographies, thereby producing leads to files which are searched by using their internal structures."*
"This is Not Domain Analysis" (Albrechtsen, 2015)	**Paradigma do Processamento de Informação** (A indexação, especialmente a automática, abordada como parte de um paradigma mecanicista tradicional da Ciência da Informação focado no processamento para recuperação, em contraste com análises de domínio mais interpretativas).	automatic indexing <isPartOf> information processing paradigm automatic indexing <developedFor> information retrieval	Automatic indexing, information retrieval (IR), information processing paradigm, domain analysis, classification.	*"From the outset, the aim was to present a methodological alternative to the information processing paradigm in information science, which was concerned with developing models for information retrieval (IR) and automatic indexing."*

"Knowledge organization trends in library and information studies" (Saumure & Shiri, 2008)	**Evolução Tecnológica e Organização Auxiliada por Máquina** (A indexação inserida na evolução histórica pré e pós-Web, marcando a transição do trabalho humano focado em resumos/índices para o uso crescente de computadores, processamento de linguagem e metadados no contexto digital).	computers <perform> automatic indexing indexing <organizes> digital information machine-assisted indexing <extracts> semantic context	Machine-assisted knowledge organization, indexing and abstracting, automatic indexing, digital information, semantic nuances.	*"During the pre-web era, steps were being taken towards the use of machines in organizing knowledge. This was perhaps most evident in the indexing and abstracting field. Creation of programs, such as FASIT... revealed that the ability of computers to index materials was being explored." "Post-web studies of indexing and abstracting are often related to the indexing of digital information."*

Quadro 10 - Dimensões políticas, infraestruturas terminológicas e indexação na *Web Semântica*

Artigo	Frame	Relações Associativas Tipadas	Termos Relacionados	Excerto do Texto

"Political versus Apolitical Epistemologies in Knowledge Organization" (Hjørland, 2020)	**Epistemologia e Dimensão Política** (A indexação abordada não como uma prática neutra, mas como um ato político e interpretativo guiado por diferentes teorias do conhecimento — empiristas, racionalistas, hermenêuticas e pragmáticas — que serve a propósitos e valores humanos específicos).	indexing <cannotBe> neutral indexing <serves> human goals hermeneutical indexing <reflects> discourse needs	Epistemological assumptions, neutral criteria, pragmatic theories, hermeneutical theories, human goals, political attitudes.	*"Pragmatic and critical theories of indexing are in agreement with the historicist point of view that subjects are relative to specific discourses but emphasize that subject analysis should support given goals and values and should consider the consequences of indexing. These theories emphasize that indexing cannot be neutral and that it is a wrong goal to try to index in a neutral way."*
"Terminology Registries for Knowledge Organization Systems: Functionality, Use, and Attributes" (Golub et al., 2014)	**Serviços de Terminologia e Automação** (A indexação inserida no contexto de infraestruturas digitais e Web Semântica, facilitada por serviços terminológicos, registros de vocabulário e processos de indexação	terminology services <support> indexing automated indexing <assigns> terms from vocabularies KOS <usedFor> indexing	Terminology services, automated indexing, metadata creation, classification, tagging, SKOS, semantic interoperability.	*"Cataloging. This includes functionality to support indexing (classification, annotation, tagging) and metadata creation activities. It could be achieved via: a cataloging application associated with the TR; direct provision of web services by the TR itself; or making available*

	automática e anotação).			*information on third party services..."*
"Unleashing the Power of Data Through Organization" (Soergel, 2015)	**Extração, Transferência e Bases de Conhecimento** (A indexação manual vista como um suporte tradicional à busca, contrastada com a necessidade de estruturar o conhecimento em si — via extração de dados — para permitir a transferência de conhecimento entre domínios e análises complexas).	manual indexing <assists> searchers indexing <supports> knowledge transfer	Manual indexing, subject cataloging, knowledge transfer, information extraction, social tagging, search.	*"There is also much discussion on whether manual indexing/subject cataloging is still needed (either way, we need KO to assist users in search) and the role of KO in social tagging... For the future we need to focus also on structuring and representing the actual data or knowledge itself..." "Breton (1991) argued for indexing documents to support transfer."*

"An evaluation of enhancing social tagging with a knowledge organization system" (Matthews et al., 2010)	**Indexação Híbrida e Folksonomias** (A indexação tradicional (manual) vista como custosa e insustentável para grandes coleções digitais, contrastada com o *social tagging* (barato, porém ruidoso), propondo-se uma integração onde o controle vocabular incentiva e melhora a indexação feita pelos usuários).	traditional subject indexing <incurs> high costs social tagging <reduces> indexing costs vocabulary-based suggestions <encourage> indexing activities	Traditional subject indexing, manual indexing, social tagging, controlled vocabulary, folksonomies, classification.	*"Traditional subject indexing and classification are considered infeasible in many digital collections... However, there are costs associated with the use of knowledge organization systems – manual indexing or classification are a significant resource... Social tagging... hold[s] the promise of reducing indexing costs by drawing end-users into contributing..."*

"Extending the Simple Knowledge Organization System for Concept Management in Vocabulary Development Applications" (Tennis & Sutton, 2008)	**Web Semântica e Gestão de Vocabulários** (A indexação inserida no ambiente da Web Semântica, vista como um processo prático de recuperação que é apoiado por modelos como o SKOS, e que depende da gestão histórica e estruturada de conceitos abstratos utilizados pelos indexadores).	SKOS model <supports> indexing abstract concepts <appliedIn> indexing processes indexers <employ> concepts	SKOS model, Semantic Web, abstract concepts, vocabulary development applications (VDA), concept instance, designers, indexers, retrieval.	*"While the SKOS model sufficiently describes entities for modeling the current state of a scheme in support of indexing and search on the Semantic Web, it lacks the expressive power to serve the needs of VDA needing to maintain scheme historical continuity." "We demonstrated that this mediating entity [...] is necessary to the useful management and application of abstract concepts in indexing processes, vocabulary development, and information retrieval systems."*

Quadro 11 - Indexação colaborativa, feminista, facetada e políticas de indexação em repositórios

Artigo	Frame	Relações Associativas Tipadas	Termos Relacionados	Excerto do Texto

"Exploitation of folksonomies in subject analysis" (Kakali & Papatheodorou, 2010)	**Indexação Colaborativa / Social Tagging** (A indexação abordada no contexto da Web 2.0, em que a descrição temática é expandida ou complementada por vocabulários gerados pelos próprios usuários, contrastando com e enriquecendo a indexação tradicional de especialistas).	social tagging <enhances> subject indexing collaborative tagging <provides> index terms users <participateIn> subject indexing	Social tagging, collaborative tagging, subject indexing, folksonomies, tags, subject descriptors, knowledge organization systems (KOS).	*"This paper presents a methodology for the exploitation of social tagging in subject indexing, and explores that method through a case study in an academic library setting. The findings reveal the characteristics of users' tagging behavior, which mainly enhances the subject description of documents."*
"Knowledge Organization for Feminism and Feminist Research" (Samuelsson, 2010)	**Epistemologia Crítica e Marginalização Discursiva** (A indexação compreendida como um processo de modelagem de significado que, ao utilizar KOS universais e objetivistas, frequentemente falha em representar e acaba marginalizando discursos e perspectivas feministas).	indexing <shapes> meaning universal KOS <marginalizes> feminist knowledge feminist research <requires> context-sensitive indexing	Knowledge organization systems (KOS), indexing practice, classification, feminist discourse, subject headings, marginalization, meaning shaping.	*"Olson describes the practice of indexing and classification as 'naming information'... every single subject heading and classification, as well as the practical use of them in indexing and classification, are seen as shaping meaning, and are in that way important."*

"The Elusive Tale: Leveraging the Study of Information Seeking..." (La Barre & Tilley, 2012)	**Indexação de Domínio Específico e Facetada** (A indexação contextualizada nas necessidades complexas de busca do domínio do folclore, onde as descrições padronizadas são insuficientes e exigem índices de tipos e motivos, além de análise de facetas).	facet analysis <informs> indexing systems motif indices <usedFor> folktale indexing subject headings <highlight> folkloric themes	Descriptive practices, subject headings, tale type, motif indices, facet analysis, information retrieval, domain resources.	*"The Library of Congress provides guidelines for assigning subject headings to folklore material... Best practice according to H-1627 is to assign subject headings that highlight the following: '(a) the ethnic, national, or occupational group... (b) the theme of the folklore; (c) the folkloric emphasis or genre'."*,
"The indexing policy in the practices of Brazilian institutional repositories... " (Fujita et al., 2023)	**Política de Indexação e Gestão de Repositórios** (A indexação concebida como um processo estratégico de análise de assunto dentro de repositórios institucionais, que depende fundamentalmente de políticas formalizadas e controle de vocabulário para garantir a precisão	indexing policy <improves> information retrieval controlled vocabulary <standardizes> subject indexing institutional repositories <require> indexing policy	Indexing policy, subject cataloging, institutional repositories, vocabulary control, controlled vocabularies, subject analysis, information retrieval.	*"Indexing is a process of subject analysis for representing the significant content of printed or digital information resources, whose goal is subject-specific and/or exhaustive access and retrieval with and without vocabulary control in information systems."*

	e a recuperação da informação).			
"Written Knowledge: A Literary Perspective on Indexing Theory" (Andersen, 2000)	**Perspectiva Literária e Epistemologia Social** (A indexação vista teoricamente como uma atividade social e literária, na qual o indexador atua de maneira semelhante a um autor, produzindo representações documentais a partir dos contextos do discurso científico).	indexer <actsAs> author indexing <produces> document representation indexing theory <isA> literary theory	Document representation, indexer as author, literary theory, written knowledge, social epistemology, statement of knowledge.	*"An indexer must also take into account these contexts when producing a representation of a text with descriptors. That way an indexer can be considered an author, because the indexer produces a statement of knowledge when representing a document and thereby reproduces a document, namely a document representation."*

Quadro 12 - Artigos sobre ética, viés e avaliação de sistemas de indexação

Artigo	Frame	Relações Associativas Tipadas	Termos Relacionados	Excerto do Texto

"Male Homosexuality in Brazilian Indexing Languages: Some Ethical Questions" (Pinho & Guimarães, 2012)	**Ética e Viés Cultural na Representação** (A indexação concebida como uma prática social e altamente subjetiva, vulnerável a crenças pessoais e vieses culturais/ideológicos, onde a escolha de rótulos afeta a justiça social e a precisão da recuperação).	librarians <perform> indexing indexing <influencedBy> cultural context indexing languages <represent> subjects	Indexing languages, subject representation, subjectivity, cultural and ideological context, precision, terminological control, bias.	*"The librarians establish labels when indexing in order to look for substitutes for knowledge; most of the time they are influenced by subjectivity. Certainly, this action is based on a cultural and ideological context, and that is why librarians are vulnerable in that their personal beliefs interfere with the representation of knowledge that is being analyzed in that moment"*.
"The Empirical Approach in the Evaluation of Information Systems" (Fugmann, 1999)	**Processo Intelectual e Previsibilidade** (A indexação vista como um processo intelectual sistemático de interpretação de texto, composto pela seleção e representação da essência documental, com o fim estrito de gerar previsibilidade na recuperação).	intellectual indexing <achieves> text interpretation indexing <provides> predictable concept representations indexing <consistsOf> essence selection	Intellectual indexing, essence selection, essence representation, consistency, predictability, concept representations, text interpretation.	*"The solution to this puzzle is that consistency is paramount only in the first step of indexing, that of essence selection. Here it is identical with the predictability of the selection. In the second step of indexing, that of essence representation,*

				only predictability is required".
"The Representation of Ethics and Knowledge Organization in the WoS and LISTA Databases" (Martínez-Ávila et al., 2015)	**Políticas de Indexação e Percepção de Domínio** (A indexação em bases de dados atuando como um filtro epistemológico, onde políticas de cobertura, escolhas algorítmicas e inconsistências moldam, silenciam ou distorcem a percepção pública de um domínio científico).	inconsistencies in indexing <misrepresent> domain indexing policies <cause> omissions databases <perform> indexing	Indexing policies, inconsistencies in indexing, databases, retrieval profile, deliberate omission, misrepresentation, indexing process.	*"...inconsistencies in indexing can misrepresent the full picture of the domain by including irrelevant documents or omitting relevant ones, sometimes inexplicably"*. *"Omissions based on indexing policies (such as coverage and format) and other omissions (such as the misrepresentation of contents) can affect the perception of the concepts and scientific configuration of sensitive topics such as 'knowledge organization' and 'ethics'"*.

6 ANÁLISE

A leitura histórica do *corpus* 1, organizado por décadas, permite observar que o conceito de indexação nunca foi nem tecnicamente neutro nem epistemologicamente estável. A cada período, o *frame* de "*indexing*" se reconfigura ao articular, de forma distinta, processos, agentes, instrumentos, critérios de qualidade e objetivos de uso, de modo que não há um "núcleo duro" imutável, mas uma sucessão de enquadramentos teórico-tecnológicos. Ao seguir essa trajetória, torna-se possível compreender a indexação não apenas como operação técnica localizada, mas como um ponto de condensação das disputas entre paradigmas da Organização do Conhecimento (OC), modelos de Representação do Conhecimento (RC) em IA e concepções de usuário e de documento.

6.1 Análise do *corpus* por décadas

A análise do primeiro *corpus* foi organizada em perspectiva histórica, com o objetivo de observar as transformações do *frame* de "*indexing*" ao longo do tempo. Esse recorte permite identificar continuidades, deslocamentos e inflexões na maneira como a literatura da área concebe a indexação, seus instrumentos, seus agentes e suas finalidades.

6.1.1 Anos 1970: a promessa probabilística e o retorno do "*about*"

Na década de 1970, o *frame* predominante de "*indexing*" é fortemente marcado pelo entusiasmo com modelos probabilísticos e estatísticos de recuperação da informação. Os trabalhos de Harter, Bookstein e Swanson, Sparck Jones e Salton evidenciam um esforço de formalização da indexação como um processo quantificável, em que regras estatísticas, distribuição de termos e pesos numéricos assumem o papel de instrumentos centrais. Expressões como "*statistical rule <instrumentFor> automatic keyword indexing*" e "*specialty word <targetOf> keyword indexing*" cristalizam essa visão, na qual o documento é fragmentado em unidades lexicais cuja especialidade pode ser inferida a partir do desvio em relação a uma distribuição aleatória.

Em paralelo, Sparck Jones introduz a questão do peso diferencial dos termos, ao propor a ponderação de termos de indexação como mecanismo de

distinção entre itens mais e menos informativos. A relação "*statistical weighting <instrumentFor> index term weighting*" e a ideia de que "*statistical information is indeed more readily obtained automatically than manually*" colocam a estatística não apenas como técnica auxiliar, mas como epistemologia implícita da relevância: aquilo que se repete com certos padrões passa a ser tomado como mais representativo do conteúdo documental. Salton e Yang reforçam essa lógica ao formalizar "*exhaustivity*" e "*specificity*" como propriedades quantificáveis do sistema, de forma que "*indexing exhaustivity <isPropertyOf> automatic indexing*" se torna uma variável de projeto, e não uma qualidade subjetiva do indexador.

Contudo, a mesma década assiste a um gesto teórico divergente, representado por Maron, que recoloca o conceito de "*aboutness*" no centro da discussão. Ao insistir que a decisão-chave na indexação é determinar de que um documento é "*about*", o autor reintroduz uma dimensão cognitiva e interpretativa que escapa à pura contagem de palavras. Nesse sentido, o *frame* dos anos 1970 já é internamente tensionado: de um lado, a aposta em modelos probabilísticos e pesos numéricos como formas de domesticar o conteúdo conceitual; de outro, a percepção de que a decisão sobre o assunto permanece ancorada em julgamentos humanos, ainda que se tente enquadrá-los em métricas formais.

Do ponto de vista da OC, essa década inaugura um primeiro descolamento entre indexação como operação sintático-estatística sobre cadeias de caracteres e indexação como prática interpretativa sobre documentos "sobre" algo. Essa fissura terá consequências duradouras: as soluções técnicas futuras vão oscilar entre aprofundar a via estatística, sofisticando seus instrumentos, e resgatar, sob novas formas, a centralidade do julgamento conceitual.

6.1.2 Anos 1980: entre o indexador cognitivo e os padrões sintáticos

Os estudos sobre consistência de indexação em "*Indexing Consistency in Medline*" explicitam a indexação manual como processo crítico de mediação, em que o *"indexer <agentOf> manual indexing*" e o "*subject heading <instrumentFor> manual indexing*" são vistos como condições para a qualidade global do sistema. A relação "*inter-indexer consistency <measures> indexing quality*" materializa uma concepção em que a OC depende de acordos cognitivos e normativos entre profissionais, e em que a variabilidade humana é tanto risco quanto métrica.

Ao mesmo tempo, Dillon e Gray (Fasit) representam uma tentativa de deslocar a fronteira da automação para além da simples frequência de termos. A ênfase em "*syntactic pattern <instrumentFor> automatic indexing*" e "*content bearing textual unit <targetOf> indexing*" indica que o foco passa a recair em unidades maiores (sintagmas, grupos nominais, estruturas de frase) tratados como portadores privilegiados de conteúdo. O objetivo de identificar unidades textuais portadoras de conteúdo sem uma análise sintática completa mostra que a indexação automática começa a aproximar-se de decisões antes reservadas ao indexador humano, embora ainda sem recorrer a semântica explícita. Fagan, ao investigar indexação por frases automática por métodos não sintáticos, complementa esse movimento ao conceber "*complex descriptor <productOf> phrase indexing*" e "*phrase descriptor <targetOf> indexing*", aproximando a indexação automática de práticas tradicionais de construção de descritores compostos.

Talvez o ponto mais significativo da década, contudo, seja o surgimento de sistemas de indexação assistida por conhecimento, como o projeto de Humphrey e Miller em "*Knowledge-Based Indexing of the Medical Literature*". Aqui, "*knowledge base <instrumentFor> interactive indexing*" e "*frame <instrumentFor> knowledge representation*" introduzem explicitamente a teoria de *frames* e a representação de conhecimento no contexto da indexação da literatura médica. O indexador permanece como "*agentOf*", mas passa a operar em diálogo com um sistema que codifica expectativas e relações típicas do domínio, antecipando a convergência, décadas depois, entre SOC, ontologias e modelos baseados em *frames*. A contribuição mais forte do texto é mostrar uma tentativa precoce de combinar tesauro controlado, representação de conhecimento e interação humano-computador para indexação documental. Isso é importante porque o artigo antecipa ideias que hoje aparecem em sistemas inteligentes de apoio à catalogação, recuperação de informação e extração de conhecimento.

A década de 1980, portanto, consolida a percepção de que a indexação é simultaneamente processo cognitivo, uso de vocabulário controlado e operação passível de alguma automação. O *frame* de *"indexing*" se bifurca entre um polo que enfatiza a expertise humana e a consistência, e outro que explora padrões formais cada vez mais complexos, incluindo estruturas frasais e representações de conhecimento. Do ponto de vista epistemológico, isso revela uma OC em situação de fronteira: dependente da interpretação humana, mas pressionada a traduzir essa

interpretação em regras, padrões e bases de conhecimento parcialmente formalizáveis.

6.1.3 Anos 1990: latência semântica, imagens e novas fronteiras cognitivas

A década de 1990 aprofunda a dimensão estatística da indexação, ao mesmo tempo em que amplia seu campo de aplicação para domínios não textuais e bibliométricos. A introdução da "*Latent Semantic Indexing*", por Deerwester *et al.*, é emblemática dessa inflexão. Ao propor "*singular value decomposition <instrumentFor> latent semantic indexing*" e "*conceptual content <targetOf> indexing*", os autores sugerem que o objetivo da indexação automática não é mais alinhar literalmente palavras de consulta e palavras de documentos, mas capturar um "conteúdo conceitual" latente que se manifesta nas coocorrências.

Em outra frente, Shatford Layne desloca o debate para o campo da imagem, ao problematizar algumas questões na indexação de imagens. Ao articular "*image attribute <targetOf> image indexing*" e "*image grouping <productOf> image indexing*", a autora mostra que, no domínio visual, a indexação exige decisões ontológicas sobre a diferença entre o que a imagem é (um tipo de objeto gráfico) e o que ela representa (uma ponte genérica ou a Ponte do *Brooklyn*). O *frame* deixa de ser apenas verbo-texto e se torna uma cena multimodal, em que atributos visuais, identidades genéricas e específicas, e agrupamentos por similaridade desafiam os modelos pensados para documentos escritos.

Simultaneamente, Garfield introduz a noção de indexação derivativa, na qual "*cited article title <instrumentFor> derivative indexing*" e "*relevant indexing term <productOf> algorithmic indexing*" expressam uma mudança epistemológica importante. Em vez de partir do texto integral do documento, a indexação passa a utilizar relações de citação como fonte de termos, assumindo que as escolhas de referências feitas pelos autores são, elas mesmas, atos de indexação implícita.

Por fim, Bates reintroduz com força a dimensão cognitiva e pragmática do indexador humano em ambientes digitais. Quando se afirma que "*human indexer <agentOf> subject indexing*" e "*information gap <targetOf> indexer anticipation*", o *frame* deixa de descrever apenas relações entre texto e termos para incorporar a lacuna de conhecimento do usuário como objeto da indexação. Indexar, aqui, é tentar adivinhar o contorno de uma ausência: o que o usuário não sabe, mas

procura, em uma *web* e em bibliotecas digitais marcadas por sobrecarga informacional e heterogeneidade de recursos.

A década de 1990, portanto, expande enormemente o *frame* de "*indexing*": da estatística de superfície à semântica latente, do texto ao visual, do documento isolado às redes de citação e às lacunas cognitivas dos usuários. Do ponto de vista da construção de um tesauro-dicionário para a OC, essa expansão sugere que as relações associativas em torno de indexação não podem ser reduzidas a poucas categorias genéricas; elas atravessam dimensões processuais, instrumentais, epistêmicas e ontológicas bastante heterogêneas.

6.1.4 Anos 2000–2010: bibliometria, escalabilidade e o acoplamento à infraestrutura

A partir dos anos 2000, a indexação passa a ser cada vez mais pensada em escala de sistemas e infraestruturas, em forte diálogo com métricas bibliométricas, *web* de dados e arquiteturas de processamento distribuído. Kontostathis e Pottenger retomam a "*Latent Semantic Indexing*" (LSI) de modo mais técnico, ao detalhar "*singular value decomposition <instrumentFor> latent semantic indexing*" e "*term co-occurrence <isConditionFor> latent semantic indexing*". O foco desloca-se para a compreensão de como a técnica efetivamente traz à tona distinções semânticas latentes ao reduzir o ruído dos dados. Indexação passa a significar, aqui, um modo de controlar dimensionalidade e destacar padrões estruturais que interessam à recuperação, o que prepara o terreno para aplicações em grandes coleções digitais.

Em Glänzel, "*multiauthored publication <targetOf> database indexing*" e "*core document <instrumentFor> search strategy*" indicam que a indexação em bases de dados passa a se apoiar em documentos "nódulos" altamente citados e coautoriados, que funcionam como pivôs para estratégias de busca. Já Banks radicaliza essa perspectiva ao propor o "*h-b index <instrumentFor> topic indexing*", em que tópicos e compostos científicos são indexados em função de indicadores de impacto derivados do h-index. Indexar deixa de ser apenas descrever o conteúdo de um documento e passa a ser também situá-lo em um espaço de relevância e "calor" temático medido por citações.

Na mesma década, avanços em indexação por palavras-chave, como de Medelyan e Witten, introduzem uma forma híbrida em que os limites entre extração

livre e atribuição controlada se tornam mais fluidos. “*Domain-specific thesaurus <instrumentFor> keyphrase indexing*” e “*machine learning model <instrumentFor> automatic indexing*” mostram que a indexação passa a ser mediada por modelos de aprendizado que operam entre um vocabulário controlado e candidatos gerados automaticamente.

Já na década de 2010, o problema deixa de ser apenas “como indexar” para se tornar “como indexar em escala”. Trabalhos sobre “*MapReduce indexing strategies*” investigam explicitamente escalabilidade e eficiência, ao articular “*text collection <patientOf> MapReduce indexing*” e “*scalability <isPropertyOf> indexing strategy*”. Em paralelo, a expansão da *web* acadêmica e dos repositórios institucionais introduz novas preocupações: Arlitsch e O’Brien discutem “*invisible institutional repositories*” a partir de erros ou lacunas na indexação por *Google Scholar*, enquanto Franceschini et al. mostram como inconsistências na indexação de DOIs (“Digital Object Identifier”) em bases bibliométricas comprometem a visibilidade e a mensuração de produção científica.

Do ponto de vista da OC, essas décadas expõem um deslocamento importante: a indexação deixa de ser uma operação circunscrita ao interior de um sistema documental e passa a ser vista como componente de uma ecologia de serviços, métricas, motores de busca e infraestruturas distribuídas. O *frame* de “*indexing*” acopla-se a indicadores bibliométricos, algoritmos de aprendizagem de máquina, arquiteturas de processamento em larga escala e plataformas de descoberta acadêmica. Ao mesmo tempo, a emergência de *folksonomias* e marcação social (“*social tagging*”), observada em trabalhos como os de Rorissa, Stvilia e Jörgensen e outros da busca ampliada, coloca em cena uma indexação democrática, em que “*users <agentOf> democratic indexing*” e “*end-user information needs <basisFor> indexing documents*”. Ainda que esse aspecto apareça com mais força na análise da busca ampliada, já se percebe, nesse período, a coexistência de três regimes: o profissional controlado, o algorítmico automatizado e o participativo distribuído.

6.1.5 Anos 2010–2020: visibilidade algorítmica, participação distribuída e infraestrutura de indexação

Na década de 2010, o *frame* de "*indexing*" sofre uma inflexão decisiva, pois deixa de ser compreendido apenas como técnica de representação temática ou como operação de apoio à recuperação e passa a ser pensado como componente estrutural de ecossistemas digitais amplos, atravessados por plataformas, motores de busca, bases bibliométricas e arquiteturas de processamento distribuído. Se, nos anos 2000, já se observava um movimento de aproximação entre indexação, aprendizado de máquina, tesauros e indicadores bibliométricos, no período de 2010 a 2020 esse processo se intensifica e se desloca para o nível infraestrutural: indexar passa a significar também garantir visibilidade em sistemas automatizados, escalabilidade computacional e confiabilidade de identificadores em ambientes massivos de circulação científica.

Um primeiro aspecto dessa transformação aparece no trabalho de Rorissa (2010), que insere de modo explícito as *folksonomias* e o "*social tagging"* no debate sobre indexação de imagens. As relações "*social tagging <instrumentFor> image indexing*", "*user <agentOf> social tagging*" e "*image attribute <targetOf> indexing*" mostram que o papel de indexador já não pertence exclusivamente ao profissional da informação, mas é parcialmente redistribuído entre usuários comuns que descrevem recursos em ambientes colaborativos da Web 2.0. Ao mesmo tempo, essa abertura reintroduz o problema clássico da OC entre controle e liberdade terminológica, pois o ganho em diversidade semântica vem acompanhado de instabilidade vocabular, variação denominativa e menor padronização.

Arlitsch e O'Brien (2012) demonstram que os repositórios institucionais podem permanecer "invisíveis" quando seus esquemas de metadados não dialogam adequadamente com os requisitos de coleta e leitura de motores de busca acadêmicos como o *Google Scholar*. As relações "*metadata schema <instrumentFor> search engine indexing*", "*Google Scholar <agentOf> repository indexing*" e "*institutional repository <targetOf> indexing"* indicam que a indexação, nesse contexto, já não é apenas uma escolha semântica interna ao sistema documental, mas uma condição de legibilidade para agentes automatizados externos.

Nesse mesmo período, McCreadie, Macdonald e Ounis (2012) levam a discussão a outro patamar ao pensar a indexação como problema de escalabilidade computacional. As relações "*MapReduce <instrumentFor> large-scale indexing*", "*posting list <productOf> inverted indexing*" e "*computational scalability*

<isPropertyOf> distributed indexing” mostram que o desafio já não consiste somente em determinar quais termos melhor representam um documento, mas em construir índices capazes de operar sobre coleções em escala de *terabytes*. O *frame* de “*indexing*” torna-se, assim, inseparável da engenharia de sistemas distribuídos, da eficiência de processamento e da arquitetura dos índices invertidos. Se nos anos 1970 e 1980 a automação buscava aproximar a máquina da decisão terminológica humana, na década de 2010 a questão passa a ser como sustentar materialmente a indexação em ambientes massivos, heterogêneos e continuamente atualizados. Há, portanto, um deslocamento do foco semântico para o foco infraestrutural, sem que o semântico desapareça: ele passa a depender das condições técnicas de processamento em larga escala.

Outro desdobramento decisivo do período aparece na crescente associação entre indexação e avaliação científica. Torres-Salinas *et al.* (2014), ao discutirem a indexação de monografias e editoras científicas no *Book Citation Index*, mostram que “*monograph <targetOf> database indexing*”, “*scientific publisher <agentOf> monograph publishing*” e “*Book Citation Index <instrumentFor> research evaluation*”. Indexar, aqui, significa inserir livros e editoras em um circuito de mensuração de impacto, especialização temática e visibilidade institucional. A indexação deixa de operar apenas como mediação entre documento e usuário para atuar também como mediação entre produção científica e regimes de avaliação. Esse ponto é particularmente importante porque mostra que a indexação participa da constituição do valor acadêmico: aquilo que é indexado passa a ser contado, comparado, ranqueado e legitimado em métricas de desempenho científico.

Franceschini, Maisano e Mastrogiacomo (2015) radicalizam esse problema ao demonstrar como erros na indexação de DOI comprometem diretamente a confiabilidade das bases bibliométricas. As relações “*DOI <instrumentFor> paper disambiguation*”, “*multiple DOI assignment <isPropertyOf> database error*” e “*scientific paper <targetOf> DOI indexing*” mostram que o identificador único se converte em peça central da infraestrutura contemporânea de indexação. Nesse cenário, a indexação não é apenas temática, lexical ou conceitual; ela é também identificadora, isto é, responsável por garantir a individualização inequívoca de objetos documentais em bases interoperáveis. Quando essa camada falha, não se produz apenas ruído técnico, mas distorções de visibilidade, contagem de citações, atribuição de autoria e avaliação institucional.

Tomados em conjunto, esses trabalhos indicam que a década de 2010 fragmenta e expande o *frame* de "*indexing*" em pelo menos quatro direções complementares. A primeira é a direção participativa, em que usuários comuns se tornam agentes de descrição por meio de *tags* e práticas colaborativas. A segunda é a direção algorítmico-infraestrutural, em que a indexação depende da legibilidade dos metadados para motores de busca e sistemas automatizados. A terceira é a direção computacional em larga escala, na qual a construção de índices passa a ser indissociável de estratégias de processamento distribuído e eficiência técnica. A quarta é a direção bibliométrico-avaliativa, em que indexar significa também atribuir identidade persistente, visibilidade e mensurabilidade aos produtos científicos.

6.1.6 Anos 2020–2025: modelos combinados de IA e a resistência das Humanidades

No período mais recente, a indexação assume um caráter explicitamente híbrido entre processos humanos e automáticos, com centralidade crescente das técnicas de aprendizado de máquina. Golub (2021), em "*Automated Subject Indexing: An Overview*", define de forma inequívoca que "*machine learning <instrumentFor> automated subject indexing*" e "*human intellectual process <isReplacedBy> automated technique*", ao descrever a substituição de etapas de leitura, análise conceitual e tradução em linguagem de indexação por métodos de linguística computacional e estatística. O *frame* de "*indexing*" é reconfigurado como uma cadeia de operações algorítmicas cujo objetivo é simular, em alguma medida, a atribuição de assuntos realizada por especialistas.

O desenvolvimento de ferramentas como Annif e Finto AI reforça essa tendência. Em Suominen *et al*., "*ensemble algorithm <instrumentFor> automated subject indexing*" e "*text classification <isConditionFor> automated indexing*" indicam que a aposta não recai mais em um único modelo, mas em combinações de abordagens lexicais, associativas e redes neurais. A qualidade da indexação torna-se, ela própria, objeto de experimentação, com "*indexing quality <patientOf> evaluation*" em diferentes configurações de "*ensemble*". Nesse cenário, a indexação é concebida como um problema de engenharia de modelos: escolher, treinar e combinar classificadores para maximizar desempenho sobre *corpora* específicos, frequentemente ancorados em SOC preexistentes.

Contudo, Golub *et al*. chamam atenção para um contraponto importante ao analisar indexação de assunto nas humanidades. A relação "*high granularity <isPropertyOf> humanities indexing*" e "*faceted vocabulary <instrumentFor> humanities indexing*" mostra que, nesse campo, a indexação automática e os serviços comerciais genéricos tendem a falhar em capturar as necessidades de especificidade e granularidade exigidas por pesquisadores. As Humanidades expõem uma "resistência semântica": ambiguidade, polissemia, metáfora e exigências de contextualização histórica e geográfica exigem decisões que não se deixam reduzir facilmente a padrões estatísticos ou modelos treinados sobre grandes volumes de dados. O *frame*, aqui, reencena a centralidade da leitura especializada e de vocabulários facetados profundos, sugerindo que a OC não pode ser inteiramente delegada à máquina.

Do ponto de vista da convergência entre OC e IA, esse período evidencia uma virada: em vez de opor indexação manual e automática, os trabalhos passam a conceber um continuum em que sistemas automatizados são treinados, avaliados e corrigidos à luz de práticas e estruturas da OC. Ferramentas de IA não operam no vazio, mas ancoradas em tesauros, esquemas de classificação e políticas de indexação pré-existentes, reforçando a importância de modelos relacionais explícitos. É nesse contexto que um tesauro-dicionário com relações associativas tipadas, como o que sua pesquisa visa a fundamentar, se torna particularmente relevante: automatizar a indexação passa a depender da qualidade das relações paradigmáticas codificadas nos SOC.

6.1.7 Consequências epistemológicas e implicações para um tesauro-dicionário

Ao percorrer essa trajetória, é possível ver que o *frame* de "*indexing*" se fragmenta em múltiplos subframes: probabilístico-estatístico, cognitivo-interpretativo, sintático-frasal, latente-semântico, imagético, bibliométrico, automatizado por aprendizado de máquina, democrático por *social tagging*, entre outros. Cada um desses subframes traz consigo um conjunto de relações associativas específicas: "*probabilistic model <instrumentFor> automatic indexing*", "*human indexer <agentOf> manual indexing*", "*image attribute <targetOf> image indexing*", "*h-b index <instrumentFor> topic indexing*", "*ensemble algorithm <instrumentFor> automated*

subject indexing", "*social tagging <enhances> subject indexing*", para citar apenas algumas. A indexação aparece, assim, como um nó em que se enredam diferentes concepções de documento, de usuário, de conhecimento e de tecnologia.

Isso significa que não existe uma "indexação em si" descolada de teorias do conhecimento, da linguagem e da sociedade. Quando um sistema privilegia pesos estatísticos, assume implicitamente uma visão de relevância ligada à frequência e à coocorrência; quando enfatiza consistência inter-indexador, apoia-se em ideais de normalização cognitiva; quando se ancora em citações, aceita que os próprios atores científicos indexem o campo ao selecionar o que citam; incorpora modelos que refletem tanto os dados de treino quanto as estruturas de SOC que servem de alvo. Em todos os casos, as relações associativas tipadas em torno de indexação materializam essas escolhas epistemológicas, tornando visíveis dependências instrumentais, finalidades, condicionantes normativos e fundamentos teóricos.

Para o projeto de transformação do "Diccionario de Organización del Conocimiento" em tesauro-dicionário, a análise histórica da literatura de indexação sugere que a simples categoria genérica "ter relacionado" é insuficiente. A diversidade de *frames* revela a necessidade de uma tipologia relacional mais fina. Ao mapear, a partir do *corpus*, como essas relações se articulam em torno de "*indexing*" ao longo do tempo, sua pesquisa produz um inventário relacional que não é apenas descritivo, mas normativo e projetivo: ele oferece um modelo de como a OC pode representar, de forma mais transparente, os compromissos teóricos embutidos em suas práticas.

Nesse sentido, o catálogo de relações associativas tipadas construídas não é apenas um "complemento" ao dicionário, mas um dispositivo de alinhamento entre OC e IA. Ao explicitar, em linguagem de *frames*, quais são os papéis e vínculos que estruturam o conceito de indexação, esse modelo se torna diretamente reutilizável em ontologias, esquemas de dados e sistemas automatizados de indexação, facilitando a interoperabilidade semântica. Em vez de a IA aprender "do zero" quais conceitos se associam a indexação, ela pode ser instruída por um tesauro-dicionário em que essas relações foram previamente analisadas, tipificadas e justificadas em perspectiva de Análise de Domínio (AD).

6.2 Eixos analíticos do *corpus* ampliado

A análise do *corpus* 2, obtido por busca temática ampliada na *Web of Science*, desloca o foco da trajetória histórica de "*indexing*" para uma cartografia conceitual da OC e de seus regimes de indexação. Diferentemente da primeira rodada de testes, em que os 28 artigos do *corpus* foram enviados em um único bloco ao NotebookLM e resultaram em respostas mais gerais e pouco densas em termos de relações tipadas, a análise aqui apresentada decorre da reconfiguração do *corpus* em blocos menores, o que tornou as respostas mais específicas e aumentou a visibilidade de relações associativas entre indexação e outros elementos. Essa estratégia, simétrica à utilizada no *corpus* 1, ampliou significativamente o número e a clareza das relações associativas identificadas, permitindo um tratamento analítico mais fino do *corpus* ampliado.

Em vez de seguir apenas a cronologia, esta etapa evidencia eixos analíticos que atravessam os *frames* construídos no *corpus* 2. Esses eixos se deixam ver nas próprias relações tipadas que emergem dos textos, tais como *"indexing <identifies> epistemological potentialities", "request-oriented indexing <supports> given goals", "indexing <uses> document properties", "semantic indexing <generates> semantic metadata", "social tagging <enhances> subject indexing", "indexing <cannotBe> neutral", "universal KOS <marginalizes> feminist knowledge"* ou *"automatic indexing <isPartOf> information processing paradigm"*. Com base nessas relações concretas, é possível agrupar o *corpus* em quatro grandes eixos: a) indexação como problema epistemológico (Hjørland), b) indexação orientada por SOC e ontologias (Tudhope, Golub e colaboradores), c) dimensão ética-discursiva da indexação (Guimarães, Samuelsson e coautores) e d) práticas híbridas de indexação em contextos de *folksonomias*, categorias de nível básico e ambientes digitais.

6.2.1 Hjørland e a indexação como problema epistemológico

Os textos de Hjørland presentes no *corpus* 2 *("Fundamentals of Knowledge Organization", "What is Knowledge Organization (KO)?", "The Importance of Theories of Knowledge: Indexing and Information Retrieval as an Example", "Reviews of Concepts in KO: Subject (of Documents)"* e *"Indexing: Concepts and Theory*") configuram um eixo em que a indexação é tematizada como questão epistemológica, e não apenas como técnica de descrição. Na quadro 7, por exemplo, "*Fundamentals of Knowledge Organization*" é enquadrado no *frame* no

qual a indexação aparece como um processo voltado a identificar o potencial informativo e epistemológico de um documento para auxiliar autores futuros, expresso em relações como "*indexing <identifies> epistemological potentialities*" e "*indexing <helps> future authors*".

Em "The Importance of Theories of Knowledge…", a indexação é descrita por relações como "*indexing <evaluates> document contribution to knowledge*" e "request-oriented indexing <supports> given goals", que situam o ato de indexar como avaliação de potenciais informativos em função de objetivos e demandas específicas. "*What is Knowledge Organization (KO)?*" reforça essa perspectiva ao apresentar "*indexing <isA> knowledge organizing process*" e "*indexing <involves> semantic labor"*, enfatizando o trabalho interpretativo e semântico. Já "*Reviews of Concepts in KO: Subject (of Documents)*" explicita a decomposição processual da indexação com relações como "*indexing <consistsOf> conceptual analysis*" e "*indexing <consistsOf> translation stage*", distinguindo a análise conceitual da etapa de tradução em rótulos.

No conjunto, essas relações tipadas constroem um *frame* em que a indexação é orientada por finalidades (ajudar futuros autores, apoiar determinados objetivos), dependente de teorias do conhecimento e de critérios de relevância e organizada em etapas analíticas distintas. O eixo epistemológico do *corpus* ampliado, portanto, não é uma abstração imposta; ele emerge diretamente das relações associativas que ligam indexação a potenciais epistemológicos, metas pragmáticas, trabalho semântico e estruturas de análise de assunto.

6.2.2 Tudhope, Golub e a indexação orientada por SOC e ontologias

Um segundo conjunto de textos enfatiza a indexação como prática acoplada a sistemas formais de OC, ontologias e infraestruturas terminológicas. Na quadro 8, o artigo de Vlachidis & Tudhope é enquadrado no *frame* em que a indexação semântica automatizada aparece em relações como "*semantic indexing <generates> semantic metadata", "semantic indexing <supports> semantic queries"* e *"information extraction <usedFor> semantic indexing"*. Aqui, a indexação é apresentada como processo que gera metadados semânticos e dá suporte a consultas complexas, sustentada por ontologias de domínio e regras de extração de informação.

Textos de Golub e colaboradores, especialmente "*Terminology Registries for Knowledge Organization Systems*" (quadro 10), constroem o *frame* com relações como *"terminology services <support> indexing", "automated indexing <assigns> terms from vocabularies"* e *"KOS <usedFor> indexing"*. Essas relações situam a indexação em um ecossistema em que serviços terminológicos e registros de vocabulário fornecem termos, identificadores e ligações que tornam possível a atribuição automática e a criação de metadados interoperáveis.

Em Matthews et al. (quadro 10), o *frame* mostra como a indexação tradicional é combinada com marcação social, por meio de relações como *"traditional subject indexing <incurs> high costs", "social tagging <reduces> indexing costs"* e *"vocabulary-based suggestions <encourage> indexing activities*". Já Tennis e Sutton enfatizam "*SKOS model <supports> indexing", "abstract concepts <appliedIn> indexing processes"* e *"indexers <employ> concepts*", articulando indexação à gestão histórica e estruturada de conceitos na *Web Semântica*.

Esse eixo tecnológico-semântico do *corpus* ampliado, portanto, se materializa em relações associativas que expressam dependência funcional de ontologias, SOC e serviços terminológicos, geração de metadados e contexto semântico e articulação entre práticas humanas e automação. Essas relações servem de base para pensar, no projeto do tesauro-dicionário, categorias associativas que sejam reaproveitáveis em grafos semânticos e vocabulários estruturados.

6.2.3 Guimarães, Samuelsson e a dimensão ética-discursiva da indexação

Um terceiro bloco de textos evidencia a indexação como prática discursiva com implicações éticas e políticas. Em "*Male Homosexuality in Brazilian Indexing Languages: Some Ethical Questions*" (quadro 12), o *frame* é construído por relações como "*librarians <perform> indexing", "indexing <influencedBy> cultural context"* e *"indexing languages <represent> subjects*". Essas relações explicitam que indexadores, ao criar rótulos, são influenciados por contextos culturais e ideológicos, e que as línguas de indexação têm papel ativo na configuração da representação de sujeitos.

No artigo "*The Representation of Ethics and Knowledge Organization in the WoS and LISTA Databases*", o *frame* traz relações como inconsistencies "*in indexing <misrepresent> domain", "indexing policies <cause> omissions"* e

“databases <perform> indexing”. Aqui, a indexação aparece como dispositivo que pode distorcer a percepção de um domínio por meio de omissões e inconsistências, vinculando práticas de indexação a efeitos de visibilidade e ocultamento.

Samuelsson, em “*Knowledge Organization for Feminism and Feminist Research*” (quadro 11), alimenta o *frame* com relações como “*indexing <shapes> meaning”, “universal KOS <marginalizes> feminist knowledge”* e *“feminist research <requires> context-sensitive indexing”*. Essas relações tornam explícito que sistemas universais podem marginalizar determinados saberes e que a indexação molda significados ao decidir quais temas e perspectivas serão destacados. Fujita *et al.* (2023) enfatizam relações como *“indexing policy <improves> information retrieval”, “controlled vocabulary <standardizes> subject indexing”* e *“institutional repositories <require> indexing policy”*, ligando diretamente políticas de indexação a precisão, padronização e visibilidade institucional.

Em conjunto, essas relações associativas tipadas mostram que o eixo ético-discursivo do *corpus* 2 não é apenas um tema de discussão teórica, mas está inscrito nas próprias formas de dizer o que a indexação faz, de que é influenciada e que efeitos produz. Para o tesauro-dicionário proposto, isso indica a necessidade de categorias relacionais capazes de explicitar dependências contextuais, marginalização e distorção representacional, em vez de esconder essas dinâmicas sob relações associativas genéricas.

6.2.4 *Folksonomias*, categorias de nível básico e práticas híbridas

Por fim, o *corpus* ampliado inclui um conjunto de trabalhos que exploram formas híbridas de indexação, nas fronteiras entre práticas profissionais, contribuições de usuários e modelos cognitivos. No quadro 8, o artigo de Hajibayova apresenta relações como “*indexing <uses> basic-level terms”, “basic-level categories <supports> image indexing”* e *“user-centered indexing <reliesOn> crosswalks”*, enfatizando o uso de categorias de nível básico para representar o conteúdo conceitual de recursos, especialmente imagens.

Textos sobre *folksonomias* e marcação social, como Wu *et al.*, Stvilia & Jörgensen e Kakali e Papatheodorou, distribuem-se nos quadros 8, 9 e 11, em *frames* como “Indexação Democrática e Folksonomias”, “Indexação de Imagens e Folksonomia” e “Indexação Colaborativa / Social Tagging”. As relações associativas

correspondentes incluem “*democratic indexing <determines> authority”, “users <perform> democratic indexing”, “social tagging <isA> democratic indexing”, “image indexing <uses> controlled vocabularies”, “nonexpert-generated tags <usedFor> image indexing”, “indexing <basedOn> end-user information needs”, “social tagging <enhances> subject indexing”, “collaborative tagging <provides> index terms”* e *“users <participateIn> subject indexing”*. Essas relações mostram a indexação como prática partilhada e negociada entre especialistas, sistemas e usuários, na qual *tags* de usuários podem complementar ou tensionar a indexação profissional, enquanto vocabulários controlados oferecem estruturas de ancoragem.

Além disso, trabalhos como os de Saumure e Shiri situam essas práticas em um panorama histórico em que “*computers <perform> automatic indexing”, “indexing <organizes> digital information”* e *“machine-assisted indexing <extracts> semantic context”*, destacando a passagem de uma indexação predominantemente manual para regimes híbridos em ambientes digitais. No conjunto, as relações tipadas que aparecem nesses artigos configuram um eixo híbrido em que a indexação é simultaneamente cognitiva, social e maquínica, articulando categorias de nível básico, *folksonomias*, controle vocabular e extração automática em diferentes combinações.

6.2.5 Integração: do ecossistema ampliado ao modelo relacional do tesauro-dicionário

A leitura do *corpus* temático ampliado reforça e complexifica a análise histórica por décadas. Enquanto o primeiro *corpus* mostrou como o *frame* de “*indexing*” se fragmenta em subframes probabilísticos, cognitivos, imagéticos, bibliométricos e algorítmicos, o segundo revela que esses subframes se articulam a escolas teóricas, infraestruturas tecnológicas e agendas ético-políticas distintas. Em termos de AD, pode-se dizer que o conceito de indexação está situado em um campo tensionado por pelo menos três grandes “formas de conhecer”: uma forma epistemológico-crítica (Hjørland), uma forma tecnológico-semântica (Tudhope, Golub, Tennis) e uma forma ético-discursiva (Guimarães, Samuelsson, Pinho).

Essa integração tem consequências diretas para o modelo de relações associativas que a dissertação propõe. No plano empírico, o *corpus* 2 mostra que a indexação é recorrentemente descrita por relações como “*indexing <identifies>*

epistemological potentialities”, “*request-oriented indexing <supports> given goals”*, “*indexing <uses> document properties”*, “*indexing <reliesOn> controlled vocabularies”*, “*semantic indexing <generates> semantic metadata”*, “*terminology services <support> indexing”*, “*indexing <cannotBe> neutral”*, “*indexing <influencedBy> cultural context”*, “*universal KOS <marginalizes> feminist knowledge”*, “*social tagging <enhances> subject indexing”* ou “*automatic indexing <isPartOf> information processing paradigm”*. Em vez de introduzir categorias formais que não aparecem nos dados, o modelo relacional partiu dessas relações concretas e as generalizou em tipos mais amplos (dependência instrumental, apoio funcional, influência epistemológica, marginalização discursiva, decomposição processual etc.), que serão reutilizados no tesauro-dicionário.

A convergência entre OC e IA observada na literatura, tanto na automação da indexação quanto na construção de ontologias e registros terminológicos, sugere que o tesauro-dicionário resultante deve ser pensado desde o início como recurso aplicável por agentes automáticos. Trabalhos sobre indexação semântica e serviços terminológicos mostram que a eficácia da automação depende da qualidade e da expressividade dos SOC subjacentes, o que se reflete em relações como “*semantic indexing <supports> semantic queries”*, “*machine-assisted indexing <extracts> semantic context”* ou “*automated indexing <assigns> terms from vocabularies”*. Isso significa que as relações associativas tipadas precisam ser definidas de forma clara e operacional, para que possam orientar tanto o trabalho de indexadores humanos quanto o de sistemas de IA na interpretação de textos, na sugestão de descritores e na navegação semântica.

A presença, no *corpus* ampliado, de textos que problematizam *folksonomias*, categorias de nível básico, feminismo, sexualidade e ética indica que o futuro tesauro-dicionário não pode ser concebido apenas como instrumento de normalização terminológica. Relações como “*indexing <shapes> meaning”*, “*inconsistencies in indexing <misrepresent> domain”*, “*indexing <basedOn> end-user information needs”*, “*democratic indexing <determines> authority”* ou “*feminist research <requires> context-sensitive indexing”* mostram que a indexação participa de disputas de visibilidade, autoridade e sentido. O tesauro-dicionário, portanto, precisa funcionar também como espaço de explicitação de tensões e alternativas, permitindo que diferentes perspectivas epistemológicas e discursivas

encontrem representação por meio de relações que tornem visíveis influências contextuais e efeitos de marginalização.

6.3 Papéis Qualia

Ao reunir as relações associativas tipadas de todos os quadros, é possível observar que o *frame* de "*indexing*" reconstruído nesta pesquisa se apoia sobretudo em três papéis Qualia: Agentivo, Constitutivo e Télico. O papel Agentivo aparece quando a indexação é descrita como ação de alguém ou de alguma instância sobre algo, como em "*indexer <evaluates> document contribution to knowledge"*, "*users <perform> democratic indexing"*, "*computers <perform> automatic indexing"* ou "*terminology services <support> indexing"*, que evidenciam a centralidade de agentes humanos, coletivos, técnicos e institucionais na atribuição de assuntos e no suporte à atividade. O papel Constitutivo se manifesta em relações que descrevem propriedades, componentes ou alvos da indexação, por exemplo "*indexing <consistsOf> conceptual analysis"*, "*indexing <consistsOf> translation stage"*, "*semantic indexing <generates> semantic metadata"* ou "*indexing <organizes> digital information"*, indicando do que o processo é feito, quais partes o compõem e sobre o que ele recai.

O papel Télico, por sua vez, concentra as relações em que instrumentos, técnicas e sistemas são apresentados como meios para alcançar determinados fins, como em "*semantic indexing <supports> semantic queries"*, "*provenance method <usedFor> content indexing"*, "*social tagging <enhances> subject indexing"*, "*manual indexing <assists> searchers"*, "*indexing <supports> knowledge transfer"* ou "*social tagging <reduces> indexing costs"*. Esses exemplos mostram que a literatura analisa a indexação sempre em função de objetivos práticos de recuperação, descoberta, avaliação, visibilidade e economia de recursos, e que os papéis Qualia permitem tornar explícito se uma relação está falando de "quem age", "do que é feito" ou "para que serve" no ecossistema de indexação modelado por esta pesquisa.

Além desses três papéis mais recorrentes, a análise também identifica, ainda que com menor frequência, papéis de cunho mais Formal e Processual. Relações como "*automatic indexing <isPartOf> information processing paradigm"* situam a indexação em relação a outras atividades e paradigmas da área, enquanto relações

de decomposição interna, como "*subject retrieval <proceedsByLinking> indexing with provenance"* ou "*abstract concepts <appliedIn> indexing processes"*, indicam que a literatura não vê a indexação como um ato único e instantâneo, mas como cadeia de operações encadeadas em fluxos de trabalho mais amplos. Esses papéis menos frequentes são importantes porque aproximam a modelagem de uma visão mais fina da prática profissional e abrem espaço para, em versões futuras do tesauro-dicionário, representar não só o que a indexação faz e com o que ela opera, mas também que lugar ela ocupa em arquiteturas de representação mais complexas

A seguir, apresenta-se um quadro que sintetiza os principais tipos de relações associativas identificadas e o papel Qualia predominante em cada uma delas.

Quadro 13 – Síntese das relações associativas tipadas do *frame* "*indexing"*

Tipo de relação tipada	**Papel Qualia predominante**	**Definição resumida**	**Exemplo (artigo)**
agentOf	Agentivo	Liga o agente (humano ou sistema) ao processo de indexação que ele executa.	human indexer <agentOf> manual indexing / indexador humano realiza a indexação em MEDLINE (Funk e Reid, 1983).
agentOf	Agentivo	Liga o usuário comum à prática de marcação social.	user <agentOf> social tagging / usuários passam a atuar como indexadores em ambientes colaborativos (Rorissa, 2010).

instrumentFor	Télico	Indica instrumentos, técnicas ou modelos usados para realizar indexação.	statistical rule <instrumentFor> automatic keyword indexing / regras estatísticas sustentam a indexação automática por palavras-chave (Harter, 1975).
instrumentFor	Télico	Liga recursos de SOC ou serviços terminológicos à indexação.	domain-specific thesaurus <instrumentFor> keyphrase indexing / tesauros de domínio orientam a atribuição automática de termos (Medelyan e Witten, 2008).
instrumentFor	Télico	Representa tecnologias e infraestruturas computacionais usadas para indexar em larga escala.	MapReduce <instrumentFor> large-scale indexing / o paradigma MapReduce viabiliza indexação de corpora em escala de terabytes (McCreadie *et al.*, 2012).
instrumentFor	Télico	Relaciona esquemas de metadados ao sucesso da indexação em buscadores.	metadata schema <instrumentFor> search engine indexing / esquemas de metadados adequados permitem a coleta por Google Scholar (Arlitsch e O’Brien, 2012).

instrumentFor	Télico	Liga algoritmos de "*ensemble"* à indexação automática assistida por SOC.	ensemble algorithm <instrumentFor> automated subject indexing / combinações de algoritmos suportam indexação temática com Annif/Finto AI (Suominen *et al.*, 2022).
isPropertyOf	Constitutivo	Expressa qualidades ou atributos do processo de indexação.	indexing quality <isPropertyOf> indexing / qualidade é atributo central da indexação (Rolling, 1981).
isPropertyOf	Constitutivo	Caracteriza atributos de granularidade e especificidade.	high granularity <isPropertyOf> humanities indexing / alta granularidade é propriedade desejada em indexação em Humanidades (Golub *et al.*, 2020).
isPropertyOf	Constitutivo	Relaciona métricas de impacto a objetos indexados.	citation impact <isPropertyOf> indexed publication / impacto de citação caracteriza publicações que funcionam como documentos-núcleo em estratégias de busca (Glänzel, 2002).

targetOf	Constitutivo	Indica o objeto diretamente indexado (documentos, tópicos, atributos).	scientific paper <targetOf> DOI indexing / artigos científicos são alvo da indexação por DOI em bases bibliométricas (Franceschini *et al.*, 2015).
targetOf	Constitutivo	Indica objetos institucionais ou coleções como alvo da indexação.	institutional repository <targetOf> indexing – repositórios institucionais são alvo de indexação por motores de busca acadêmicos (Arlitsch e O'Brien, 2012).
targetOf	Constitutivo	Relaciona atributos de imagens ao processo de indexação imagética.	image attribute <targetOf> image indexing / atributos genéricos e específicos de imagens são objeto da decisão de indexação (Shatford Layne, 1994).
productOf	Processual	Representa resultados gerados pela indexação ou por técnicas específicas.	complex descriptor <productOf> phrase indexing / descritores complexos resultam da indexação por frases (Fagan, 1989).
productOf	Processual	Indica estruturas derivadas do processo de indexação.	posting list <productOf> inverted indexing / listas de postagem são produtos da construção de índices

			invertidos (McCreadie *et al.*, 2012).
isConditionFor	Processual	Liga condições necessárias à possibilidade ou eficácia da indexação.	term co-occurrence <isConditionFor> latent semantic indexing / coocorrência de termos é condição para a LSI (Kontostathis e Pottenger, 2006).
isConditionFor	Processual	Relaciona lacunas de informação a decisões de indexação.	information gap <isConditionFor> subject indexing / lacunas de conhecimento dos usuários condicionam a escolha de termos pelo indexador (Bates, 1998).
measures	Télico	Liga indicadores e métricas aos aspectos da indexação que quantificam.	inter-indexer consistency <measures> indexing quality / consistência entre indexadores mede a qualidade da indexação (Rolling, 1981).
basisFor	Télico	Indica fundamentos conceituais ou necessidades que orientam a indexação.	end-user information needs <basisFor> indexing documents / necessidades informacionais dos usuários fundamentam a indexação de imagens (Stvilia e Jörgensen, 2010).

basisFor	Télico	Liga perspectivas teóricas a práticas de indexação.	epistemological theory <basisFor> indexing practice / teorias de domínio sustentam escolhas de indexação (Hjørland, diversos textos).
enhances	Télico	Representa recursos ou práticas que reforçam ou complementam a indexação.	social tagging <enhances> subject indexing / marcação social complementa a indexação temática tradicional (Rorissa, 2010; Matthews *et al.*, 2010).
supports	Télico	Liga serviços terminológicos e registros SOC ao trabalho de indexação.	terminology registry <supports> KOS management / registros terminológicos apoiam a gestão de SOC usados na indexação (Golub *et al.*, 2014).
influences	Constitutivo	Expressa dependências conceituais entre indexação e contextos epistemológicos ou institucionais.	epistemological assumptions <influence> indexing practice / pressupostos epistemológicos de OC influenciam políticas de indexação (Hjørland, 2011; 2020).

sharesObjectiveWith	Télico	Indica processos ou sistemas que compartilham objetivos com a indexação.	automated subject indexing <sharesObjectiveWith> manual indexing / indexação automática e manual compartilham o objetivo de representar o assunto com precisão (Golub, 2021).

6.4 "*Indexing*" e a costura entre ramos da árvore de domínio

A partir desse conjunto de relações associativas tipadas, foi possível retornar à árvore de domínio do "Diccionario de Organización del Conocimiento" e explicitar como o termo "*indexing*" se articula com diferentes ramos da OC. Nesse sentido, o quadro a seguir sintetiza um subconjunto representativo de relações em que a indexação se conecta a teorias, campos relacionados, SOC, agentes e contextos socioculturais, evidenciando que o processo de indexar atravessa os níveis conceitual, procedimental e infraestrutural do domínio.

Quadro 14 – Relações de "*indexing*" na árvore de domínio do dicionário

Relação analisada	Relação tipada	Ramo da árvore (termo relacionado)	Tipo de ligação entre ramos
Indexing – Indexer	<agentOf>	5. Persons (Indexer)	Pessoas (5) → Processos (2.3 Indexing)
Indexing – Subject heading	<instrumentFor>	3.2.1 Types of KOS (Subject headings list)	Sistemas (3.2.1) → Processos (2.3)
Indexing – Controlled vocabularies	<reliesOn>	3.2.1 Types of KOS (Controlled vocabulary / Thesaurus)	Processos (2.3) ↔ Sistemas (3.2.1)

Indexing – Domain analysis	<isBasedOn>	1.1 Fundaments and theories (Domain analysis)	Processos (2.3) ↔ Teorias (1.1)
Indexing – Epistemological potentialities	<identifies>	1.1 Fundaments and theories	Processos (2.3) → Teorias / potenciais (1.1)
Indexing – Information retrieval	<hasGoalOrProcess>	1.3 Related fields (Information retrieval)	Processos (2.3) ↔ Campos relacionados (1.3)
Indexing – Statistical rule / model	<instrumentFor>	1.3 Related fields (probabilistic IR)	Modelos externos (1.3) → Processos (2.3)
Indexing – Machine learning / ensemble	<instrumentFor>	1.3 Related fields (Artificial Intelligence)	IA (1.3) → Processos (2.3)
Indexing – Semantic indexing / metadata	<generates>	3.4 Representation (Semantic metadata)	Processos (2.3) → Elementos de representação (3.4)
Indexing – Terminology services / KOS	<supportedBy>	3.2.1 Types of KOS (terminology services, KOS)	Processos (2.3) ↔ Sistemas / serviços (3.2.1)
Indexing – Indexing policy / warrants	<isControlledBy>	3.1.5 Warrant (Ethical warrant, User warrant)	Princípios / warrants (3.1.5) → Processos (2.3)
Indexing – Cultural context	<influencedBy>	(Contexto sociocultural ligado a 1.1/1.3)	Contexto sociocultural (1.x) → Processos (2.3)

Indexing – Social tagging / democratic indexing	<enhancedBy>	2.3 Indexing / 3.2.1 Folksonomy	Processos (2.3) ↔ Práticas colaborativas / KOS (3.2.1)
Indexing – Image attribute	<targetOf>	1.2 Basic concepts (Document / Attribute)	Conceitos de documento (1.2) → Processos (2.3)
Indexing – DOI / identifiers	<instrumentFor>	1.3 Related fields (Bibliometrics / identification)	Infraestrutura identificadora (1.3) → Processos (2.3)

A análise dos relacionamentos permitiu retomar a árvore de domínio do "Diccionario de Organización del Conocimiento" e evidenciar como o termo "*indexing*" se articula com diferentes ramos desse domínio. Com base nos resultados, observa-se, em primeiro lugar, a presença de relações intra-ramo no âmbito de "*Knowledge organization processes (KOP)*" (ramo 2), especialmente entre "*indexing*", "*aboutness*" e "*subject analysis*", o que reforça o papel desses conceitos como núcleo procedimental da OC. Ao mesmo tempo, emergem relações inter-ramo que conectam "*indexing*" a diversos "*knowledge organization systems (KOS)*" (ramo 3), como "*classification system*", "*thesaurus*", "*taxonomy*" e "*subject headings list*", explicitando a dependência mútua entre processos de indexação e as linguagens e sistemas documentários de autoridade usados para representar os assuntos.

Para além da articulação entre processos e sistemas, os resultados também evidenciam vínculos entre "*indexing*" e métodos específicos (ramo 4, "*Methods*"), por exemplo no diálogo com "*vocabulary control*", "*faceted methodology*" e "*systematic indexing*", bem como com o exercício profissional representado por "*indexer*" e "*user*" no ramo 5, "*Persons*". Ao situar esses relacionamentos nos diferentes ramos da árvore, de "*Knowledge organization in general"* (ramo 1) e seus fundamentos teóricos, passando por "*Knowledge organization processes (KOP)"* (ramo 2), "*Knowledge organization systems (KOS)*" (ramo 3) e "*Methods"* (ramo 4), até chegar a "*Persons*" (ramo 5) –, torna-se possível fechar o ciclo da pesquisa, mostrando que "*indexing*" não se limita a um procedimento isolado, mas constitui um ponto de

convergência entre teorias, processos, sistemas, métodos e agentes que estrutura a própria configuração do domínio da OC.

7 CONCLUSÕES

As análises realizadas com apoio de modelos de linguagem evidenciaram limites e potencialidades do uso de IA na Organização do Conhecimento (OC), reforçando o caráter situado de qualquer recurso desse tipo quando aplicado à leitura de *corpus* científicos. Em ambos os *corpora*, o NotebookLM foi operado com o mesmo *prompt* analítico, fundamentado em Ciência da Informação, OC, Teoria de *Frames* e Terminologia Baseada em *Frames*, e recebeu os artigos em blocos menores (cinco textos por vez, e três ao final), o que permitiu manter a consistência da extração de *frames* e relações associativas tipadas ao longo de todo o conjunto. Essa simetria metodológica mostra que a qualidade das respostas da IA depende diretamente de decisões de projeto de *corpus*, de segmentação e de formulação de *prompts*, e não apenas das capacidades "genéricas" do modelo, exigindo curadoria cuidadosa e interpretação crítica por parte da pesquisadora.

A articulação entre o *corpus* histórico, organizado por décadas, e o *corpus* temático ampliado foi decisiva para compreender a evolução e a complexificação do *frame* de "*indexing"* no campo da OC. O primeiro *corpus* permitiu reconstruir, em perspectiva diacrônica, a passagem de modelos probabilísticos e estatísticos para abordagens sintático-lexicais, sistemas baseados em conhecimento, indexação semântica latente, métricas bibliométricas e infraestruturas de processamento distribuído. O segundo *corpus*, por sua vez, deslocou o foco para debates epistemológicos, ético-políticos e tecnológicos mais amplos, nos quais a indexação aparece articulada a teorias de domínio, sistemas de organização do conhecimento, folksonomias, pesquisas feministas, estudos de proveniência e políticas de indexação em repositórios institucionais. Em conjunto, os dois *corpora* mostram que não há um "núcleo duro" estável de "*indexing"*, mas uma sucessão de enquadramentos em que processos, agentes, instrumentos, critérios de qualidade e objetivos se reorganizam historicamente.

Ao integrar os resultados da tipificação das relações associativas com a árvore de domínio do "Diccionario de Organización del Conocimiento", a dissertação demonstra, de forma concreta, como o *frame* de "*indexing*" atravessa os principais ramos da OrC: parte de "*Knowledge organization in general*" e de seus fundamentos teóricos, percorre "*Knowledge organization processes (KOP)*", conecta-se a "*Knowledge organization systems (KOS)*" e a "*Methods*", e se materializa nas

práticas de “*Persons*” como “*indexers*” e “*users*”. Ao explicitar essas costuras no Quadro 14, em que relações como <*agentOf*>, <*instrumentFor*>, <*reliesOn*>, <*isControlledBy*> e <*enhancedBy*> ligam “*indexing*” a serviços terminológicos, métricas e contextos sociopolíticos, a pesquisa reforça a ideia central das conclusões: não há uma indexação neutra ou isolada, mas um nó relacional que condensa disputas epistemológicas, tecnológicas e éticas e que oferece base concreta para a futura implementação de um tesauro-dicionário da OC ancorado em *frames* e interoperável com aplicações de IA.

Do ponto de vista teórico, as análises permitiram desmontar, de forma concreta, a ilusão de neutralidade que historicamente permeia muitos sistemas de informação. Vários Sistema de Organização do Conhecimento (SOC) foram concebidos sob premissas empiristas ou racionalistas que supõem que a indexação é um espelho objetivo da realidade, capaz de representar uma verdade universal sobre os documentos. Contudo, *frames* como o da indexação orientada por questões, no contexto de pesquisas feministas, e o da indexação de assuntos relacionados à homossexualidade evidenciam, na prática, teorias pragmáticas e críticas da indexação: nomear a informação, escolher descritores e estabelecer relações não são atos neutros, mas atos de poder que favorecem certos recortes e usos da informação em detrimento de outros. Ao tipificar relações associativas em torno desses casos, o modelo de *frames* torna explícitas as intencionalidades e ideologias embutidas nos sistemas, mostrando que, por trás de cada técnica ou arquitetura de indexação (seja tesauro, ontologia ou folksonomia), há grupos sociais, épocas e visões de mundo específicas tentando dar sentido e impor uma certa ordem à realidade informacional.

Em conjunto, os quatro artigos brasileiros destacados na análise da busca ampliada, dentro do conjunto de 28 textos, refletem a maturidade de uma verdadeira “escola brasileira” de OC. Longe de se limitar à formulação de regras mecânicas para ordenar acervos, essa produção mostra uma OC comprometida com a análise dos impactos sociais da classificação (particularmente em temas éticos e de minorias), com a crítica aos vieses das plataformas e bases de dados, com a articulação interdisciplinar entre OC, arquivologia e estudos de proveniência e com a aplicação dessas reflexões à infraestrutura científica nacional, como nos repositórios institucionais. Esses trabalhos dão concretude à tese central defendida por autores como Hjørland: classificar e indexar são atos humanos, situados e atravessados por

intencionalidades sociais e políticas. Ao explicitar essas dimensões por meio de *frames* e relações associativas tipadas, a dissertação demonstra que um tesauro-dicionário da OC orientado por esse paradigma pode funcionar não apenas como instrumento técnico, mas como meio de expor, problematizar e negociar os valores que estruturam a OC.

Do ponto de vista metodológico, a pesquisa também demonstra que o uso de IA em estudos de OC não pode ser pensado como mera automação da leitura, mas como procedimento analítico que exige desenho rigoroso do *corpus*, formulação explícita de instruções e validação interpretativa contínua. A experiência com os dois *corpora* mostrou que a segmentação dos textos em blocos menores, aliada ao uso do mesmo *prompt* analítico e do mesmo referencial teórico, foi decisiva para aumentar a clareza e a densidade das relações associativas tipadas extraídas pelo NotebookLM. Isso reforça que, em pesquisas dessa natureza, o resultado não depende apenas da ferramenta utilizada, mas da articulação entre *corpus*, estratégia de entrada, teoria e revisão crítica da pesquisadora.

Em termos conceituais, a dissertação confirma que o *frame* de "*indexing"* não se limita à atribuição de termos a documentos, mas se expande em múltiplas direções: envolve agentes humanos e automatizados, instrumentos estatísticos, linguísticos e computacionais, critérios de qualidade, políticas institucionais, infraestruturas de descoberta e disputas epistemológicas e éticas. A modelagem das relações associativas tipadas permitiu tornar visível essa rede de dependências, oferecendo uma descrição mais fina do conceito do que aquela normalmente encontrada em relações associativas genéricas de tesauros tradicionais. Nesse sentido, o estudo contribui para aproximar a OC de abordagens semânticas mais explícitas, capazes de representar não apenas proximidade entre termos, mas a natureza específica dos vínculos que os conectam.

Como desdobramento, o modelo desenvolvido pode ser aplicado a outros conceitos centrais da OC, ampliando progressivamente o tesauro-dicionário e permitindo observar como diferentes *frames* se articulam em uma rede conceitual mais ampla. Além disso, futuras pesquisas podem comparar o comportamento de diferentes modelos de linguagem na extração de *frames* e relações tipadas, bem como testar o procedimento em outros domínios especializados, bases de dados ou conjuntos documentais multilíngues. Com isso, abre-se a possibilidade de consolidar uma metodologia replicável para a análise semântica de conceitos em

OC com apoio de IA, sem abrir mão do controle teórico e da responsabilidade interpretativa do pesquisador.

REFERÊNCIAS

ABRAHAMSEN, Knut Tore. Indexing of musical genres: an epistemological perspective. **Knowledge Organization**, v. 30, n. 3–4, 2003. Disponível em: https://www.researchgate.net/publication/287004548_Indexing_of_musical_genres_An_epistemological_perspective. Acesso em: 09 mar. 2026.

ALBRECHTSEN, Hanne. This is Not Domain Analysis. **Knowledge Organization**, v. 42, n.8, 2015. Disponível em: https://www.researchgate.net/publication/288642582_This_is_Not_Domain_Analysis. Acesso em: 09 mar. 2026.

ALMEIDA, Bruno. Terminologia e organização do conhecimento: linguagens, vocabulários e sistemas. **LinhaD'Água**: São Paulo, v. 34, n. 02, p. 26-46, mai.-ago. 2021. Disponível em: http://dx.doi.org/10.11606/issn.2236-4242.v34i2p26-46. Acesso em: 28 jan. 2025.

ANDERSEN, Jack. Written knowledge: a literary perspective on indexing theory. **Knowledge Organization**, v. 27, n. 4, 2000. Disponível em: https://www.researchgate.net/publication/292826024_Written_knowledge_A_literary_perspective_on_indexing_theory. Acesso em: 09 mar. 2026.

ARLITSCH, K.; O'BRIEN, P. S. Invisible institutional repositories: addressing the low indexing ratios of IRs in Google Scholar. **Library Hi Tech**, v. 62, n. 3, 2011. Disponível em: https://www.emerald.com/lht/article-abstract/30/1/60/264469/Invisible-institutional-repositoriesAddressing-the?redirectedFrom=fulltext. Acesso em: 10 mar. 2026.

BANKS, M. G. An extension of the Hirsch index: indexing scientific topics and compounds. **Scientometrics**, v. 57, n. 2, 2006. Disponível em: https://www.researchgate.net/publication/220364922_An_extension_of_the_Hirsch_Index_Indexing_scientific_topics_and_compounds. Acesso em: 10 mar. 2026.

BARITÉ, Mario. Organización del conocimiento: un nuevo marco teórico-conceptual en bibliotecologia y documentacion. CARRARA, Kester. **Educação, Universidade e Pesquisa**. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP. 2001.

BARITÉ, Mario (*et al*). **Diccionario de Organización del Conocimiento**: clasificación, indización, terminología. 6. ed. revisada y ampliada. Montevideo: CSIC, 2015.

BARITÉ, Mario. Los árboles de dominio. In: CATALÁ, Sara Álvares; BARITÉ, Mario (coord.). **Teoría y praxis en terminología**. Montevideo: CSIC, 2017.

BARITÉ, Mário; LAIPELT, Rita do Carmo Ferreira; PETROCCELLI, Patricia Daniela; GABRIEL JÚNIOR, Rene Faustino. Rumo a um tesauro-dicionário on-line da Organização do Conhecimento: primeiro design. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 2025, Rio de Janeiro, RJ. **Anais**… Rio de Janeiro: [s.n.], 2025.

BARSALOU, Lawrence W. Frames, concepts, and conceptual fields. In: LEHRER, Adrienne; KITTAY, Eva F. (ed.). **Frames, Fields, and Contrasts***: New Essays in Semantic and Lexical Organization*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992. p. 21–74. Disponível em: https://barsaloulab.org/Online_Articles/1992-Barsalou-chap-frames.pdf. Acesso em: 15 set. 2025.

BATES, M. J. Indexing and access for digital libraries and the internet: human, database, and domain factors. **Journal of the American Society for Information Science**, v. 49, n. 11, p. 991–998, 1998. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-4571%281998110%2949%3A13%3C1185%3A%3AAID-ASI6%3E3.0.CO%3B2-V. Acesso em: 10 mar. 2026.

BRÄSCHER, Marisa. Semantic relations in knowledge organization systems. **Knowledge Organization**, vol. 41, n. 2, p. 175-180, 2014. Disponível em: https://www.imrpress.com/journal/KO/41/2/10.5771/0943-7444-2014-2-175. Acesso em: 10 mar. 2026.

BOOKSTEIN, A.; SWANSON, D. Probabilistic models for automatic indexing. **Journal of the American Society for Information Science**, v. 25, n. 3, p. 143–149, 1974. Disponível em: https://dblp.org/rec/journals/jasis/BooksteinS74.html. Acesso em: 10 mar. 2026.

DEERWESTER, S. *et al.* Indexing by latent semantic analysis. **Journal of the American Society for Information Science**, v. 41, n. 3, p. 143–149, 1990. Disponível em: https://www.researchgate.net/publication/303802749_Indexing_by_Latent_Semantic_Analysis. Acesso em: 10 mar. 2026.

DIAS, C. A. Terminologia: conceitos e aplicações. Brasília, **Ciência da Informação,** v. 29, n. 1, jan./abr. p. 90-92, 2000. Disponível em: https://doi.org/10.1590/S0100-19652000000100009. Acesso em: 09 ago. 2024.

DILLON, M.; GRAY, A. S. FASIT: a fully automatic syntactically based indexing system. **Journal of the American Society for Information Science**, v. 34, n. 3, p. 199–207, 1983. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.4630340202. Acesso em: 10 mar. 2026.

FABER, Pamela; LEÓN-ARÁUZ, Pilar. Specialized Knowledge Representation and the Parameterization of Context. **Frontiers in Psychology**, February , Volume 7, Article 196, 2016. Disponível em: https://www.researchgate.net/publication/349516437_Specialized_Knowledge_Representation_from_Terms_to_Frames. Acesso em: 30 set. 2025.

FAGAN, J. L. The effectiveness of a nonsyntactic approach to automatic phrase indexing for document retrieval. **Journal of the American Society for Information**

Science, v. 40, n. 3, p. 199–207, 1989. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-4571(198903)40:2%3C115::AID-ASI6%3E3.0.CO;2-B. Acesso em: 10 mar. 2026.

FILLMORE, Charles J. Frame semantics. In: The Linguistic Society of Korea (ed.). **Linguistics in the Morning Calm**. Seoul: Hanshin, 1982. p. 111–137. Disponível em: https://brenocon.com/Fillmore%201982_2up.pdf. Acesso em: 15 set. 2025.

FRANCESCHINI, F.; MAISANO, D.; MASTROGIACOMO, L. Errors in DOI indexing by bibliometric databases. **Scientometrics**, v. 65, n. 5, 2014. Disponível em: https://dl.acm.org/doi/abs/10.1007/s11192-014-1503-4. Acesso em: 10 mar. 2026.

FUGMANN, Robert. The empirical approach in the evaluation of information systems. **Knowledge Organization**, v. 26, n. 1, 1999. Disponível em: https://www.inlibra.com/document/download/pdf/uuid/6f13ba98-7a79-3b98-9bdd-3fd9da9f1d6c. Acesso em: 09 mar. 2026.

FUJITA, Mariângela Spotti Lopes; PANUTO, Jessica Cristina; TARTAROTTI, Roberta Cristina Dal'Evedove. The indexing policy in the practices of Brazilian institutional repositories: a diagnostic study. **Journal of Librarianship and Information Science**, v. 50, n. 2, 2023. Disponível em: https://www.researchgate.net/publication/367173092_The_indexing_policy_in_the_practices_of_Brazilian_institutional_repositories_A_diagnostic_study_from_the_perception_of_managers_and_indexers. Acesso em: 09 mar. 2026.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60–90, 2003. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2089. Acesso em: 15 mar. 2026.

FUNK, M. E.; REID, C. A.; MCGOOGAN, L. S. Indexing consistency in MEDLINE. **Bulletin of the Medical Library Association,** v. 71, n. 3, p. 280–286, 1983. Disponível em: https://pubmed.ncbi.nlm.nih.gov/6344946/. Acesso em: 10 mar. 2026.

GARFIELD, E.; SHER, I. H. KeyWords Plus algorithmic derivative indexing. **Journal of the American Society for Information Science**, v. 44, n. 1, p. 1–9, 1993. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199306)44:5%3C298::AID-ASI5%3E3.0.CO;2-A. Acesso em: 10 mar. 2026.

GLANZE, W. Coauthorship patterns and trends in the sciences (1980-1998): a bibliometric study with implications for database indexing and search strategies. **Library Trends**, v. 50, n. 3, 2002. Disponível em: https://www.researchgate.net/publication/249465176_Coauthorship_Patterns_and_Trends_in_the_Sciences_1980-1998_A_Bibliometric_Study_With_Implications_for_D

atabase_Indexing_and_Search_Strategies/link/0f31752e646c7d3f7c000000/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19. Acesso em: 10 mar. 2026.

GIUNCHIGLIA, Fausto; DUTTA, Biswanath; MALTESE, Vincenzo. From knowledge organization to knowledge representation. **Knowledge Organization**, v. 41, n. 1, 2014. Disponível em: https://www.researchgate.net/publication/260073124_From_Knowledge_Organization_to_Knowledge_Representation. Acesso em: 09 mar. 2026.

GREEN, Rebecca. Relationships in the Organization of Knowledge: An Overview. In: BEAN, A.; GREEN, Rebecca (ed.). **Relationships in the Organization of Knowledge**. Springer Science & Business Media, 2001. Disponível em: https://link.springer.com/chapter/10.1007/978-94-015-9696-1_1. Acesso em: 10 abr. 2026.

GREEN, Rebecca. Relationships in knowledge organization. **Knowledge Organization**, Würzburg, v. 35, n. 3–4, p. 150–159, 2008. Disponível em: https://storage.imrpress.com/imr/journal/KO/article/505441/1752847190274.pdf. Acesso em: 20 fev. 2026.

GOLUB, Koraljka. Automated subject indexing: an overview. **Knowledge Organization**, Würzburg, v. 49, n. 3, p. 173–192, 2021. Disponível em: https://www.tandfonline.com/doi/full/10.1080/01639374.2021.2012311. Acesso em: 16 fev. 2026.

GOLUB, Koraljka *et al*. Subject indexing in humanities: a comparison between a local university repository and an international bibliographic service. **Journal of Documentation**, v. 76, n. 6, p. 1193-1214, 2020. Disponível em: https://www.researchgate.net/publication/341843847_Subject_indexing_in_humanities_a_comparison_between_a_local_university_repository_and_an_international_bibliographic_service. Acesso em: 20 fev. 2026.

GOLUB, Koraljka *et al.* Terminology registries for knowledge organization systems: functionality, use, and attributes. **Journal of the Association for Information Science and Technology,** v. 65, n. 9, p. 1901-1916, 2014. Disponível em: https://www.researchgate.net/publication/262026247_Terminology_registries_for_knowledge_organization_systems_Functionality_use_and_attributes. Acesso em: 09 mar. 2026.

GORDON, M. D.; DUMAIS, S. Using latent semantic indexing for literature based discovery. **Journal of the American Society for Information Science**, v. 49, n. 3, p. 280–286, 1998. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199806)49:8%3C674::AID-ASI2%3E3.0.CO;2-T. Acesso em: 10 mar. 2026.

HAJIBAYOVA, Lala. Basic-level categories: a review. **Journal of Information Science**, v. 39, n. 5, 2013. Disponível em:

https://www.researchgate.net/publication/258152827_Basic-level_categories_A_review. Acesso em: 09 mar. 2026.

HARTER, S. P. A probabilistic approach to automatic keyword indexing. **Journal of the American Society for Information Science**, v. 26, n. 5, p. 280–286, 1975. Disponível em: https://www.semanticscholar.org/paper/A-probabilistic-approach-to-automatic-keyword-Part-Harter/9bb0dcd80292262cf60e33259b91e03930bfd0f2. Acesso em: 10 mar. 2026.

HJØRLAND, Birger. Concepts of knowledge organization. **Knowledge Organization, Würzburg,** v. 43, n. 6, p. 475–484, 2016. Disponível em: https://www.researchgate.net/publication/324722232_Knowledge_Organization_KO. Acesso em: 02 fev. 2026.

HJØRLAND, Birger. Knowledge organization= information organization. **Advances in knowledge organization**, v. 13, p. 8-14, 2012. Disponível em: https://www.researchgate.net/profile/Birger-Hjorland/publication/357074802_Is_Knowledge_Organization_Information_Organization/links/64dde48ead846e28829c3a5f/Is-Knowledge-Organization-Information-Organization.pdf. Acesso em 15 fev. 2026.

HJØRLAND, Birger; BARROS, Thiago Henrique Bragato. Análise de domínio. **Em Questão**, v. 30, p. e-140568, 2024. Disponível em: https://seer.ufrgs.br/index.php/EmQuestao/article/view/140568. Acesso em: 8 ago. 2024.

HJØRLAND, Birger. What is Knowledge Organization (KO)?. **Knowledge Organization**, v. 35, n. 2/3, p. 86-101, 2008. Disponível em: https://www.researchgate.net/publication/277803483_What_is_Knowledge_Organization_KO. Acesso em: 08 ago. 2024.

HJØRLAND, Birger. Fundamentals of knowledge organization. **Knowledge Organization**, v. 30, n. 2, p. 87–111, 2003. Disponível em: https://www.researchgate.net/publication/285929806_Fundamentals_of_Knowledge_Organization. Acesso em: 09 mar. 2026.

HJØRLAND, Birger. Reviews of Concepts in Knowledge Organization. Subject (of Documents). **Knowledge Organization**, v. 44, n. 1, 2017. Disponível em: 10.5771/0943-7444-2017-1-55. Acesso: 09 mar. 2026.

HJØRLAND, Birger. The importance of theories of knowledge: indexing and information retrieval as an example. **Journal of the American Society for Information Science and Technology**, v. 62, n. 1, p. 72–77, 2011. Disponível em: https://doi.org/10.1002/asi.21451. Acesso em: 09 mar. 2026.

HJØRLAND, Birger. Indexing: concepts and theory. **Knowledge Organization**, v. 45, n. 7, 2018. Disponível em: https://www.researchgate.net/publication/329449826_Indexing_Concepts_and_Theory. Acesso em: 09 mar. 2026.

HJØRLAND, Birger. Political versus apolitical epistemologies in knowledge organization. **Knowledge Organization**, v. 47, n. 3, 2020. Disponível em: https://www.imrpress.com/journal/KO/47/6/10.5771/0943-7444-2020-6-461. Acesso em: 09 mar. 2026.

HOLMA, Baiba. The reference system of relational semantics in knowledge organization systems. In: GARCIA, J.G.; MARTÍNEZ, F. B.; SERRA, A. P. (coord.) **La dimensió humana de l'organització del coneixement.** Facultat de Biblioteconomia i Documentació, 2005. Disponível em: https://www.researchgate.net/publication/28296253_The_reference_system_of_relational_semantics_in_knowledge_organization_systems. Acesso em: 31 out. 2025.

HUMPHREY, S. M.; MILLER, N. E. Knowledge based indexing of the medical literature: the indexing aid project. **Journal of the American Society for Information Science**, v. 38, n. 2, p. 89–98, 1987. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1097-4571(198705)38:3%3C184::AID-ASI7%3E3.0.CO;2-F. Acesso em: 10 mar. 2026.

KAKALI, Constantia; PAPATHEODOROU, Christos. Exploitation of folksonomies in subject analysis. **Library & Information Science Research**, v. 32, n. 2, p. 87–97, 2010. Disponível em: https://www.sciencedirect.com/science/article/pii/S0740818810000368. Acesso em: 09 mar. 2026.

KREBS, Luciana; LAIPELT, Rita do Carmo Ferreira. Teorias da linguística cognitiva para pensar a categorização no âmbito da Ciência da Informação. **Transinformação**, [S. l.], v. 30, n. 1, 2018. Disponível em: https://periodicos.puc-campinas.edu.br/transinfo/article/view/5949. Acesso em: 11 set. 2025.

KRIEGER, M. da G.; SANTIAGO, M. S.; CABRÉ, M. T. Terminologia em foco: uma entrevista comentada com Maria Teresa Cabré. **Calidoscópio**, [S. l.], v. 11, n. 3, p. 328–332, 2013. Disponível em: https://revistas.unisinos.br/index.php/calidoscopio/article/view/cld.2013.113.11. Acesso em: 22 abr. 2025.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2788:1986.** Documentation – Guidelines for the establishment and development of monolingual thesauri. Geneva: International Organization for Standardization, 1986.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 5964:1985.** Documentation – Guidelines for the establishment and development of multilingual thesauri. Geneva: International Organization for Standardization, 1985.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-1:2011** — Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-2:2013** — Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies. Geneva: International Organization for Standardization, 2013.

LA BARRE, Kathryn A.; TILLEY, Carol L. The elusive tale: leveraging the study of information seeking and knowledge organization to improve access to and discovery of folktales. **Journal Of The American Society For Information Science And Technology,** 63(4):687–701, 2012. Disponível em: https://ideas.repec.org/a/bla/jamist/v63y2012i4p687-701.html. Acesso em: 09 mar. 2026.

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

MACULAN, Benildes *et al.* **Aplicação de reengenharia de tesauro: modelagem do THESAGRO**. 2015. Disponível em: https://iskoiberico.org/wp-content/uploads/2015/11/0d9bd-224_maculan.pdf. Acesso em: 02 Fev. 2026.

MACULAN, Benildes Coura Moreira. **Metodologia para refinamento semântico de relações em tesauros**. Belo Horizonte: Editora UFMG, 2024.

MAI, Jens-Erik. Folksonomies and the new order: authority in the digital age. **Knowledge Organization**, Würzburg, v. 38, n. 2, 2011. Disponível em: https://www.researchgate.net/publication/228802528_Folksonomies_and_the_New_Order_Authority_in_the_Digital_Disorder. Acesso em: 09 mar. 2026.

MATTHEWS, Brian; JONES, Catherine; PUZON, Bartłomiej; MOON, Jim; TUDHOPE, Douglas; NIELSEN, Marianne Lykke. An evaluation of enhancing social tagging with a knowledge organization system. **Emerald Publishing**, v. 36, n. 5, 2010. Disponível em: https://www.emerald.com/ajim/article-abstract/62/4-5/447/122084/An-evaluation-of-enhancing-social-tagging-with-a?redirectedFrom=fulltext. Acesso em: 09 mar. 2026.

MCCREADIE, R.; MACDONALD, C.; OUNIS, I. MapMap indexing strategies: studying scalability and efficiency. **Information Processing & Management**, v. 62, n. 11, 2011/2012. Disponível em: https://www.researchgate.net/publication/228353403_MapReduce_indexing_strategies_Studying_scalability_and_efficiency. Acesso em: 10 mar. 2026.

MARON, M. E. On indexing, retrieval and the meaning of about. **Journal of the American Society for Information Science**, v. 28, n. 1, p. 1–9, 1977. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.4630280107. Acesso em: 10 mar. 2026.

MARTÍNEZ-ÁVILA, Daniel; GUIMARÃES, José Augusto Chaves; PINHO, Fabio Assis; FOX, Melodie J. The representation of ethics and knowledge organization in

the WoS and LISTA databases. **Knowledge Organization**, v. 42, n. 4, 2015. Disponível em: https://www.researchgate.net/publication/289988270_The_Representation_of_Ethics_and_Knowledge_Organization_in_the_WoS_and_LISTA_Databases. Acesso em: 09 mar. 2026.

MEDEIROS, Marisa Bräscher Basílio. Terminologia brasileira em ciência da informação: uma análise. **Ciência da Informação**, Brasília, 15(2): 135-42, jul./dez. Disponível em: https://doi.org/10.18225/ci.inf.v15i2.234. Acesso em: 28 jan. 2025.

MEDELYAN, O.; WITTEN, I. H. Thesaurus based automatic keyphrase indexing. **JCDL**: Joint Conference on Digital Libraries, v. 57, n. 3, p. 280–286, 2006. Disponível em:https://www.researchgate.net/publication/224061561_Thesaurus_based_automatic_keyphrase_indexing. Acesso em: 10 mar. 2026.

MEDELYAN, O.; WITTEN, I. H. Domain-independent automatic keyphrase indexing with small training sets. **Journal of the American Society for Information Science**, v. 59, n. 3, 2008. Disponível em: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.20790?__cf_chl_f_tk=zlEjLhKZY8xtQLsJUu91s9H6om31cG5Me6K_DzazdBw-1782948010-1.0.1.1-CrjUp5Sa7TN3319v31WTp1hSNcYKS7kkYL63YS8iLvo. Acesso em: 10 mar. 2026.

PETERS, Isabella; WELLER, Katrin. Paradigmatic and Syntagmatic Relations in Knowledge Organization Systems. **Information--Wissenschaft und Praxis**, v. 59, n. 2, 2008. Disponível em: https://scispace.com/pdf/paradigmatic-and-syntagmatic-relations-in-knowledge-20w4ez1tc3.pdf. Acesso em: 06 fev. 2026.

PINHO, Fabio Assis; GUIMARÃES, José Augusto Chaves. Male homosexuality in Brazilian indexing languages: some ethical questions. **Knowledge Organization**, v. 39, n. 4, 2012. Disponível em: https://repositorio.unesp.br/entities/publication/47c53dd8-610e-4fb3-912e-d5c4345d2611:. Acesso em: 09 mar. 2026.

QIN, Jian. Knowledge Organization and Representation under the AI Lens. J. **Data Inf. Sci.**, 2020, vol. 5, no 1, p. 3-17. Disponível em: https://www.researchgate.net/publication/341656740_Knowledge_Organization_and_Representation_under_the_AI_Lens. Acesso em: 08 mar. 2026.

ROLLING, L. Indexing consistency, quality and efficiency. **Information Processing & Management**, v. 7, n. 2, p. 89–98, 1981. Disponível em: https://www.sciencedirect.com/science/article/pii/0306457381900285. Acesso em: 10 mar. 2026.

SALTON, G.; YANG, C. T. On the specification of term values in automatic indexing. **Journal of Documentation**, v. 29, n. 4, p. 358–372, 1973. Disponível em:

https://www.semanticscholar.org/paper/On-the-Specification-of-Term-Values-in-Automatic-Salton-Yang/c7fd9037b571862246e366189f1fc1c58fe0dd9c. Acesso em: 10 mar. 2026.

SAMUELSSON, Jenny. Knowledge organization for feminism and feminist research: a discourse-oriented study of systematicity. **Knowledge Organization**, v. 37, n. 1, 2010. Disponível em: https://philpapers.org/rec/SAMKOF?__cf_chl_f_tk=RXsiG5SZca4D3E7AIeYTLK8ivlMF9MeFSQHLw.lyAro-1782936841-1.0.1.1-_Cq24kMLS3lwL.ZLB3PUkMYTdo5VruaICVHxrNSOEQM. Acesso em: 09 mar. 2026.

SAUMURE, Kristie; SHIRI, Ali. Knowledge organization trends in library and information studies: a preliminary comparison of the pre- and post-web eras. **Journal of Information Science**, v. 34 (5), 2008. Disponível em: https://journals.sagepub.com/doi/10.1177/0165551507084300?__cf_chl_f_tk=4Iuf32qsDei31ioSs9.mJ_tP7e.uEHhjfm4I3gn61yo-1782934757-1.0.1.1-5zqqWq4b9CChpewqh8r8jZd6k7.YtaBieNFW_md2NH8. Acesso em: 09 mar. 2026.

SOERGEL, D. **Indexing languages and thesauri**: construction and maintenance. Los Angeles: Melville Publishing Company, 1974. Disponível em: https://www.researchgate.net/publication/308310252_Indexing_languages_and_thesauri_Construction_and_maintenance. Acesso em: 15 mar. 2025.

SOERGEL, Dagobert. Unleashing the power of data through organization: structure and connections for meaning, learning, and discovery. **Knowledge Organization**, Würzburg, v. 42, n. 6, 2015. Disponível em: https://www.researchgate.net/publication/293027204_Unleashing_the_Power_of_Data_Through_Organization_Structure_and_Connections_for_Meaning_Learning_and_Discovery. Acesso em: 09 mar. 2026.

SPARCK JONES, K. Index term weighting. **Information Storage and Retrieval**, v. 9, n. 11, p. 619–633, 1973. Disponível em: https://www.semanticscholar.org/paper/Index-term-weighting-Jones/a075b6f5597f3cee0b70c473e6a298e08c81b636. Acesso em: 10 mar. 2026.

SUOMINEN, Osma; INKINEN, Juho; LEHTINEN, Mona. Annif and Finto AI: developing and implementing automated subject indexing. **Cataloging & Classification Quarterly**, Abingdon, v. 59, n. 1, p. 27–54, 2021. Disponível em: https://urn.fi/URN:NBN:fi-fe2021080942632. Acesso em: 10 mar. 2026

STVILIA, Besiki; JÖRGENSEN, Corinne. Member activities and quality of tags in a collection of historical photographs in Flickr. **Journal of the American Society for Information Science and Technology**, v. 61, n. 12, 2010. Disponível em: https://ideas.repec.org/a/bla/jinfst/v61y2010i12p2477-2489.html. Acesso em: 09 mar. 2026.

SVENONIUS, Elaine. The intellectual foundation of information organization. Cambridge: The MIT Press, 2000. 255p. **Perspectivas em Ciência da Informação**, [S. l.], v. 7, n. 2, 2007. Disponível em:

https://periodicos.ufmg.br/index.php/pci/article/view/23434. Acesso em: 22 mar. 2024.

TENNIS, Joseph T. Forms, Knowledge, and Forms of Knowing: Correct and Useful Epistemology and Knowledge Organization. In: GUIMARÃES, José Augusto Chaves, DOBEDEI, Vera (org.). **Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade**. Marília: ISKO-Brasil : FUNDEPE, 2012.

TENNIS, Joseph T.; SUTTON, Stuart A. Extending the Simple Knowledge Organization System for concept management in vocabulary development applications. **Journal of the American Society for Information Science and Technology**, v. 59, n. 1, 2008. Disponível em: https://www.researchgate.net/publication/220432821_Extending_the_Simple_Knowledge_Organization_System_SKOS_for_concept_management_in_vocabulary_development_applications. Disponível em: 09 mar. 2026.

TOGNOLI, Natalia; GUIMARÃES, José Augusto Chaves. Provenance as a knowledge organization principle. **Knowledge Organization**, v. 46, n. 8, 2019. Disponível em: https://www.researchgate.net/publication/338472753_Provenance_as_a_Knowledge_Organization_Principle. Disponível em: 09 mar. 2026.

VLACHIDIS, Andreas; TUDHOPE, Douglas. A knowledge based approach to information extraction for semantic interoperability in the archaeology domain. **Journal of the Association for Information Science and Technology**, v. 38, n. 1, 2012. Disponível em: https://www.researchgate.net/publication/273890216_A_knowledge-based_approach_to_Information_Extraction_for_semantic_interoperability_in_the_archaeology_domain. Disponível em: 09 mar. 2026.

VÉZINA, Robert (*et al*). **Escrevendo definições terminológicas**. Quebec: Escritório de Quebec da língua francesa, 2009.

VOGEL, Michely Jabala Mamede. **A noção de estrutura lingüística e de processo de estruturação e sua influência no conceito e na elaboração de linguagens documentárias**. 2007. Dissertação (Mestrado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2007. doi:10.11606/D.27.2007.tde-17032008-104245. Acesso em: 16 out. 2024.

WU, Dan; HE, Daqing; QIU, Jin; LIN, Ruonan; LIU, Yang. Comparing social tags with subject headings on annotating books: a study comparing the information science domain in English and Chinese. **Journal of Information Science**, v. 39, n. 3, 2012. Disponível em: https://dl.acm.org/doi/abs/10.1177/0165551512451808?__cf_chl_f_tk=nh4QDjGMQibhg.Yt6rdmRrEKnXZwjRo7L8M0ehR3zTk-1782934985-1.0.1.1-RiUxz6JjFBrWOWCi9XcxYzwN5IkFwPMcsgIORHaeJ3s. Acesso em: 09 mar. 2026.

ZENG, Marcia Lei. Knowledge organization systems (KOS). **Knowledge Organization**, Würzburg, v. 35, n. 2–3, p. 160–182, 2008. Disponível em:

https://www.researchgate.net/publication/297530633_Knowledge_Organization_Systems_KOS. Acesso em: 20 mar. 2026.

ANEXO 1

Árvore de domínio do "Diccionario de Organización del Conocimiento"

1. **Knowledge Organization in general**

Nota de inclusão: inclui termos correspondentes a conceitos fundamentais, teorias ou aspectos teóricos específicos do OC.

1.1 Fundaments and theories

Bibliographic classification
Domain analysis
Faceted theory
Integrative levels theory
Knowledge organization (Knowledge representation and organization) (KO)
(Knowledge representation and organization)
(KO)

1.2. Basic concepts

Data
Discipline
Document
Domain
Information
Knowledge
Subject
Work

1.3. Related fields

Exclusion: Specific terms from related areas will not be included.
Artificial Intelligence
Bibliometrics
Conceptology (Theory of concept)
Information organization
Information retrieval
Knowledge management
Knowledge representation
Linguistics
Logic
Natural Language Processing
(NLP), v. Natural Language Processing
Terminology
(Theory of concept), v. Conceptology.

2. KO Processes (KOP)

Nota de inclusión: inclui termos representativos dos procedimentos utilizados para executar as funções e finalidades do OC, bem como aqueles que indicam os seus princípios ou elementos.

2.1. Concepts, techniques and methods

Aboutness

Abstracting (2)
Attribute
Content analysis
Disambiguation
Facet analysis
Fiction
Non fiction
Subject analysis
Technical reading

2.2. Classification
Broad classification
Call number
Citation order
Close classification
Reclassification

2.3. Indexing
Automatic indexing
Chain indexing
(Consistency)
(Free indexing), v. Keyword indexing
Indexing consistency
Keyword indexing (Free indexing)
Social indexing
Subject indexing
Tagging

2.4. Abstracting
Abstract

2.5. Coding
Icon

2.7. Ordering
Alphabetical order
Numeric order
Systematic order

3. Knowledge organization systems (KOS)
Nota de inclusion: inclui termos relacionados com as diferentes linguagens e sistemas utilizados no OC, bem como aqueles que indicam os seus princípios, elementos ou partes.

3.1. Principles
Canon

3.1.1. Hierarchy
Hierarchical force
Polyhierarchy

3.1.2. Helpful sequence

Alphabetical order (2)
Canonical order
Decreasing concreteness
Increasing concreteness
Later-in-evolution
Later-in-time
Spatial contiguity

3.1.3. Mutual exclusivity (Mutual exclusion)

3.1.4. Specificity

3.1.5. Warrant

Academic warrant (Scholarly warrant)
Common usage
Cultural warrant
Epistemic warrant
Ethical warrant
Literary warrant
Organizational warrant
Semantic warrant
Structural warrant
Use warrant
User warrant

3.2. Typologies

Language

3.2.1. Types of KOS

Artificial language
Classaurus
Classification system (Bibliographic classification)
Ad hoc Classification
Analytic-synthetic classification (Analytico-synthetic classification)
(Analytico-synthetic classification)
Folk classification (=Naïve classification?)
Naïve classification (=Folk classification?)
Decimal classification
Specialized classification
Universal classification
Code
Conceptual map
Controlled vocabulary (Standardized vocabulary)
Dictionary
Folksonomy
Glossary
Index

- Asyndetic index
- KWAC index
- KWIC index
- KWOC index
- Onomastic index
- Syndetic index

Indexing system (Indexing language)
(Indexing language)
List
- Authority list
- Descriptor list
- Faceted list
- Subject headings list

Ontology
Synonyms ring
Taxonomy
Thesaurofacet
Thesaurus
- Corporate thesaurus
- Enriched thesaurus
- Faceted thesaurus
- Macrothesaurus
- Microthesaurus
- Monolingual thesaurus
- Plurilingual thesaurus

Vocabulary

3.2.2. Specific KOS

(BC)
Bibliographic Classification
(CC)
Classification System for Knowledge Organization literature
Colon Classification
(CSKOL)
(DDC)
Dewey Decimal Classification
(ICC)
International Coding Classification
(LCC)
Library of Congress Classification
Library of Congress Subject Headings
(LCSH)
Uniterm
(UDC)
Universal Decimal Classification

3.3. Parts of KOS

Auxiliary table
Chain index
Class

Facet
Identifier
Index (2)
    Chain index
    Permuted index (Rotated index)
    Relative index
Main Class
Manual
Note
    Add note
    Content note
    Definition note
    Scope note (SN)
PMEST
Scheme
Subclass
Subfacet
Summary

3.4. Representation
Abstract
Accession number
Alternative location
Array
Arrow
Authoritative subject heading
(Basis of division), v. Characteristic
Bias
Broader term (BT)
(BT), v. Broader term
Category
Characteristic (Basis of division)
Class
Class number
Classification number
Call number
Compound term
Cross classification
Cross reference
Descriptor
Discontinued term
Focus
Fundamental category
Genre
Heading
Index term
Instance
Inverted heading
Keyword
Narrowed term (NT)

Natural language
(NT)
Non descriptor
Orphan term
Paradigmatic relation(ship)
Post-coordination
Pre-coordination
Preferred term
Qualifier
Related term
Species
Subject heading
Symbol
Syntagmatic relation(ship)
Tag
Taxon
Term

3.5. Structure and Elements of KOS
Abridged edition
Absolute syntax
Alphabetical order
Associative relation(ship) (Related term relation(ship)
Equivalence relation(ship)
Full edition
(Generic relation)
Genus-species relation(ship)
Hierarchical relation(ship)
Notation
Standard edition
Synonymic relation
Syntax
Systematic order
Use
Use for
Whole-part relation(ship)

4. Methods
Nota de inclusão: inclui termos relacionados com métodos específicos ou aspetos metodológicos da OC.
Analytic-synthetic method (Analytico-synthetic method)
(Analytico-synthetic method)
Authority control
Bottom-up
Faceted methodology
Interoperability
Synonimy control
Systematic indexing
Systematifier
Top-down

Vocabulary control

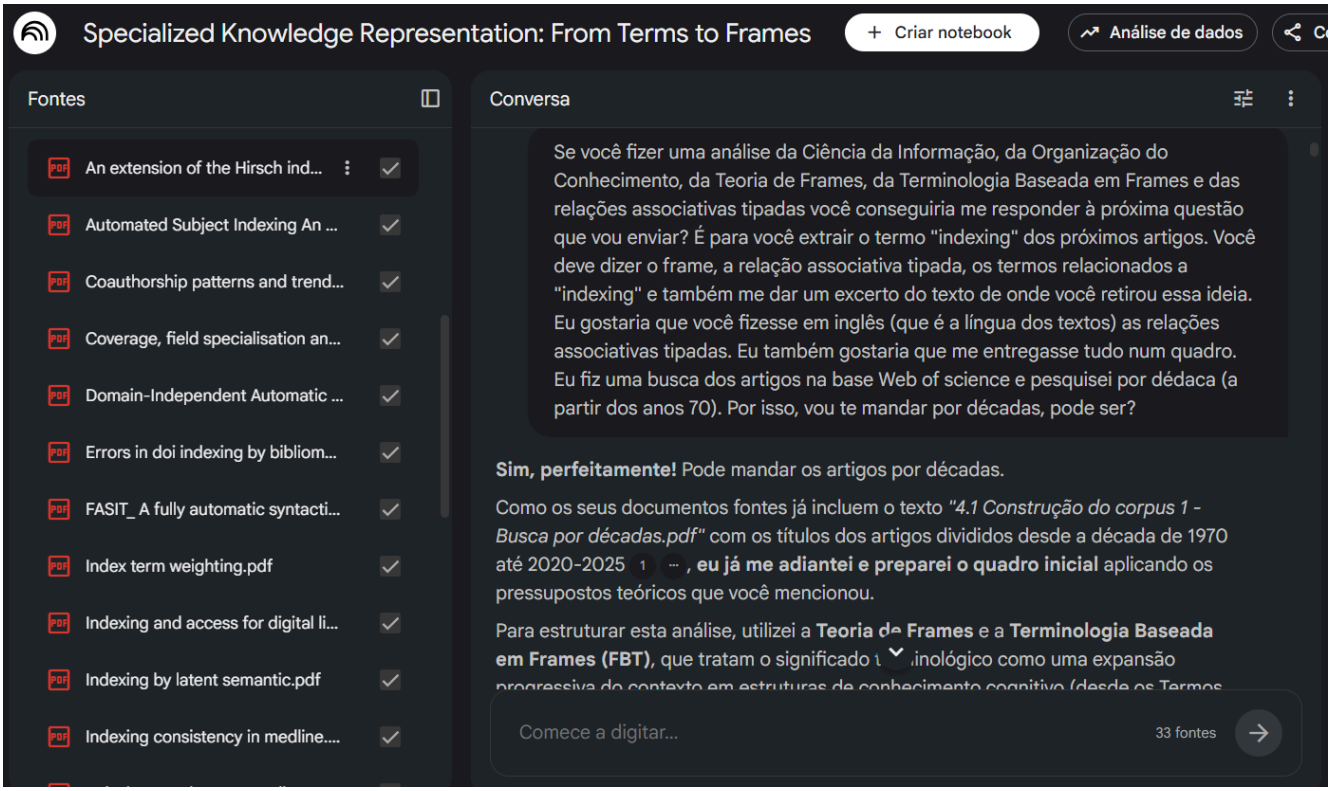
5. Persons
Nota de inclusion: inclui termos representativos do exercício profissional na área de OC.
Cataloguer
Classificacionist
Classifier
Community
Expert
Indexer
Reader
User

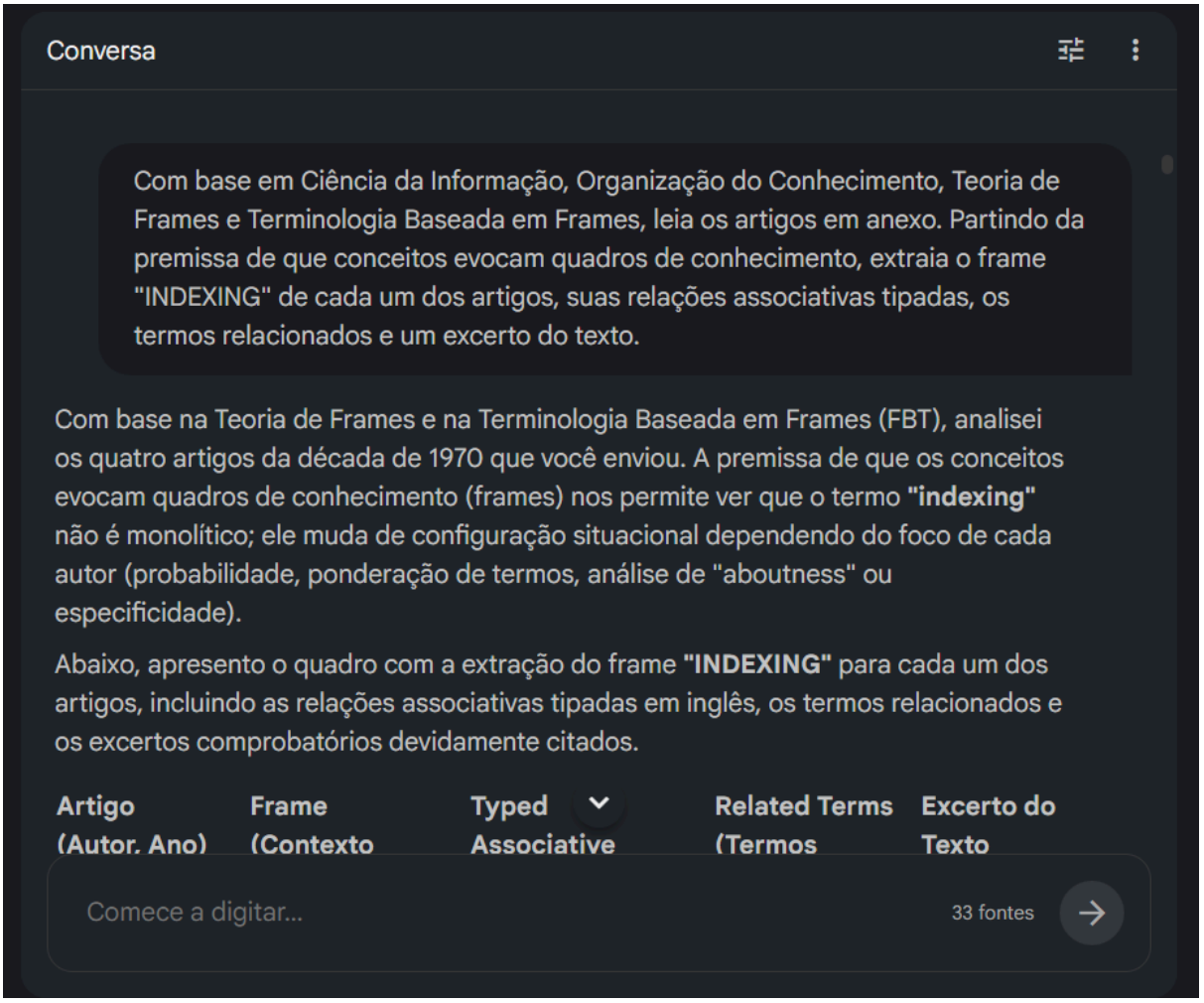
6. Institutions
International Society for Knowledge Organization
Library of Congress
UDC Consortium

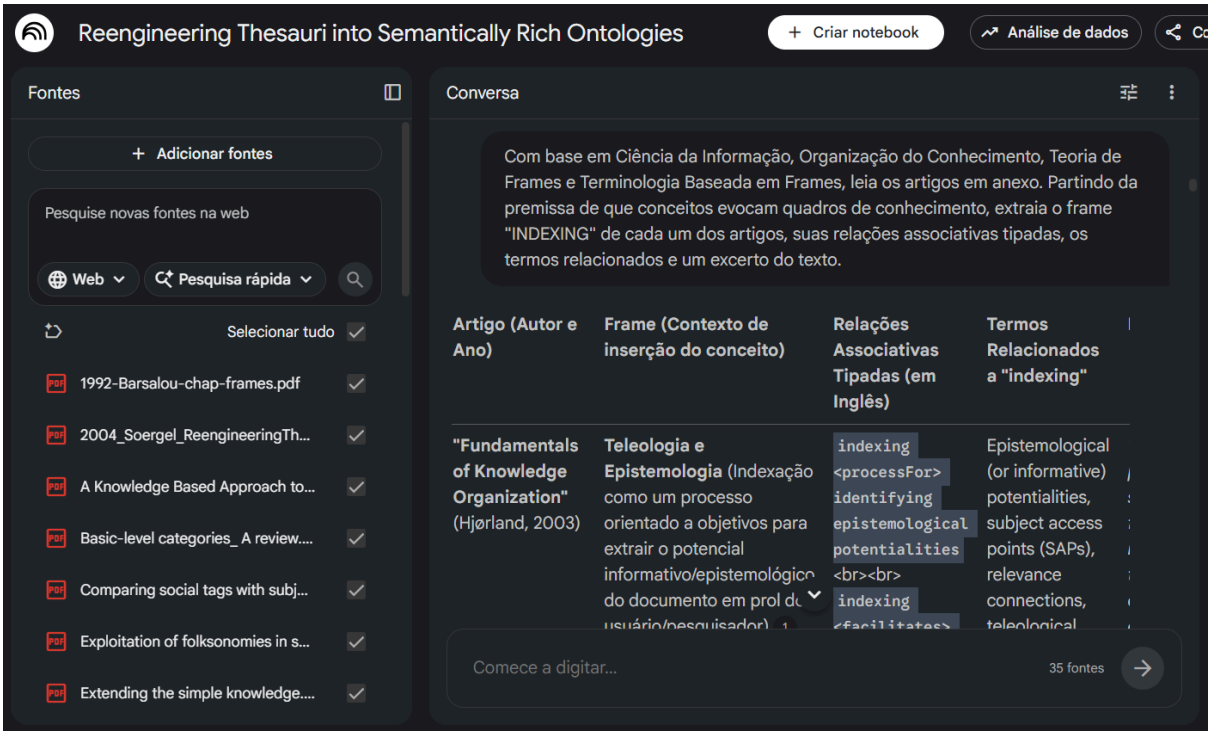
7. Doubtful cases
Library classification
Simple Knowledge Organization System
(SKOS)

ANEXO 2

Extração Semântica e Modelagem de Frames (NotebookLM).




Conversa
Com base em Ciência da Informação, Organização do Conhecimento, Teoria de Frames e Terminologia Baseada em Frames, leia os artigos em anexo. Partindo da premissa de que conceitos evocam quadros de conhecimento, extraia o frame "INDEXING" de cada um dos artigos, suas relações associativas tipadas, os termos relacionados e um excerto do texto.
Com base na Teoria de Frames e na Terminologia Baseada em Frames (FBT), analisei os quatro artigos da década de 1970 que você enviou. A premissa de que os conceitos evocam quadros de conhecimento (frames) nos permite ver que o termo "indexing" não é monolítico; ele muda de configuração situacional dependendo do foco de cada autor (probabilidade, ponderação de termos, análise de "aboutness" ou especificidade).
Abaixo, apresento o quadro com a extração do frame "INDEXING" para cada um dos artigos, incluindo as relações associativas tipadas em inglês, os termos relacionados e os excertos comprobatórios devidamente citados.
Artigo (Autor, Ano)
Frame (Contexto
Typed Associative
Related Terms (Termos
Excerto do Texto
Comece a digitar...
33 fontes


Reengineering Thesauri into Semantically Rich Ontologies
+ Criar notebook
Análise de dados
Fontes
+ Adicionar fontes
Pesquise novas fontes na web
Web
Pesquisa rápida
Selecionar tudo
1992-Barsalou-chap-frames.pdf
2004_Soergel_ReengineeringTh...
A Knowledge Based Approach to...
Basic-level categories_ A review....
Comparing social tags with subj...
Exploitation of folksonomies in s...
Extending the simple knowledge....
Conversa
Com base em Ciência da Informação, Organização do Conhecimento, Teoria de Frames e Terminologia Baseada em Frames, leia os artigos em anexo. Partindo da premissa de que conceitos evocam quadros de conhecimento, extraia o frame "INDEXING" de cada um dos artigos, suas relações associativas tipadas, os termos relacionados e um excerto do texto.
Artigo (Autor e Ano)
Frame (Contexto de inserção do conceito)
Relações Associativas Tipadas (em Inglês)
Termos Relacionados a "indexing"
"Fundamentals of Knowledge Organization" (Hjørland, 2003)
Teleologia e Epistemologia (Indexação como um processo orientado a objetivos para extrair o potencial informativo/epistemológico do documento em prol d
indexing <processFor> identifying epistemological potentialities

 indexing
Epistemological (or informative) potentialities, subject access points (SAPs), relevance connections, teleological
Comece a digitar...
35 fontes