

Controle de vocabulário no tesauro UNESP

Samantha Augusta dos Santos de Jesus ¹

<https://orcid.org/0000-0003-4479-0836>

¹Universidade Estadual Paulista, Marília, SP, Brasil

Resumo: A indexação é uma prática que visa à recuperação da informação e se divide nos seguintes processos: análise, síntese e representação. O último é considerado primordial e, para ser empregado, requer um sistema de controle de vocabulário, como o tesauro. O tesauro é compreendido como um sistema especializado de termos relacionados. Ele representa os termos identificados no processo de indexação pelo controle de vocabulário. O Tesauro UNESP tem como objetivo controlar os termos utilizados por seus usuários da comunidade acadêmica; no entanto, apresenta alguns problemas. Dessa forma, busca-se compreender a importância do controle de vocabulário, identificar, discutir e propor melhorias. Para isso, foi realizado um levantamento na literatura, nas bases de dados LISTA e BRAPCI, sendo o tesauro analisado conforme as recomendações da norma ISO 25964-1 e a literatura pertinente. Diversos problemas foram identificados no tesauro e possíveis soluções foram propostas. Conclui-se que o tesauro é um instrumento muito importante e que o controle de vocabulário é vital, sendo necessárias melhorias no instrumento de linguagem.

Palavras-chave: controle de vocabulário; tesauro; tesauro UNESP

1 Introdução

A indexação é uma ação que descreve e identifica um documento de acordo com um assunto (Silva; Fujita, 2004). O conceito de indexação foi cunhado a partir da elaboração de índices e está relacionado à prática da análise de assunto.

O processo da indexação pode ser compreendido, sinteticamente, em três etapas (Fujita; Neves; Dal'Evedove, 2017):

- a) leitura, identificação e seleção dos conceitos, realizadas pelo leitor profissional;
- b) construção do texto documentário por meio dos conceitos selecionados;
- c) tradução dos conceitos por meio das linguagens documentárias.

Todas as etapas apresentadas são de extrema importância para a prática da análise documentária, principalmente no que diz respeito ao controle de vocabulário, pois possibilita uma representação fidedigna da linguagem da comunidade usuária do sistema e, conseqüentemente, uma recuperação da informação eficaz.

O tesauro é um instrumento de controle de vocabulário, baseado em conceitos, termos preferidos e não preferidos, e relações, aplicado a um campo específico do conhecimento e utilizado no controle terminológico para a indexação de documentos (Van der Laan; Ferreira, 2000). Isso significa que o controle de vocabulário é uma atividade intrínseca ao tesauro, pois reflete na representação e na recuperação da informação:

[O] tesauro é um sistema de organização do conhecimento científico [...], utilizado como instrumento para organização, indexação e recuperação da informação em bases de dados. O Tesauro da UNESP tem vocabulário com termos especializados das áreas de conhecimento de suas atividades [...] [e] é utilizado para representar o conteúdo da informação (Universidade Estadual Paulista, [2013-]).

O tesauro é um vocabulário controlado que auxilia no processo da indexação de documentos, realizada pelos bibliotecários nas unidades da instituição, servindo também como exemplo de representação para uso em outros tesouros e demais sistemas de controle de vocabulário. Com o Tesauro UNESP não é diferente, além de ser utilizado pelos usuários na busca de documentos no sistema. No entanto, esse tesauro possui diversos problemas de estrutura e de controle de vocabulário, como identificado pelo uso e por alguns apontamentos em estudos anteriores. Esses problemas podem não ser tão bem-vistos e podem prejudicar futuras indexações e buscas no sistema.

Dessa maneira, como forma de se aprofundar nesses estudos anteriores, a problemática da pesquisa reside em identificar quais são esses problemas de controle de vocabulário apresentados no Tesauro UNESP e como a literatura orienta na resolução desses problemas.

Para isso, a pesquisa objetivou avaliar o Tesauro UNESP, mais especificamente (Fujita; Neves; Dal'Evedove, 2017):

- a) identificar e discutir os problemas e;
- b) propor melhorias de controle de vocabulário.

Para a revisão bibliográfica, realizou-se um levantamento bibliográfico nas bases de dados *Library Information Science Technology Abstract* (LISTA) e Base de Dados em Ciência da Informação (BRAPCI), identificando os artigos representativos da literatura na área, nos anos 2012 a 2024.

A análise no Tesauro UNESP foi realizada no conceito “Deus” por meio da literatura levantada e, igualmente, por meio das normas de construção de tesauro, a ISO 25964-1 (International Organization for Standardization, 2011).

A pesquisa foi dividida em seções. Após a introdução com a contextualização da problemática e os objetivos, foi apresentada a seção da indexação e do vocabulário controlado; depois, a metodologia, que abordou o Tesauro UNESP e todos os procedimentos de coleta e análise; os resultados, com a aplicação do método e todos os procedimentos metodológicos; e, por fim, a conclusão com reflexões sobre a proposta inicial e os resultados.

2 Indexação e controle de vocabulário

O conceito de indexação foi cunhado a partir da elaboração de índices e está relacionado a prática da análise de assunto. A indexação pode ser conceituada como um processo de sistematização que objetiva a padronização para uma recuperação eficaz, através de mecanismos de busca (Martines; Almeida, 2023).

Essa padronização ocorre por meio da ação que descreve e identifica um documento de acordo com um assunto (Oliveira; Santos, 2024; Silva; Fujita, 2004) e se caracteriza por estágios e etapas que se sobrepõem, com o propósito de recuperar e acessar a informação do conteúdo do documento (Klann; Ramos, 2023; Mafra; Cordeiro, 2022).

A indexação é o processo mais importante da análise documentária, ou seja, ela “[...] condiciona o valor de um sistema documentário” (Chaumier, 1988,

p. 63). Considera-se a indexação como o processo de escrever e caracterizar um documento por meio da representação de conceitos. A representação desses conceitos envolve a “[...] transcrição dos conceitos em uma linguagem documentária, após tê-los extraído do documento através da análise” (Chaumier, 1988, p. 64), tendo em vista que essa transcrição ocorre graças aos instrumentos de indexação (classificações, tesouros e dentre outros).

Na análise documentária, a indexação é uma prática de representação com objetivo de recuperar a informação, sendo o passo mais importante, pois ela auxilia na eliminação de 90% dos ruídos e silêncios em um sistema documentário (Chaumier, 1988; Silva; Fujita, 2004).

De acordo com Lancaster (2004), dificilmente o indexador lerá um documento integralmente, em razão da quantidade de trabalho exigida. O que acontece, normalmente, é um misto entre uma leitura atenta a uma leitura *scanning*. A primeira leitura mencionada somente é realizada em partes do texto em que se terá o máximo de conteúdo representativo, como no título, resumo, sinopse, conclusões, títulos de seções, legendas de ilustrações ou de tabelas. No entanto, deve-se levar em conta o documento integral, isto é, as partes vistas de relance, considerando que “[...] os termos atribuídos precisam refletir o todo” (Lancaster, 2004, p. 24).

O processo da indexação, compreende-se, essencialmente, três etapas (Fujita; Neves; Dal’evedove, 2017; Rubi, 2009):

- a) análise;
- b) síntese;
- c) representação.

Na primeira etapa, o leitor profissional realiza a leitura, a identificação e a seleção de conceitos; na segunda, constrói-se um texto documentário por meio dos conceitos selecionados, etapa relacionada com a elaboração de resumos; e por último, no processo central da indexação de assuntos, os conceitos são traduzidos por meio das linguagens documentárias (Fujita; Neves; Dal’evedove, 2017; Rubi, 2009), ou seja, ocorre “[...] a análise conceitual inicial em uma descrição formal

do assunto ou representação do assunto” (Holstrom, 2024, p. 523, tradução própria).

Frequentemente, as expressões “controle de vocabulário” e “vocabulário controlado” são confundidas entre si. Para distinção, define-se o controle de vocabulário como o processo de um objetivo, enquanto o vocabulário controlado como o resultado desse processo (Aguiar; Tálamo, 2012).

Além disso, o controle de vocabulário é um processo normativo que qualifica as atividades através dos princípios da terminologia. Esse processo objetiva a singularidade terminológica e conceitual entre o significante e o significado (Smit, 2018).

Barité (2014) define o controle de vocabulário como um conjunto de técnicas e procedimentos que podem ser aplicadas para a resolução de problemas de ambiguidade, relação entre termos que expressam conceitos e denominações.

Para o estabelecimento de um controle de vocabulário, é necessária a padronização gramatical, a identificação dos sinônimos, dos quase-sinônimos e dos antônimos (Smit, 2018).

Isto é, o controle de vocabulário é dotado de uma sequência de ações baseadas na terminologia e na semântica, no conhecimento do vocabulário utilizado pela instituição e pela bibliografia, além de ser essencial a clareza nas motivações das decisões na operação do controle (Smit, 2018).

Sintetizando, “[...] o processo de controle de vocabulário pressupõe intervenções no: controle da normalização gramatical – a forma dos termos; controle semântico dos termos ou controle do significado” (Aguiar; Tálamo, 2012, p. 119).

A aplicação do vocabulário compreende que a linguagem natural está presente em diversos documentos, no entanto, em comunidades especializadas sua aplicação pode ser considerada um desvio das ciências que requerem maior consolidação e avanço (Barité, 2014):

O controle de vocabulário é utilizado como ferramenta comum no processo de criação, desenvolvimento ou revisão [...] de qualquer tipo de sistema de organização do conhecimento, mas também como

suporte para a organização formal de conteúdos (Barité, 2014, p. 98, tradução própria).

Os vocabulários controlados oferecem mecanismos para remediar a ambiguidade generalizada e a sinonímia presente na linguagem natural (Liu; Wacholder, 2017) por meio de uma representação que diminua o uso de diversas nomenclaturas para o mesmo conceito e, conseqüentemente, reduza os ruídos informacionais (Davanzo; Moreira, 2019).

Ao mesmo tempo que o controle de vocabulário é um recurso essencial na indexação (Alves; Tartarotti; Fujita, 2022a, 2022b), deve-se atentar ao seu viés, pois, muitas vezes, não se adapta às mudanças da linguagem e da cultura (Holstrom, 2024). Além disso, a indexação é um método subjetivo, ou seja, pode ocorrer de duas ou mais pessoas selecionem termos distintos para se referir ao mesmo documento (Camossi; Fujita; Tartarotti; Rodas, 2023).

Paralelamente, o tesauro é um tipo de instrumento de controle de vocabulário dinâmico que objetiva colaborar com o processo da organização da informação, fornecer uma navegação no sistema mais fluida e simplificada, além de servir como ponte entre o usuário e o sistema (Davanzo; Moreira, 2019). Ele permite a recuperação da informação eficaz e orienta indexadores e buscadores na seleção de termos para expressar conceitos adequados em suas consultas (Mazzocchi, 2017). Além disso, como instrumento de indexação, o uso do tesauro pode ser vantajoso na “[...] condução do processo por vocabulário controlado, bem como a correlação semântica ser orientada pelas relações dos termos” (Martines; Almeida, 2023, p. 28).

3 Tesauro

Os tesauros são esquemas de controle terminológico utilizados em sistemas de informação. Atendem a uma especialidade do conhecimento, podem ser estruturados como macrotesauros (conceitos amplos) ou microtesauros (conceitos específicos) e, além disso, têm abrangência multidisciplinar ou específica (Santos; Costa; Barros; Vital, 2020).

O tesauro pode ser considerado uma ferramenta de representação que se apresenta por diferentes níveis de terminologia, padronização, além de possibilitar o estabelecimento de relação entre o usuário e o conteúdo de um documento (Santos; Barros; Laipelt, 2021).

O tesauro se constitui como um conjunto de termos descritores “[...] que formam um sistema de conceitos inter-relacionados, sendo desenvolvidos com base em conceitos” (Maculan; Aganette, 2017, p. 106). Ele se divide em quatro elementos básicos: léxico, estrutura gramatical, rede paradigmática e rede sintagmática.

De acordo com Mazzocchi (2017, p. 368, tradução própria):

A estrutura relacional de um tesauro baseia-se em métodos que conectam termos com significados relacionados. [...] Embora não sejam projetados para fornecer uma especificação abrangente dos significados dos termos, podem ser considerados ferramentas semânticas operacionais. Isso se deve ao fato de serem compostos por relações semânticas, refletindo, de certa forma, a estrutura conceitual de um campo de conhecimento específico.

De acordo com o Instituto Brasileiro de Informação em Ciência e Tecnologia (1984, p. 1-2), o tesauro possui quatro finalidades principais:

- a) controlar os termos usados na indexação mediante um instrumento que traduza a linguagem natural [...] numa linguagem mais controlada [...];
- b) assegurar, mediante uma linguagem controlada, uma prática consistente [...];
- c) limitar o número de termos necessários atribuídos aos documentos [...];
- d) servir como auxiliar de busca na estratégia de recuperação.

O tesauro pode ser considerado tanto um instrumento de recuperação da informação quanto de organização do conhecimento. Sua estrutura é estabelecida por meio de relações semânticas e funcionais.

A estrutura do tesauro é estabelecida por meio da sistematização das relações entre os conceitos. Essas relações podem ser semânticas – determinadas

pela análise das características dos conceitos – ou conceituais, envolvendo movimentos classificatórios (Santos; Costa; Barros; Vital, 2020).

4 Metodologia

A presente pesquisa é de natureza qualitativa, do tipo exploratório e bibliográfico.

A pesquisa é qualitativa por investigar o fenômeno do Tesauro UNESP como objeto de pesquisa, preocupando-se com a qualificação dos dados e não com a quantificação.

Além disso, pode ser considerada uma pesquisa exploratória por explorar o Tesauro UNESP a fim de obter uma visão geral e uma aproximação do objeto de estudo, identificando o controle de vocabulário. É um tipo de pesquisa que não exige rigidez no planejamento, o que possibilita o esclarecimento e a modificação das ideias a respeito do tesauro.

4.1 Tesauro UNESP

O tesauro foi escolhido para a análise por ser um instrumento multidisciplinar. A UNESP é uma instituição que abrange diversas áreas do conhecimento e, por ser *multicampi*, possui uma diversidade intrínseca. Essa diversidade exige maior controle, uma vez que, por meio do uso do sistema, é possível identificar suas necessidades, limitações e problemas relacionados ao controle de vocabulário.

O Tesauro UNESP, objeto de pesquisa a ser analisado, é um vocabulário controlado da Universidade Estadual Paulista. Foi construído pelo Grupo de Linguagem UNESP em 2013, devido à necessidade de uma linguagem “[...] atualizada e hierarquizada que representasse as diferentes áreas temáticas presentes no acervo de bibliotecas da Universidade” (Fujita, 2016a).

Seus objetivos são oferecer uma interface integrada com as bases de dados da UNESP (Santos, Moreira e Fujita, 2018) e estabelecer conformidade e unificação para os usuários do sistema, visando à adequação para usos futuros (Fujita, 2016b).

O Tesauro UNESP possibilita ao usuário optar entre a generalização ou a especificidade na busca, além de oferecer ao bibliotecário uma linguagem que

permite remeter o termo da linguagem natural para uma linguagem controlada no sistema (Fujita, 2016a).

O tesauro é multidisciplinar, está nos idiomas português e inglês, e possui fundamentação na *Library of Congress Subject Headings* (LCSH) com termos da Lista de Cabeçalhos de Assunto da Rede Bibliodata (LCARB) (Fujita; Cruz; Patrício, 2017; Fujita; Moreira; Santos; Cruz; Ribas, 2018).

4.2 Percurso metodológico da pesquisa

Para fundamentar e embasar o referencial teórico, foi realizado um levantamento bibliográfico na literatura, por meio das bases de dados BRAPCI e LISTA, no período de 2012 a 2024. A pesquisa se concentrou na indexação, mais especificamente sobre o controle de vocabulário, ou seja, a tradução da linguagem natural para o vocabulário controlado.

As bases de dados BRAPCI e LISTA foram escolhidas por serem da área da biblioteconomia e ciência da informação. Consequentemente, o levantamento trará resultados mais específicos, tendo em vista que o termo indexação é utilizado em diversas áreas, como na economia e ciência da computação. O uso de bases de dados específicas evitou esse tipo de ambiguidade e resultou em uma recuperação da informação mais eficiente.

No levantamento realizado na BRAPCI, a estratégia de busca foi conduzida da seguinte forma: utilizou-se o termo “indexação” como palavra-chave no campo de busca, com delimitação para os anos de 2012 a 2024, e foram selecionadas as coleções de revistas brasileiras e estrangeiras.

Na LISTA, a estratégia de busca foi definida da seguinte forma: KW index* (KW é a abreviação de *keywords*, que significa palavras-chave em português), com os seguintes limitadores: texto completo, revistas acadêmicas (analisadas por especialistas), data de publicação entre janeiro de 2012 e dezembro de 2024, tipo de publicação ‘revista acadêmica’ (*academic journal*), tipo de documento ‘artigo’ (*article*) e idiomas português, espanhol e inglês.

O total dos artigos levantados na BRAPCI foram de 241 e na LISTA 828. Por se tratar de um número grande para ser tratado em um curto período, foram

selecionados apenas artigos que se pretendia utilizar no referencial teórico, ou seja, aqueles que abordavam a definição, o uso, a descrição, os conceitos e as relações dos conceitos trabalhados na presente pesquisa, visando ao contexto geral, ao aprofundamento e à sistematização da literatura.

Para a escolha dos trabalhos que seriam utilizados no referencial teórico, realizou-se a leitura dos resumos, títulos e palavras-chave, como forma de identificar se esses textos eram compatíveis com as temáticas que se pretendia abordar no presente artigo. Depois, para uma segunda verificação, realizou-se uma leitura *scanning* nos textos selecionados e, assim, sua posterior condensação e apresentação crítica no trabalho.

Esse levantamento foi apresentado em síntese por meio do referencial teórico, enfatizando a importância do controle de vocabulário. Os textos utilizados estavam publicados em periódicos nos idiomas português e inglês.

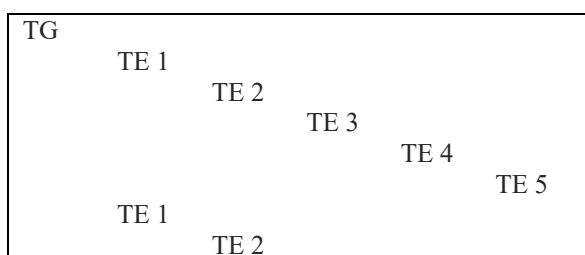
Foi explorado e detalhado o conceito “Deus” no Tesauro UNESP e suas respectivas ramificações, considerando a norma de controle de vocabulário estabelecida pela ISO 25964-1 (International Organization for Standardization, 2011) para vocabulário controlado, especificamente para tesauros de recuperação da informação.

A escolha do termo analisado justifica-se por sua extrema relevância para a discussão sobre as representações nos sistemas de organização do conhecimento. Os sistemas tradicionais representam predominantemente a cultura ocidental, que, no entanto, muitas vezes tem uma visão distorcida da realidade. Consequentemente, essa parcialidade se reflete em um sistema que não representa grupos e culturas não alinhados a ele (Miranda; Silva, 2019). As religiões nesses sistemas são representadas sob um viés cristão-católico, inclusive no Tesauro UNESP. Isso, de acordo com Trivelato e Moura (2017, local. 2), expressa o “[...] visível desequilíbrio das Religiões e das alteridades não hegemônicas nos Sistemas de Organização do Conhecimento (SOC)”. Dessa forma, reforça-se a necessidade de discutir o tema, considerando sua estrutura e uma visão crítica de sua representação.

As ramificações analisadas foram de até cinco níveis, considerando os termos específicos. Isso significa que foram analisadas cinco especificações ou cinco subordinações do termo geral.

Mais especificamente, como é apresentada na Figura 1, o TG é o termo geral, TE é o termo específico e os recuos representam a subordinação do termo inferior para com o termo superior:

Quadro 1 - Representação dos níveis analisados



Fonte: Elaborado pela autora.

Essas ramificações foram identificadas no próprio tesauro e analisadas conforme instruções da norma ISO 25964-1. A análise se baseou em verificar se o tesauro utilizou as recomendações da International Organization for Standardization (2011). Essas orientações indicam que, na escolha do termo preferido, por exemplo, o controle de vocabulário é essencial, pois um conceito pode ser expresso de muitas formas. No entanto, é importante levar em consideração o contexto, ou seja, os aspectos culturais e éticos.

Outra recomendação da ISO 25964-1, verificada no Tesauro UNESP, é se os termos e significados estão em consonância ao escopo do sistema. Uma das formas de se verificar é por meio dos próprios termos e das relações estabelecidas entre eles. Quando a estrutura não é suficiente para representar o escopo, recomenda-se o uso de uma nota de escopo para a definição do significado escolhido (International Organization for Standardization, 2011). Sob uma perspectiva crítica, as notas de escopo podem definir conforme o contexto pretendido.

A ISO 25964-1 recomenda a seleção de um termo preferido quando há mais de um para representar um conceito, colocando os demais como ponto de

acesso a ele (International Organization for Standardization, 2011). É necessário levar em consideração o contexto que se pretende representar, pois podem existir termos preferidos distintos dependendo da cultura. O contexto a ser analisado do termo “Deus” e de suas ramificações é o do Tesauro UNESP, ou seja, um contexto universitário e acadêmico. Dessa maneira, serão observados o contexto da aplicação deste termo e suas representações, além das escolhas por termos preferidos e termos não preferidos.

Outra recomendação da ISO 25964-1, que será verificada no Tesauro UNESP é a padronização da apresentação dos termos quanto ao uso do singular e plural, adequando-se ao padrão do idioma; quanto à variação da grafia, utiliza-se a mais aceita; e a preferência pelo uso de letras minúsculas, exceto para nomes próprios, siglas e abreviaturas (International Organization for Standardization, 2011). Além disso, verificar-se-á se o Tesauro UNESP segue essa recomendação e se faz uso de sinais de pontuação ou caracteres especiais, principalmente nos termos preferidos, para que isso não prejudique o processamento do sistema.

Além da escolha do termo preferido, o estabelecimento das relações é algo muito importante que será analisado no tesauro, pois essas relações orientam os usuários na escolha do termo apropriado (International Organization for Standardization, 2011). Ademais, será analisada a semântica que essas relações oferecem ao tesauro para determinar se são ou não suficientes.

Verificar-se-á como se deu as relações estabelecidas no tesauro, como se estabeleceu a relação de equivalência entre termos preferidos e não preferidos; a relação de hierarquia, com os conceitos superordenados e os subordinados; e as relações associativas entre conceitos que não possuem relação hierárquica, mas semântica ou conceitual (International Organization for Standardization, 2011).

Discutiram-se, ainda, os problemas do tesauro relacionados ao controle de vocabulário e apresentaram-se as possíveis melhorias embasadas na norma, tendo em vista a estrutura, as relações do tesauro e o controle dos termos.

Discutiram-se, igualmente, as dificuldades de padronização para o controle de vocabulário, tendo em vista a literatura estudada e discutida no referencial teórico. De modo geral, essas discussões e análises não seguiram uma estrutura definida, limitando-se a verificar se os termos, a relação e a estrutura do

Tesauro UNESP estão de acordo com a ISO 25964-1 (International Organization for Standardization, 2011), além de verificar se esse mesmo tesauro converge com aspectos culturais e éticos necessários a um sistema de organização do conhecimento.

5 Resultados e discussões

O Tesauro UNESP foi exportado de seu portal na data de 13 de setembro de 2022, apresentando-se da seguinte maneira em sua estrutura hierárquica (Quadros 2). Logo adiante, apresentam-se, igualmente, as análises e discussões realizadas.

Quadro 2 - Recorte Tesauro UNESP - 1

DEUS
TÉRMINOS GENÉRICOS
TG Filosofia
TÉRMINOS ESPECÍFICOS
TE7 Confiança em Deus
TE8 Fé
TE9 Evidencia
TE7 Espírito Santo
TE8 Batismo no Espírito Santo
TE8 Pentecostes
TE8 Repouso no espírito
TE7 Feminilidade de Deus
TE7 Milagres
TE7 Mito
TE7 O Sagrado
TE8 Espaço sagrado
TE9 Santuários
TE10 Relicários
TE10 Santuários confucionistas
TE10 Santuários cristãos
TE10 Santuários domésticos
TE8 Santidade
TE9 Consagração
TE7 Politeísmo
TE7 Teologia
TE8 Monoteísmo
TE8 Secularismo

TE9 Ateísmo
TE10 Cristianismo e ateísmo
TE9 Humanismo secular
TE8 Teísmo
TE9 Monoteísmo
TE9 Politeísmo
TE8 Teologia natural
TE9 Criação
TE9 Teleologia
TE10 Criação
TE8 Teologia prática
TE9 Adoração (Religião)
TE10 Advinhação
TE11 Feng-shui
TE11 Geomancia
TE11 Leitura da sorte por nomes
TE11 Leitura da sorte por runas
TE11 Varinha mágica
TE10 Ano litúrgico
TE11 Advento
TE11 Ano litúrgico Sermões
TE11 Corpus Christi
TE11 Cristo Rei, Festa de
TE11 Festa da Apresentação de Nossa Senhora
TE11 Festa da Apresentação do Senhor
TE11 Festa da Ascensão de Nossa Senhora
TE11 Festa da Ascensão do Senhor
TE11 Festa da Exaltação da Santa Cruz
TE11 Festa da Transfiguração do Senhor
TE11 Festa de Pentecostes
TE11 Imaculada Conceição, Festa da
TE11 Natal
TE11 Páscoa
TE11 Quaresma
TE11 Rosário, Nossa Senhora do, Festa de
TE11 Tempo Comum (Ano litúrgico)
TE11 Tempo pascal
TE10 Deus Adoração e amor
TE10 Feticismo
TE11 Sexo Aspectos religiosos
TE10 Fetichismo
TE11 Animismo
TE11 Ídolos e imagens
TE10 Ídolos e imagens Culto
TE11 Iconoclasmo

TE10 Sacrifício
TE11 Auto-sacrifício
TE11 Bode expiatório
TE11 Ex-votos
TE11 Jesus Cristo Expição
TE11 Sacrifício humano
TE9 Arte sacra
TE10 Arquitetura de igrejas
TE11 Arquitetura beneditina
TE11 Arquitetura jesuítica
TE11 Catedrais
TE11 Templos
TE10 Arte agostiniana
TE10 Arte franciscana
TE10 Arte jesuítica
TE10 Árvore da vida
TE10 Cartões religiosos
TE10 Contra-reforma na arte
TE10 Custódias
TE10 Heráldica eclesiástica
TE10 Ícones
TE11 Pintura de ícones
TE10 Ícones bizantinos
TE10 Ícones russos
TE10 Imagem (Teologia)
TE11 Imagem de Deus
TE10 Presépios na arte e tradição cristã
TE11 Natal Teatro (Literatura)
TE11 Presépios
TE11 Santos (Arte)
TE10 Reforma e arte
TE11 Humanismo na arte
TE10 Simbolismo na Bíblia
TE11 Quatro cavaleiros do apocalipse
TE11 Tipologia (Teologia)
TE9 Educação cristã
TE10 Catequese
TE11 Confirmação Preparo e ensino
TE11 Questionamento
TE9 Evangelização
TE10 Cinema na evangelização
TE10 Igreja Crescimento
TE10 Jesus Cristo Métodos de evangelização
TE10 Obras da Igreja junto aos seus ex-membros
TE10 Reavivamentos

TE10 Testemunhos (Cristianismo)
TE10 Videoteipes na evangelização
TE9 Festas religiosas
TE10 Ano litúrgico
TE10 Epifania
TE11 Reis Magos
TE10 Festa do Divino
TE10 Festas juninas
TE10 Folia de Reis
TE10 Natal
TE10 Páscoa
TE10 Rosário, Nossa Senhora do, Festa de
TE9 Igreja Administração
TE10 Igreja Administração de pessoal
TE10 Igreja Finanças
TE11 Assistência social da igreja
TE11 Clero Pensões
TE11 Clero Salários, etc
TE11 Contribuição cristã
TE11 Dízimos
TE10 Igrejas (Edifícios) Manutenção e reparos
TE9 Missões
TE10 Arquitetura missioneira espanhola
TE10 Igreja Crescimento
TE10 Missionários
TE11 Filhos de missionarios
TE11 Missionárias
TE11 Missionários leigos
TE9 NT
TE10 Catamarãs
TE10 Culinária para viagens marítimas
TE10 Dízimos
TE10 Iatismo
TE10 Igreja Administração Legislação
TE10 Literatura didática
TE11 Exemplos
TE11 Fábulas
TE10 Nomes comerciais
TE11 Marca de produtos
TE10 Nomes geográficos
TE11 Nomes de ruas
TE10 Nomes pessoais
[...]
TE10 Onomástica
TE11 Etnologia dos nomes

TE11 Nomes geográficos
TE10 Nomes pessoais
TE11 Anônimos e pseudônimos
TE11 Apelidos
TE11 Leitura da sorte por nomes
TE10 Onomástica
TE11 Etnologia dos nomes
TE11 Nomes geográficos
TE11 Nomes pessoais
TE10 Palavras e expressões
TE11 Citações
TE11 Clichês
TE11 Fraseologia
TE11 Jargão (Terminologia)
TE11 Provérbios
TE11 Slogans
TE10 Rafting (Esporte)
TE10 Remo (Esporte)
TE10 Tribunais eclesiásticos
TE7 Teologia natural
TERMOS RELACIONADOS
TR God
TR Metafísica
TR Monoteísmo
TR Reino de Deus
TR Religião
TR Santíssima Trindade
TR Teísmo

Fonte: Universidade Estadual Paulista ([2013-]).

No tesauro, os termos específicos são referidos por sua sigla TE, acompanhada de uma numeração progressiva de acordo com sua especificidade (por exemplo: TE1, TE2, TE3, e assim por diante). No recorte analisado, essa numeração está incorreta. No primeiro nível de especificidade, os termos estão apresentados como TE7, e os seguintes níveis estão de acordo com essa sequência, ou seja, TE8, TE9 e assim sucessivamente. No entanto, deveria ser iniciado como no exemplo (TE1, TE2, TE3...).

Além disso, observam-se alguns problemas de hierarquia, como, por exemplo, o termo “Politeísmo”, que deveria estar no mesmo nível hierárquico que

o termo “Monoteísmo”. O termo “Páscoa”, por exemplo, deveria estar subordinado ao termo “Tempo pascal”, por fazer parte dele.

A expressão “Arte sacra” ou mesmo “Igreja Administração” subordinada a “Teologia prática” demonstra incoerência, pois a subordinação indica que o subordinado é um tipo de ou faz parte de algo. No entanto, não se pode considerar que “Arte sacra” ou “Igreja Administração” estejam subordinadas à “Teologia prática”, ou mesmo que “Arte sacra” compartilhe características com “Igreja Administração”.

Da mesma maneira, quais são as relações entre os termos “Catamarãs”, “Dízimos”, “Literatura didática”, “Nomes comerciais”, “Rafting (Esporte)” ou mesmo “Tribunais eclesiásticos”? Eles estão no mesmo nível hierárquico e, quando isso acontece, subentende-se que compartilham, de forma similar, aspectos e características.

Ainda em relação à estrutura, alguns exemplos de adversidades no tesauro são apresentados com o termo “Onomástica”, que se refere ao estudo de nomes de lugares e pessoas, isto é, não há necessidade de apresentar os termos “Nomes geográficos” e “Nomes pessoais”. “Catedrais”, citando como caso análogo, poderia estar subordinada a “Templos”, já que se trata de um tipo de templo. A relação entre “Arquitetura beneditina” e “Arquitetura jesuítica”, da mesma forma, não devem ser do mesmo nível que “Catedrais” e “Templos”, pois catedrais e templos podem ter esses tipos de arquiteturas citados. Dessa maneira, são relações de subordinação.

Algumas expressões apresentam palavras desconexas umas das outras em diferentes níveis do tesauro (por exemplo: “Deus Adoração e amor”, “Ídolos e imagens Culto”, “Jesus Cristo Métodos de evangelização”), enquanto outras apresentam conexões (“Batismo no Espírito Santo”, “Presépios na arte e tradição cristã”, “Obras da Igreja Junto aos seus ex-membros”). Ou seja, não há uma padronização se as expressões serão difundidas com ou sem conexões.

Uma sugestão de correção em relação a essas expressões desconexas é que elas podem se referir a termos que possuem atributos, ou seja, uma característica de divisão que serve para especificar um conceito. No Tesauro UNESP, os atributos poderiam ser expressos da seguinte maneira:

Quadro 3 - Sugestão ao Tesauro UNESP

Deus Adoração e amor, seria Deus (Adoração e amor)
Ídolos e imagens Culto, seria Ídolos e imagens (Culto)
Jesus Cristo Métodos de evangelização, seria Jesus Cristo (Métodos de evangelização)

Fonte: Elaborado pela autora.

No entanto, o Tesauro UNESP ainda assim apresenta dificuldades referentes à padronização, pois esses atributos dos termos não foram transmitidos dentro dos parênteses, como mostrado no exemplo.

Outra falta de padronização foi identificada no Quadro 4, a seguir, entre os termos “términos genéricos”, “términos específicos” e “termos relacionados”. As duas primeiras expressões (términos) apresentam uma padronização, porém, na terceira (termos), há distinção quando se referem ao mesmo conceito.

Quadro 4 - Recorte do Tesauro: classe “Deus”

Deus	TERMOS RELACIONADOS
TÉRMINOS GENÉRICOS	TR God
TG Filosofia	TR Metafísica
	TR Monoteísmo
TÉRMINOS ESPECÍFICOS	TR Reino de Deus
TE7 Confiança em Deus	TR Religião
TE8 Fé	TR Santíssima Trindade
TE9 Evidência	TR Teísmo

Fonte: Universidade Estadual Paulista ([2013-]).

A respeito da apresentação dos termos, igualmente, há falta de padronização. Referente às letras maiúsculas e minúsculas, em alguns momentos, a segunda e as subsequentes palavras de uma expressão são apresentadas com a letra minúscula (“Ícones bizantinos”, “Cinema na evangelização”, “Árvore da vida”, “Arquitetura missionária espanhola”, “Culinária para viagens marítimas”). Em outros momentos, com a letra maiúscula (“Igreja Crescimento”, “Igreja Administração Legislação”, “Igreja Administração de pessoal”, “Igreja Finanças”, “Igrejas (Edifícios) Manutenção e reparos”).

Outra falta de padronização refere-se a termos que apresentam erros ortográficos e de acentuação (“Evidencia”, “Idolos”, “Filhos de missionários”, “Fabulas”), e erro de escrita (“Assisteência social da igreja”).

Observam-se problemas relacionados a sinônimos, como “Contribuição cristã” e “Dízimos”, “Clero Pensões” e “Clero Salários, etc”, que poderiam ser consolidados em uma única expressão. Além disso, não há conexão lógica em o termo “Missionárias” estar subordinado ao termo “Missionários”, ou mesmo em o termo “NT” e suas subdivisões (“Culinária para viagens marítimas”, “Nomes comerciais”, “Rafting (Esporte)” e outros) estarem dentro da classe “Deus”.

Em tesauros, normalmente, o termo NT refere-se a *narrower term* (ou, em português, TE, termo específico). Em contextos religiosos, mais especificamente cristão, NT é a sigla para a segunda parte da Bíblia, o Novo Testamento. Compreende-se, que não se trata de nenhum dos dois, pois os termos apresentados não são específicos de “Teologia prática”, apesar de estarem subordinados na estrutura hierárquica, e não se trata de um termo bíblico, já que não refletem esse contexto.

Em última análise, a temática do recorte analisado se refere a um termo e suas respectivas subdivisões que remetem à religião, de forma geral. Como mencionado na justificativa do trabalho, grande parte dos tesauros e sistemas de organização do conhecimento tradicionais, no geral, classificam os termos “Deus” e “Religião” como exclusivamente católico-cristão, o que é um grande problema em um tesauro e, igualmente, é um problema apresentado no Tesauro UNESP.

No tesauro analisado apenas uma subdivisão (“Teologia”) apresenta termos que englobam outras manifestações de crenças, como o “Secularismo”, que inclui o “Ateísmo”; o termo “Teísmo”, que inclui o “Politeísmo”; em “Teologia prática”, dentro de “Adoração (Religião)” há o termo “Advinhação”, que inclui termos específicos que não são do catolicismo-cristão; ademais, os termos “Falicismo” e “Fetichismo”. Entretanto, há diversas subdivisões e detalhamentos referentes aos termos “Arte sacra”, com “Arquitetura de igrejas”, “Árvore da vida”, “Cinema na evangelização”, “Igreja Crescimento”, “Jesus Cristo Métodos de evangelização”, “Páscoa”, “Rosário, Nossa Senhora do, Festa de”, por exemplo.

A escolha dos termos demonstra uma preferência de representação, e essa escolha exclui denominações cristãs que não sejam católicas e exclui religiões que não são cristãs. A construção de um único discurso válido fortalece “[...] toda espécie de discursos de ódio, oriundos das tendências das instituições que lidam com a informação a disponibilizarem as mesmas segundo a perspectiva dominante” (Miranda; Silva, 2019, p. 95).

6 Conclusão

O controle de vocabulário é a essência do tesauro. Sem ele, não há tesauro. Sua principal função é apoiar na recuperação da informação para o pesquisador e o indexador na escolha do mesmo termo para um conceito, o que sugere a necessidade de padronizar os termos de um campo por meio da representação dos conceitos, de forma que as relações estejam explícitas.

O Tesauro UNESP apresenta diversos problemas, como a falta de padronização dos termos (em um momento, utiliza-se o termo de uma forma e, em outro, de outra), falta de padronização nos símbolos utilizados (TE), expressões desconexas, inconsistência na apresentação dos termos em letras maiúsculas e minúsculas, ausência de controle de sinônimos, falta de padronização nas hierarquias, erros de ortografia, entre outros.

Alguns problemas, como erros de ortografia, são considerados básicos e deveriam estar ausentes em um sistema de controle de vocabulário, que tem como função servir de exemplo-padrão para um domínio específico.

Para esses problemas, foram sugeridas correções ao longo do trabalho, assim como sugestões para a adoção de padrões e normas na construção de tesouros, como a ISO 25964-1 (International Organization for Standardization, 2011). As normas podem orientar o tesauro e, quando aplicadas, também podem orientar um sistema com controle de vocabulário, pois “[...] a ausência de um vocabulário controlado, sua falta de atualização ou até mesmo um vocabulário que não represente a linguagem da comunidade usuária, geram falhas nessa comunicação e concorrem para a insatisfação na recuperação da informação” (Fujita; Neves; Dal’Evedove, 2017, p. 161-162).

Outro problema, não tão básico, mas considerado recorrente nos tesauros, é a preferência cultural ocidental. Isso ainda é um grande desafio para os SOCs em geral, pois, muitos tesauros seguem outros sistemas tradicionais que possuem um viés ocidentalizado. Dessa maneira, essa disseminação possui um encadeamento muito mais profundo. No entanto, ainda é necessário haver esforços das instituições para a implementação de sistemas mais plurais e que incluam identidades marginalizadas.

Esses problemas apresentados não são exclusivos do Tesauro UNESP, mas são desafios frequentes em tesauros institucionais. Outra recomendação que pode ajudar a sanar e minimizar esses problemas é a necessidade da implementação de políticas de revisão, atualização e inclusão de identidades marginalizadas, consistindo na apresentação de um tesauro que busque melhorias estruturais e, igualmente, culturais. Assim, diminuem-se, inclusive, as preferências por uma cultura em detrimento de outras.

O Tesauro UNESP é um sistema especializado que serve de apoio à pesquisa, ensino e extensão da universidade. Sua melhoria proporciona uma recuperação eficaz da informação no sistema de busca das bibliotecas UNESP pelos usuários, além de facilitar o autoarquivamento das monografias pelos alunos e a indexação pelos bibliotecários indexadores da instituição. Dessa maneira, reconhece-se a importância e o escopo do Tesauro UNESP, assim como a relevância do controle de vocabulário.

Referências

- AGUIAR, F. L.; TÁLAMO, M. F. G. M. O controle de vocabulário da linguagem orgânico-funcional. *Acervo*, Rio de Janeiro, v. 25, n. 1, p. 117-138, 2012.
- ALVES, L. S. S.; TARTAROTTI, R. C. D. E.; FUJITA, M. S. L. Avaliação de vocabulário controlado para a representação e recuperação de teses e dissertações em repositório institucional. *RICI: Revista Ibero-Americana de Ciência da Informação*, Brasília, v. 15, n. 1, p. 283-294, 2022a. Disponível em: <https://doi.org/10.26512/rici.v15.n1.2022.42496>. Acesso em: 4 set. 2025.
- ALVES, L. S. S.; TARTAROTTI, R. C. D. E.; FUJITA, M. S. L. Avaliação do uso de vocabulário controlado em repositórios institucionais.

Informação@Profissões, Londrina, v. 11, n. 1, p. 52-77, 2022b. Disponível em: <https://doi.org/10.5433/2317-4390.2022v11n1p52>. Acesso em: 4 set. 2025.

BARITÉ, M. Control de vocabulario: Orígenes, evolución y proyección. **Ciência da Informação**, Brasília, v. 41, n. 1, p. 95-119, 2014. Disponível em: <https://doi.org/10.18225/ci.inf.v43i1.1421>. Acesso em: 4 set. 2025.

CAMOSSO, G.; FUJITA, M. S. L.; TARTAROTTI, R. C. D. E.; RODAS, C. M. As técnicas de search engine optimization e os elementos da indexação no processo de ranqueamento em mecanismos de busca. **Em Questão**, Porto Alegre, v. 29, p. 1-25, 2023. Disponível em: <https://doi.org/10.1590/1808-5245.29.124785>. Acesso em: 4 set. 2025.

CHAUMIER, J. Indexação: conceito, etapas e instrumentos. Tradução de José Augusto Chaves Guimarães. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n.1/2, p. 63-79, 1988.

DAVANZO, L.; MOREIRA, W. Vocabulário controlado para arquivos: análise de viabilidade e propostas. **Ágora**, Florianópolis, v. 1, n. 2, p. 1-19, 2019.

FUJITA, M. S. L. A linguagem documentária na negociação de uma política para bibliotecas universitárias: procedimentos e estratégias da pesquisa-ação integral. **Revista Conhecimento em Ação**, Rio de Janeiro, v. 1, n 1, p. 2-17, 2016b. Disponível em: <https://doi.org/10.47681/rca.v1i1.3555>. Acesso em: 4 set. 2025.

FUJITA, M. S. L.; CRUZ, M. C. A.; PATRÍCIO, B. O. M. A construção de tesauros na perspectiva dos manuais de indexação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais [...]**. Marília: UNESP, 2017.

FUJITA, M. S. L.; MOREIRA, W.; SANTOS, L. B. P.; CRUZ, M. C. A.; RIBAS, R. R. B. Construction and evaluation of hierarchical structures of indexing languages for online catalogs of libraries: an experience of the São Paulo State University. **Knowledge Organization**, Baden-Baden, v. 45, n. 3, p. 220-231, 2018. Disponível em: <https://doi.org/10.5771/0943-7444-2018-3-220>. Acesso em: 4 set. 2025.

FUJITA, M. S. L.; NEVES, D. A. B.; DAL'EVEDOVE, P. R. (org.). **Leitura documentária: estudos avançados para a indexação**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2017.

FUJITA, M. S. L. (org.). **Política de indexação para bibliotecas: elaboração, avaliação e implantação**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2016a.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E
TECNOLOGIA. **Diretrizes para elaboração de tesauros monolíngües.**
Brasília: IBICT, 1984.

HOLSTROM, C. Critical control: how different forms of vocabulary control aid and hinder novice indexers aiming to support racial and social justice. **Knowledge Organization**, Baden-Baden, v. 51, n. 7, p. 521-542, 2024. Disponível em: <https://doi.org/10.5771/0943-7444-2024-7-521>. Acesso em: 4 set. 2025.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-1: Information and documentation - thesauri and interoperability with other vocabularies: part 1 thesauri for information retrieval.** Geneve: ISO, 2011.

KLANN, A.; RAMOS, F. B. Percepções de bibliotecários sobre indexação e políticas de indexação. **BiblioCanto**, Natal, v. 9, n. 1, p. 1-22, 2023. Disponível em: <https://doi.org/10.21680/2447-7842.2023v9n1ID29835>. Acesso em: 4 set. 2025.

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

LIU, Y. H.; WACHOLDER, N. Evaluating the impact of MeSH (Medical Subject Headings) terms on different types of searchers. **Information Processing and Management**, Amsterdam, v. 53, n. 4, p. 851-870, 2017. Disponível em: <https://doi.org/10.1016/j.ipm.2017.03.004>. Acesso em: 4 set. 2025.

MACULAN, B. C. M. S.; AGANETTE, E. C. Desambiguação de relações em tesauros e o seu reuso em ontologias. **Ciência da Informação**, Brasília, v. 46, n. 1, p. 102-119, 2017. Disponível em: <https://doi.org/10.18225/ci.inf.v46i1.4017>. Acesso em: 4 set. 2025.

MAFRA, H. F.; CORDEIRO, R. I. N. Indexação de livros juvenis: a etapa da análise de assunto. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 36, n. 2, p. 48-68, 2022. Disponível em: <https://doi.org/10.14295/biblos.v36i2.14521>. Acesso em: 4 set. 2025.

MARTINES, A. R.; ALMEIDA, C. C. Fundamentos linguísticos da indexação: uma revisão. **Scire: Representación y Organización del Conocimiento**, Zaragoza, v. 29, n. 2, p. 25-37, 2023. Disponível em: <https://doi.org/10.54886/scire.v29i2.4913>. Acesso em: 4 set. 2025.

MAZZOCCHI, F. Relations in KOS: is it possible to couple a common nature with different roles? **Journal of Documentation**, Leeds, v. 73, n. 2, p. 368-383, 2017. Disponível em: <https://doi.org/10.1108/JD-05-2016-0063>. Acesso em: 4 set. 2025.

MIRANDA, M. L. C.; SILVA, F. G. Religião e cultura periféricas: a representação do islamismo na classificação decimal de Dewey. **Logeion: Filosofia da Informação**, Rio de Janeiro, v. 5, n. 2, p. 86-120, 2019. Disponível em: <https://doi.org/10.21728/logeion.2019v5n2.p86-120>. Acesso em: 4 set. 2025.

OLIVEIRA, T. N.; SANTOS, R. F. Indexação e recuperação em repositórios institucionais brasileiros: recomendações de melhorias para a difusão do conhecimento científico. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 38, n. 1, p. 139-160, 2024. Disponível em: <https://doi.org/10.63595/biblos.v38i1.17881>. Acesso em: 4 set. 2025.

RUBI, M. P. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. *In*: FUJITA, M. S. L. (org.). **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias**. [...]. São Paulo: Cultura Acadêmica, 2009.

SANTOS, A.; COSTA, A.; BARROS, C. M.; VITAL, L. P. Representação terminológica da população negra em tesouros. **Informação e Informação**, Londrina, v. 25, n. 1, p. 254-275, 2020. Disponível em: <https://doi.org/10.5433/1981-8920.2020v25n1p254>. Acesso em: 4 set. 2025.

SANTOS, A. C. R.; BARROS, T. H. B.; LAIPELT, R. C. F. Organização do conhecimento e gestão dos tesouros na web semântica. **Convergências em ciência da informação**, Aracaju, v. 4, p. 1-31, 2021. Disponível em: <https://doi.org/10.33467/conci.v4i.16493>. Acesso em: 4 set. 2025.

SANTOS, J. C. F.; MOREIRA, W.; FUJITA, M. S. L. Tesauro Unesp: integração do registro de autoridade para o TemaTres. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais** [...]. Londrina: UEL, 2018.

SILVA, M. R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas metodológicas. **Transinformação**, Campinas, v. 16, n. 2, p. 133-161, 2004.

SMIT, J. W. Vocabulário controlado e controle de vocabulário em arquivos. **Acervo**, Rio de Janeiro, v. 31, n. 3, p. 46-56, 2018.

TRIVELATO, R. M. S.; MOURA, M. A. Identidade, religião e a formação discursiva nos sistemas de classificação bibliográfica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais** [...]. Marília: UNESP, 2017.

VAN DER LAAN, R. H.; FERREIRA, G. I. S. Tesouros e terminologia. *In*: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E

DOCUMENTAÇÃO, 19., 2000, Porto Alegre. **Anais [...]**. Porto Alegre: Lume Repositório Digital UFRGS, 2000.

UNIVERSIDADE ESTADUAL PAULISTA. **Tesauro Unesp**. São Paulo: UNESP, [2013-].

Vocabulary control in the UNESP Thesaurus

Abstract: Indexing is a practice aimed at information retrieval and is divided into the following processes: analysis, synthesis, and representation. The latter is considered paramount and, to be employed, requires a vocabulary control system, such as a thesaurus. A thesaurus is understood as a specialized system of related terms. It represents the terms identified in the indexing process through vocabulary control. The UNESP Thesaurus aims to control the terms used by its academic community members; however, it presents some issues. Therefore, the goal is to understand the importance of vocabulary control, identify, discuss, and propose improvements. To this end, a literature review was conducted in the LISTA and BRAPCI databases, and the thesaurus was analysed in accordance with the ISO 25964-1 standard and relevant literature. Several problems were identified in the thesaurus, and possible solutions were proposed. It is concluded that the thesaurus is a very important tool and that vocabulary control is vital, with improvements needed in the language tool.

Keywords: vocabulary control; thesaurus; UNESP thesaurus

Declaração de autoria

Concepção e elaboração do estudo: Samantha Augusta dos Santos de Jesus

Coleta de dados: Samantha Augusta dos Santos de Jesus

Análise e interpretação de dados: Samantha Augusta dos Santos de Jesus

Redação: Samantha Augusta dos Santos de Jesus

Revisão crítica do manuscrito: Samantha Augusta dos Santos de Jesus

Declaração de disponibilidade de dados

Todo o conjunto de dados que dá suporte aos resultados deste estudo foi publicado no próprio artigo.

Autoria para correspondência

Samantha Augusta dos Santos de Jesus

sas.jesus@unesp.br

Editor-chefe

Thiago Henrique Bragato Barros

Como citar

JESUS, Samantha Augusta dos Santos de. Controle de vocabulário no tesauro UNESP. **Em Questão**, Porto Alegre, v. 31, e-145272, 2025. DOI: <https://doi.org/10.1590/1808-5245.31.145272>.

Recebido: 16/01/2025

Aceito: 04/08/2025

