



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília



**CULTURA
ACADÊMICA**
Editora

Taxonomia de áreas do conhecimento da Unesp: organização, desenvolvimento e aplicação de estruturas hierárquicas

Mariângela Spotti Lopes Fujita; Luciana Beatriz Piovezan Rio Branco;
José Carlos Francisco dos Santos; Maria Marlene Zaniboni;
Minervina Teixeira Lopes

Como citar: FUJITA, M. S. L. *et al.* Taxonomia de áreas do conhecimento da Unesp: organização, desenvolvimento e aplicação de estruturas hierárquicas. *In:* FUJITA, M. S. L.; MOREIRA, W. **Manual do planejamento, construção e manutenção do Tesouro Unesp para bibliotecas:** do conceitual a práxis. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021. p. 149-192.
DOI: <https://doi.org/10.36311/2021.978-65-5954-069-3.p149-192>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-Non Commercial-ShareAlike 3.0 Unported.

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição - Uso Não Comercial - Partilha nos Mesmos Termos 3.0 Não adaptada.

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 3.0 Unported.

5. TAXONOMIA DE ÁREAS DO CONHECIMENTO DA UNESP: DESENVOLVIMENTO E APLICAÇÃO DE ESTRUTURAS HIERÁRQUICAS

Mariângela Spotti Lopes Fujita

Luciana Beatriz Piovezan Rio Branco

José Carlos Francisco dos Santos

Maria Marlene Zaniboni

Minervina Teixeira Lopes

5.1 Introdução à Taxonomia

5.2 Desenvolvimento das estruturas hierárquicas das áreas de conhecimento da taxonomia

5.2.1 Identificação de áreas de conhecimento

5.2.2 Identificação e seleção de estruturas hierárquicas de linguagens de indexação

5.2.2.1 Identificação de estruturas hierárquicas de linguagens de indexação

5.2.2.2 Seleção de linguagens de indexação e elaboração de metodologia de construção das estruturas hierárquicas de áreas de conhecimento

5.2.2.3 Formação das estruturas hierárquicas de áreas de conhecimento com uso das linguagens

5.2.2.3.1 Fase de compatibilização das estruturas hierárquicas das linguagens de indexação VocaUSP, LCSH e TBN:

5.2.2.3.2 Fase de consolidação da estrutura hierárquica do tesouro Unesp.

5.3 Aplicação das estruturas hierárquicas de áreas de conhecimento para a classificação dos termos do tesouro

5.3.1 Classificação automática dos descritores e não descritores retrospectivos em categorias e subcategorias (metatermos)

5.3.2 Classificação intelectual dos descritores e não descritores retrospectivos em categorias e subcategorias (metatermos)

5.4 Recomendações para uso da Taxonomia de áreas de conhecimento da Unesp aos desenvolvedores do Tesouro Unesp

Apêndice C - Manual de compatibilização de termos com a taxonomia

Referências

5.1 INTRODUÇÃO À TAXONOMIA

O desenvolvimento e aplicação da Taxonomia de áreas do conhecimento foi realizado em virtude da necessidade de criação de termos novos e a consequente definição de relações de equivalência hierárquicas e associativas. A taxonomia é uma ferramenta que auxilia na descoberta da área de conhecimento de pertencimento de termos. Segundo Campos e Gomes (2008, p.1) “Taxonomia é, por definição, classificação, sistemática e está sendo conceituada no âmbito da Ciência da Informação como ferramenta de organização intelectual”.

A taxonomia significa, no âmbito da pesquisa sobre “Linguagem de indexação para bibliotecas na perspectiva da política de indexação”, o conjunto de estruturas hierárquicas responsável por classificar o vocabulário que abrange as áreas do conhecimento existentes na Unesp, relativas às atividades de ensino, pesquisa e extensão. Nesse sentido, há concordância com o fato de que “[...] a taxonomia é uma estrutura hierarquizada que representa parte do conhecimento para atender a objetivo específico de organização e representação das estruturas de conhecimento” (SHINTAKU; FUJITA; SCHIESSEL, 2019, p.34).

Cada estrutura hierárquica, composta por níveis hierárquicos, representa uma área de conhecimento disciplinar existente na universidade e é ordenada categoria e subcategoria correspondentes à essa área de conhecimento. Em cada categoria é classificado o vocabulário correspondente composto dos termos e suas relações sintáticas e semânticas que representam os objetos de conhecimento.

No processo de elaboração da taxonomia foram discutidas as etapas de desenvolvimento e sua aplicação com acompanhamento de progressos e dificuldades cujos resultados (FUJITA; MOREIRA; SANTOS, L.; CRUZ; RIBAS, 2019) propiciaram aspectos de avaliação das estruturas hierárquicas das linguagens utilizadas referentes à organização dos descritores superordenados e subordinados que servirão para a sistematização de condutas operacionais do manual de linguagem de indexação para catálogos on-line de bibliotecas. Para isso, foram realizadas reuniões regulares do Grupo de linguagem da Rede de Bibliotecas da UNESP para discussão das dúvidas e soluções à medida que as estruturas hierárquicas foram elaboradas. As dúvidas sobre o desenvolvimento e aplicação da Taxonomia foram compartilhadas em arquivo criado para essa finalidade e acessível por todos.

5.2 DESENVOLVIMENTO DAS ESTRUTURAS HIERÁRQUICAS DAS ÁREAS DE CONHECIMENTO DA TAXONOMIA

A divisão do trabalho entre os catalogadores e pesquisadores pautou-se pela experiência e familiaridade de trabalho com as áreas de conhecimento e proximidade com os especialistas para contato quando fosse necessário solucionar dúvidas sobre o desenvolvimento das estruturas hierárquicas. Assim, os catalogadores das bibliotecas com coleções de engenharia, por exemplo, se encarregaram de desenvolver as estruturas hierárquicas dessa área de conhecimento.

As etapas de desenvolvimento das estruturas hierárquicas são três: identificação de áreas de conhecimento da universidade; identificação e seleção de taxonomias de linguagens de indexação; e, formação das estruturas hierárquicas de áreas de assunto com uso das linguagens.

5.2.1 IDENTIFICAÇÃO DE ÁREAS DE CONHECIMENTO

Na primeira etapa, o início da construção da taxonomia foi marcado pela discussão sobre a divisão de trabalho entre os membros do Grupo de Linguagem da Rede de Bibliotecas da Unesp e qual o parâmetro que poderia identificar esse marco divisório.

A escolha do parâmetro de ordenação por nomes de disciplinas teve como critérios a facilidade de identificação da área de conhecimento e vocabulário correspondente pelos usuários e pelos catalogadores, além de possibilitar a visualização da organização do conhecimento nas bibliotecas universitárias. Dessa forma, será possível consultar a Linguagem Unesp pela Taxonomia, por uma lista sistemática e lista alfabética.

A Unesp possui 60 áreas de conhecimento em sua estrutura universitária formada por cursos de graduação e de pós-graduação. Realizamos a organização dessa estrutura universitária pela classificação das grandes áreas do conhecimento das agências governamentais de fomento à pesquisa no Brasil, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Esse foi o primeiro parâmetro adotado para a definição das áreas do conhecimento cujas estruturas hierárquicas passaram a ser desenvolvidas como se visualiza no Quadro 1:

Quadro 1 - Divisão por áreas do conhecimento antes do desenvolvimento das estruturas hierárquicas

Grandes áreas de conhecimento (7) (CNPq/CAPES)	Áreas de conhecimento (60) (Cursos de graduação da Unesp)
Engenharias (10)	Engenharia Aeronáutica, Engenharia Civil, Engenharia de Controle e Automação, Engenharia de Energia, Engenharia de Materiais, Engenharia de Produção, Engenharia de Telecomunicações, Engenharia Elétrica, Engenharia Mecânica, Engenharia Química.
Ciências Humanas (11)	Arte – Teatro e Artes Cênicas, Artes Visuais, Ciências Sociais, Educação Musical e Música, Filosofia, Geografia, História, Letras, Pedagogia, Psicologia, Relações Internacionais.

Ciências Sociais Aplicadas (14)	Administração Pública, Administração, Arquitetura e Urbanismo, Arquivologia, Biblioteconomia, Ciências Econômicas, Design, Direito, Jornalismo, Museologia, Radialismo, Relações Públicas, Serviço Social, Turismo.
Ciências Biológicas (6)	Ciências Biológicas, Ecologia, Engenharia Ambiental, Engenharia Biotecnológica, Engenharia de Bioprocessos e Biotecnologia, Engenharia de Biosistemas.
Ciências Exatas e da Terra (9)	Ciência da Computação, Engenharia Cartográfica, Estatística, Física, Geologia, Matemática, Meteorologia, Química, Sistemas de Informação.
Ciências Agrárias (7)	Engenharia agrônômica, Engenharia de Alimentos, Engenharia de Pesca, Engenharia Florestal, Engenharia Industrial Madeireira, Medicina Veterinária, Zootecnia.
Ciências da Saúde (11)	Ciências Biomédicas, Educação Física, Enfermagem, Farmácia-Bioquímica, Física Médica, Fisioterapia, Fonoaudiologia, Medicina, Nutrição, Odontologia, Terapia Ocupacional.

Fonte: Elaborado pelos autores.

5.2.2 IDENTIFICAÇÃO E SELEÇÃO DE ESTRUTURAS HIERÁRQUICAS DE LINGUAGENS DE INDEXAÇÃO

A segunda etapa desdobrou-se em duas fases, a primeira de busca pela identificação de estruturas hierárquicas de linguagens de indexação: Terminologia de Assuntos da Biblioteca Nacional (TBN), Library of Congress Subject Headings (LCSH) e Vocabulário Controlado da Universidade de São Paulo (VocaUSP) e a segunda fase pela seleção de linguagens de indexação e elaboração de metodologia de construção.

5.2.2.1 IDENTIFICAÇÃO DE ESTRUTURAS HIERÁRQUICAS DE LINGUAGENS DE INDEXAÇÃO

A busca por parâmetros iniciou-se com a TBN e depois com a LCSH. A linguagem de indexação TBN é disponibilizada pela Biblioteca

Nacional do Brasil e, assim como a LCARB, adota a LCSH como matriz e base terminológica na construção e atualização da linguagem (GRINGS, 2015). Em reunião realizada com responsável pela equipe de construção da TBN, confirmamos que a estrutura hierárquica não existe de modo sistematizado em uma estrutura à parte e que era necessário buscá-la e identificá-la no uso dos seus descritores. Ao procurarmos a taxonomia da LCSH descobrimos que no vocabulário a taxonomia não é aparente, tal como na TBN e que existe uma estrutura classificatória adotada para a definição de sua estrutura hierárquica, disponível no site da Biblioteca Nacional do Brasil. Tal estrutura é disponível pelo serviço Library of Congress classification web ao qual subscrevemos por assinatura para o desenvolvimento do trabalho.

5.1.2.2 SELEÇÃO DE LINGUAGENS DE INDEXAÇÃO E ELABORAÇÃO DE METODOLOGIA DE CONSTRUÇÃO DAS ESTRUTURAS HIERÁRQUICAS DE ÁREAS DE CONHECIMENTO

Para isso, os pesquisadores do Grupo de Linguagem da Rede de Bibliotecas da Unesp realizaram pesquisa exploratória sobre estruturas hierárquicas de linguagens de indexação. Adotou-se o método de comitê ascendente da Diretriz ANSI/NISO-2005

(NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005) para compilação de vocabulários controlados com uso do Modelo metodológico integrado, realizando a comparação das hierarquias da LCSH, TBN, Lista de Cabeçalhos de assuntos da rede BIBLIODATA (LCARB) e VocaUSP nas áreas de Física e Matemática (FUJITA; SANTOS, L. 2016).

Foram confirmados os seguintes resultados: impossibilidade de utilizar a LCARB por não possuir a visualização de relacionamentos hierárquicos entre termos específicos e gerais; a viabilidade das demais linguagens quanto ao uso de suas estruturas hierárquicas, definição de procedimentos de como realizar o desenvolvimento das estruturas hierárquicas das áreas de assunto. A pesquisa definiu, também, o uso da estrutura hierárquica de categorias e subcategorias do Vocabulário do SIBI/USP, por ser esta mais assemelhada à estrutura da organização do conhecimento na universidade pública paulista e, sobretudo pela garantia

cultural das áreas de conhecimento, embora não tenha a mesma matriz terminológica da LCSH, como é o caso da LCARB e da TBN.

No caso do Vocabulário do SIBI/USP (VocaUSP), a taxonomia é visível e disponível para consulta no site da linguagem de indexação.

Para as áreas de Ciências da Saúde manteve-se a consulta complementar às linguagens de indexação especializadas, Medical Subject Headings (MeSH) e Descritores em Ciências da Saúde (DeCS), ambas com mesmo fundamento linguístico e estrutural. A linguagem trilingue DeCS é mantida e atualizada pela Biblioteca Virtual de Saúde (BVS) do Ministério da Saúde do Brasil a partir do MeSH para a indexação e recuperação da literatura científica brasileira.

5.2.2.3 FORMAÇÃO DAS ESTRUTURAS HIERÁRQUICAS DE ÁREAS DE CONHECIMENTO COM USO DAS LINGUAGENS

A construção das estruturas hierárquicas, portanto, iniciou-se pela definição das categorias e subcategorias que formam a Taxonomia da linguagem de indexação utilizando como parâmetros as taxonomias da LCSH, da TBN e do VocaUSP.

Para compreensão do uso do serviço Library Congress Classification web foi elaborado um manual de orientação resumido com base no tutorial disponível no site <https://classificationweb.net/>. O manual resumido contém procedimentos de uso do serviço com a finalidade de consulta de categorias (classes e subclasses), relações hierárquicas entre descritores, subdivisões, notas, número de classificação, associação de descritores aos números de classificação da CDD, busca de registro de autoridade com finalidade de importação pelo software ALEPH utilizado pela rede de bibliotecas universitárias na construção do catálogo online.

A criação e construção das estruturas hierárquicas do Tesauro Unesp foram operacionalizadas em duas fases: compatibilização das estruturas hierárquicas das linguagens de indexação VocaUSP, LCSH e TBN e consolidação da estrutura hierárquica do tesauro Unesp.

5.2.2.3.1 FASE DE COMPATIBILIZAÇÃO DAS ESTRUTURAS HIERÁRQUICAS DAS LINGUAGENS DE INDEXAÇÃO VOCAUSP, LCSH E TBN:

A) Orientações gerais de compatibilização com a planilha Excel: A compatibilização foi realizada mediante uso do recurso “agrupar” na planilha Excel com as seguintes orientações:

- Para usar o recurso “agrupar”:
 - Incluir linha (número de linhas referentes à quantidade de descritores subordinados);
 - Colar os descritores na coluna seguinte (próximo nível hierárquico);
 - No cabeçalho do Excel, aba “Dados”, utilizar o recurso “Agrupar”.
- Criar a Estrutura Hierárquica de cada área em uma planilha Excel usando o recurso “agrupar”. Cada coluna da planilha Excel conterá um nível hierárquico, na coluna B ficará o primeiro nível (categorias), na coluna C o segundo nível e na coluna D o terceiro nível, e assim sucessivamente.
- Para criação da hierarquia dos termos, utilizar as estruturas hierárquicas das linguagens de indexação na seguinte sequência: primeiro o VocaUSP, em segundo momento, a análise é da LCSH e, caso necessário, a TBN.
- Código de cores para identificação das linguagens de indexação (Quadro 2): manter um mesmo padrão de cor para cada linguagem (Azul para LCSH; Vermelho para TBN, Verde para VocaUSP e pink para a consolidação da estrutura hierárquica do Tesouro Unesp); quando o termo constar em mais de uma linguagem, colocar lado a lado na mesma célula, cada termo com sua respectiva cor.

Quadro 2 - Exemplo de uso do código de cores

DIREITO AÉREO - DIREITO AÉREO	
	CÓDIGO BRASILEIRO DE TELECOMUNICAÇÕES - CÓDIGO BRASILEIRO DE TELECOMUNICAÇÕES

	DIREITO AÉREO MILITAR - DIREITO AÉREO MILITAR
	DIREITO ESPACIAL – Space law –Direito espacial - Direito espacial

Fonte: Elaborado pelos autores.

B) Orientações específicas da compatibilização nas áreas de conhecimento:

A compatibilização de estruturas hierárquicas foi realizada conforme orientações gerais para todas as áreas de conhecimento, porém, destacamos casos descritos, a seguir, com abordagens diferenciadas como exemplo para situações similares que, eventualmente, venham a ocorrer. Servem, também, de exemplificação de procedimentos e decisões em áreas de conhecimento tão diferentes entre si.

B1) Caso da área de Engenharia:

Algumas dessas áreas de conhecimento são extensas como é o caso de Engenharia que inclui 13 subáreas de Engenharia: Engenharia Civil, Engenharia de aeronaves, Engenharia de Minas e de petróleo, Engenharia de produção, Engenharia elétrica, Engenharia Geotermal, Engenharia Mecânica, Engenharia Metalúrgica, Engenharia naval e oceânica, Engenharia Química, Engenharia ambiental, Engenharia aeroespacial e Engenharia de pesca. Cada uma dessas subáreas tem suas especificidades e subdivisões hierárquicas como se nota no exemplo de Engenharia Química, considerada como área interdisciplinar porque se compõe de duas disciplinas, Engenharia e Química. O mesmo ocorre com Engenharia de pesca, Engenharia ambiental.

Ainda na área de Engenharia (Quadro 3), constatou-se que a LCSH participa de 7 subáreas em combinação com o VocaUSP no primeiro nível hierárquico (Engenharia Civil, Engenharia de produção, Engenharia elétrica, Engenharia Mecânica, Engenharia Química, Engenharia ambiental e Engenharia aeroespacial) e uma subárea sem combinação (Engenharia Geotermal). Enquanto que o VocaUSP não possui combinação no primeiro nível hierárquico com a LCSH em

5 áreas (Engenharia de Aeronaves, Engenharia de minas e de petróleo, Engenharia Metalúrgica, Engenharia Naval e Oceânica e Engenharia de Pesca). Entretanto, nos segundo e terceiros níveis hierárquicos observam-se combinações das três linguagens.

Quadro 3 - Exemplo da estrutura hierárquica de Engenharia de minas e petróleo, que só se correspondem nos segundo e terceiro níveis das Linguagens

Engenharia de minas e de petróleo		
	Engenharia de Minas / Mining engineering	
		Borehole mining
		Boring
		Detonação / Blasting
		Electricity in mining
		Engenharia do petróleo / Petroleum engineering
		Gas engineering
		Ground control (Mining)
		Hydraulic mining
		Iluminação de minas
		In situ processing (Mining)
		Lasers in mining
	Engenharia de petróleo / Petroleum engineering	
		Avaliação de formações
		Classificação do petróleo
		Completação
		Composição do petróleo
		Constituintes do petróleo
		Electricity in petroleum engineering
		Elevação de poços

Fonte: Elaborado pelos autores.

B2) Caso entre as áreas de Geografia e Cartografia:

A junção das áreas de Geografia e Cartografia em uma única estrutura hierárquica é um exemplo de decisão entre manter a autonomia da Cartografia como área de assunto porque na Unesp é um Curso de Graduação independente do Curso de Geografia ou obedecer a subordinação hierárquica da Cartografia com a Geografia na linguagem VocaUSP ou com Mathematical geography dentro de Geography conforme linguagem LCSH. Prevaleceu as opções de subordinações hierárquicas das linguagens com preferência para a proposta da linguagem VocaUSP por ser a que mais se aproxima do conteúdo de termos do campo semântico da Unesp. Esta decisão define uma importante conduta operacional em relação às opções de subordinação hierárquicas disponíveis pelas linguagens, ou seja, consideramos que devemos, preferencialmente, preservar a autonomia da área de assunto enquanto decisão acadêmica, porém, se na análise das estruturas hierárquicas das linguagens a área estiver subordinada em mais de uma linguagem, principalmente no VocaUSP por ser uma linguagem construída conforme cultura universitária, a subordinação deverá ser mantida.

B3) Casos de subordinação e superordenação de termos nas áreas de assunto: definição de níveis hierárquicos

O primeiro caso é o da hierarquia “Levantamentos/Survey” que na LC e BN é subordinada à Mathematical geography em terceiro nível:

GEOGRAPHY
.MATHEMATICAL GEOGRAPHY
..SURVEY
...AERIAL SURVEY
...AGRICULTURAL SURVEY
...ARCHIVAL SURVEYS

Este exemplo é característico de uma faceta e seus desdobramentos com termos de diferentes áreas de assunto que parecem não ter relação

específica com a Cartografia ou Geografia. No VocaUSP o levantamento se refere apenas aos tipos específicos de levantamentos cartográficos: Levantamentos básicos (cartografia); Levantamentos geodésicos; Levantamentos topográficos; e Reambulação.

Decidimos por definir o termo “Levantamentos/Survey” como segundo nível hierárquico dentro da categoria “Geografia” no mesmo nível de Cartografia. Assim, colocamos o termo “levantamento cartográfico” no terceiro nível e seus tipos no quarto nível. Abaixo de “Levantamento cartográfico” mantivemos os diferentes tipos de “Levantamentos/Survey” até que seja possível deslocá-los para outras categorias. Consideramos que não temos suficiente conhecimento sobre a terminologia desses termos e que seria necessário realizar uma pesquisa termo a termo para realizar deslocamentos dessa natureza. Além disso, o termo “Levantamento/Survey” é um termo que exemplifica um tipo de faceta que pode ser comum em várias categorias como demonstra o exemplo abaixo. Dessa forma, recomendamos a mesma decisão para outros casos como esse.

Geografia / Geography

.Cartografia / Cartography

.Levantamentos / Surveys

..Levantamentos Cartográficos

...Levantamentos Básicos (Cartografia)

...Levantamentos Geodésicos

...Levantamentos Topográficos

...Reambulação

..Aerial surveys / Levantamentos aéreos

..Levantamentos agrícolas / Agricultural surveys

..Architectural surveys

..Levantamentos arquivísticos / Archival surveys

..Arts surveys

..Communication surveys

..Levantamentos ecológicos / Ecological surveys

- ..Levantamentos econômicos / Economic surveys
- ..Employee attitude surveys

O segundo caso para decisão é o exemplo demonstrado pelos seguintes termos:

Geografia agrária: descritor do VocaUSP

Geografia agrícola: descritor da LCSH e da Terminologia da BN

Geografia rural: descritor da LCSH e da Terminologia da BN

Todos os três termos estão no primeiro nível de suas hierarquias, subordinados diretamente à Geografia.

A nossa dúvida é com relação à existência de possíveis relações hierárquicas de subordinação e superordenação e entre quais termos, uma vez que, não sabemos se cada termo pertence ao mesmo nível hierárquico, se são subordinados entre si e quais níveis pertencem, ou mesmo, se são sinônimos. Nesse caso, decidimos consultar o especialista da área de Geografia da Unesp no Campus de Presidente Prudente por intermédio do catalogador.

O Professor do Campus de Presidente Prudente explicou que há diferentes interpretações e que o uso de um termo ou de outro vai depender da abordagem. Conforme o referido professor, uma das linhas teóricas de pesquisa divide a Geografia entre física e humana. Subordinada a Geografia Física terá a Geografia Agrícola que estuda os meios de produção no campo, aproveitamento do solo para plantação e manejo de animais. A Geografia humana vai estudar, dentre outras coisas, a Geografia Rural que aborda as relações entre o campo e a cidade.

Entretanto, as linguagens VocaUSP e a LCSH não possuem essa estrutura com a divisão entre Geografia Física e Geografia Humana. A LCSH apresenta os dois termos, Geografia Agrícola e Geografia Rural, no primeiro nível subordinados diretamente à Geografia. O VocaUSP não tem nenhum desses dois termos, por outro lado, tem o termo Geografia agrária no primeiro nível e nenhuma subdivisão abaixo desse. No VocaUSP,

o estudo do aproveitamento do solo e dos animais (economicamente) está na área de Agronomia.

Apesar da explicação do Professor do Campus de Presidente Prudente sobre Geografia humana e física fazer mais sentido atualmente, essa não é a abordagem demonstrada pelas estruturas hierárquicas do VocaUSP e da LCSH, consideradas as linguagens fontes principais para a elaboração das estruturas hierárquicas de áreas de conhecimento para o Tesouro Unesp.

Decidimos pela incorporação dos três descritores no primeiro nível hierárquico de Geografia, contudo sem inclusão de outros termos subordinados:

GEOGRAFIA

Geografia Agrária
Geografia rural / Rural geography
Geografia agrícola / Agricultural geography

B4) Caso da área de Ciências da Saúde com uso da estrutura hierárquica das linguagens de indexação MeSH e DeCS:

Para as áreas relacionadas a Ciências da Saúde foram adotadas como quartas linguagens de combinação de estruturas hierárquicas, o Medical Subject Headings (MeSH) e o Descritores em Ciências da Saúde (DeCS), para fazer face à especificidade de áreas como Medicina que, tal como a Engenharia, é composta de várias subáreas. Portanto, na Odontologia, o DeCS funcionou como quarta linguagem de indexação para decisões de combinações importantes nas 67 subáreas, dentre estas, 30 foram determinadas pelo DeCS. A estrutura hierárquica da Medicina, ainda em desenvolvimento, utiliza o MeSH como quarta linguagem de combinação que como o DeCS tem grande influência na determinação de estruturas hierárquicas.

B5) Caso de hierarquias sem combinação ou correlação:

Com relação aos nomes de hierarquias sem combinação ou sem correlação nas linguagens de indexação VocaUSP e LCSH, como por

exemplo Indústria de equipamentos e Companhia de cruzeiros marítimos (Quadro 4), dentro da área Turismo, foi necessário realizar uma pesquisa desses nomes em outras fontes tais como Wikipedia, glossários, tesouros e verificar se podem ser encaixados em níveis já existentes. Outros termos como, por exemplo car renting business pode, também, ser colocado em outra estrutura hierárquica. Esse termo é usado pelo VocaUSP e está na categoria de Turismo e subcategoria de Indústria do Turismo, mas, na LCSH o termo preferido é Leasing and rental services que está na categoria Service Industries. A escolha foi pelo termo car renting business presente no VocaUSP que é mais adequado à área de conhecimento de Turismo na Unesp.

Quadro 4 - Exemplo da estrutura hierárquica de Turismo

TURISMO			
	INDÚSTRIA DO TURISMO	EMPRESAS PRESTADORAS DE SERVIÇOS TURÍSTICOS	EMPRESAS LOCADORAS DE EQUIPAMENTOS TURÍSTICOS
			EMPRESAS LOCADORAS DE VEÍCULOS
			EMPRESAS ORGANIZADORAS DE EVENTOS
			SERVIÇOS DE ENTRETENIMENTO
			SERVIÇOS DE INFORMAÇÕES TURÍSTICAS
			Outfitting industry / Indústria de equipamentos(incluído por proximidade)
			Cruise lines / Companhias de cruzeiros marítimos(incluído por proximidade)

Fonte: Elaborado pelos autores.

Definiu-se ainda que os nomes de categorias e subcategorias sem correlação poderão ser avaliados quanto à proximidade de significado dos demais nomes de categorias e subcategorias existentes como, por exemplo, na área de Turismo, os termos Garden Tours, Dark Tourism, Lake Tourism que existem no primeiro nível da LCSH devem ser incluídos no nível “Tipos de Turismo” do VocaUSP. Porém, quanto a nomes tais como “Indian Tourism” que são muito específicos, ou nomes compostos que ligam dois ou mais assuntos, tais como “Architecture and tourism” não devem ser incluídos na taxonomia, mas devem constar de uma lista separada de “Nomes não utilizados” (Quadro 5). Outros nomes compostos que incluem qualificativos, tais como, “Woman in tourism”, “Woman in Physics”, também, não devem ser incluídos.

Quadro 5 - Exemplo de lista com nomes não utilizados na área de Turismo

Termos não utilizados
Architecture and tourism / Arquitetura e turismo
Culture and tourism / Cultura e turismo
Holocaust memorial tours
Music and tourism / Música e turismo
Sports and tourism / Esportes e turismo
Women in tourism
Indian tourism
Customs administration and tourism

Fonte: Elaborado pelos autores.

B6) Caso de termos poli-hierárquicos:

Durante a combinação das linguagens de indexação para elaboração das estruturas hierárquicas foi observada a ocorrência de termos poli-hierárquicos, ou seja, termos subordinados a mais de um termo genérico, como por exemplo, o termo Courts e sua tradução, “Tribunais”, está subordinado à quatro hierarquias na estrutura da LCSH e da TBN (Quadro 6), respectivamente:

Quadro 6 - Subordinação poli-hierárquica

Courts (LCSH)	Tribunais (TBN)	Tribunais (VocaUSP)
BT Dispute resolution (Law) Judicial districts Law Procedure (Law)	TG Comarcas TG Direito TG Direito Processual TG Resolução de disputa (Direito)	TG Direito Processual

Fonte: Elaborado pelos autores.

O caso de poli-hierarquias revelou que a linguagem de indexação VocaUSP não tem subordinação poli-hierárquica. A construção das estruturas hierárquicas foi decidida com a combinação das estruturas das 3 linguagens iniciando-se pelo VocaUSP, o que não impede a aceitação de poli-hierarquias da LCSH e da TBN. Assim, o termo (Tribunais/ Courts), subordinado a outros 4 termos genéricos na LCSH e na TBN, foi incluído como subordinado nas quatro hierarquias. O exemplo do Quadro 7 demonstra a decisão de aceitar a poli-hierarquia como proposta de diminuição de restrições semânticas para aumento do campo semântico mais diversificado e hospitaleiro.

Quadro 7 - Termo Tribunais subordinado a vários termos

Direito			
	Tribunais		
	Direito Processual		
		Tribunais	
	Organização judiciária		
		Tribunais	
		Comarcas	
			Tribunais

Fonte: Elaborado pelos autores.

B7) Caso de níveis hierárquicos em Biociências

No estudo da hierarquia de Biociências incluíram-se os primeiros níveis: Botânica, Biologia, Imunologia, Microbiologia e Zoologia. No nível de Biologia incluíram-se, dentre outros, os segundos níveis hierárquicos Anatomia e Fisiologia por serem áreas básicas de áreas aplicadas como por exemplo, Medicina e Medicina Veterinária. Este exemplo de uma grande área básica como Biociências demonstra a necessidade de pesquisa terminológica para assegurar a composição de toda a estrutura hierárquica.

Dessa forma, é preciso definir os termos que denominam os níveis hierárquicos para identificar a composição completa da estrutura, além de verificar a estrutura departamental que sustentam as grandes áreas básicas. Por exemplo, no Instituto de Biociências de Botucatu da Unesp estão os departamentos de Anatomia, Botânica, Fisiologia, Genética, Microbiologia e Imunologia, Morfologia, Parasitologia e Zoologia, dentre outros. São parâmetros importantes, o primeiro terminológico e o segundo organizacional da cultura acadêmica que complementam a fusão das estruturas hierárquicas das linguagens de indexação. Nesse sentido, recomenda-se a investigação terminológica do termo que denomina a área e a verificação da estrutura funcional de departamentos que denominam as áreas de níveis hierárquicos para uma decisão estrutural com base na cultura acadêmica.

B8) Caso da área de História do Brasil:

Na comparação entre as linguagens LCSH, TBN e VocaUSP para a elaboração da estrutura hierárquica da área de História, observou-se que o primeiro nível denominado “História do Brasil” não existe na LCSH e nem na TBN. Os termos do segundo e terceiros níveis foram localizados nessas duas linguagens como subdivisão de local geográfico e não como termo tópico. Os termos apresentam-se como cabeçalhos de assunto no exemplo a seguir como se encontra no VocaUSP/TBN/LCSH: “INDEPENDÊNCIA DO BRASIL / Brazil - History - Declaration of Independence, 1822 / Brasil - História - Independência, 1822”. O exemplo apresenta a necessidade de adotarmos apenas o termo tópico dos cabeçalhos das linguagens TBN e VocaUSP tal como foi adotado pelo VocaUSP. Portanto, o comparativo sem as subdivisões ficará da seguinte forma:

“INDEPENDÊNCIA DO BRASIL / Declaration of Independence, 1822 / Independência, 1822”.

Entretanto, nesse processo existem cabeçalhos que, se modificados os termos tópicos, serão alterados, como por exemplo: “Descobrimiento da América/América - descobertas e explorações/América - discovery and exploration”. O termo tópico desse cabeçalho na TBN e LCSH, “América - descobertas e explorações/América - discovery and exploration”, é o termo “descobertas e explorações/discovery and exploration” que seria muito alterado e sem o local geográfico não teria significado junto à área de História. Nesse caso, é preciso manter da forma como se apresentou inicialmente.

A maioria dos cabeçalhos e termos da área de História apresentam datas referentes à períodos históricos, como por exemplo: **VICE-REINADO (1720-1822) / Brasil - História - Vice-reinado, 1762-1808; REGÊNCIA DE D. JOÃO (1808-1816) / Brasil - História - João VI, 1808-1821**. Nestes dois cabeçalhos existem diferenças entre períodos históricos que precisarão ser verificados sob critério de garantia cultural em consulta à especialista dessa área na Unesp - Campus de Assis.

B9) Outros casos: poderão ser relatados e enviados para linguagemUnesp@listas.Unesp.br

5.2.2.3.2 FASE DE CONSOLIDAÇÃO DA ESTRUTURA HIERÁRQUICA DO TESAURO UNESP.

Quando finalizada a compatibilização de cada área, é necessário definir a escolha dos nomes dos níveis hierárquicos para consolidação da estrutura hierárquica do tesauro Unesp. Por isso, a denominação selecionada do Tesauro Unesp tem que estar à frente das demais em pink. Em seguida é preciso copiar a estrutura hierárquica em outra aba denominada “Tesauro Unesp” e apagar as denominações das linguagens de indexação VocaUSP, LCSH, TBN e, no caso de áreas de ciências da saúde, MeSH. As denominações em pink deverão ser mantidas para a consolidação da estrutura hierárquica do tesauro Unesp.

A decisão para quando uma área ou subárea não tem a combinação das duas ou três linguagens é de que prevaleça a estrutura hierárquica da linguagem em que existe a área, acompanhada de pesquisas em glossários, Wikipédia e consulta à especialistas.

O Tesouro Unesp pode decidir por suas próprias denominações na estrutura hierárquica, quando for o caso, determinado por uma questão cultural ou com base nas demais linguagens quando for diferente no VocaUSP. Por exemplo: a categoria denominada “História Geral, História do Brasil e Geografia” no VocaUSP não é representativa porque a área de “Geografia” está representada em categoria independente. Portanto, a categoria foi denominada “História” com base, também, na denominação da LC, History, e na denominação da TBN, “História”.

Ao final do processo de formação das estruturas hierárquicas de áreas de conhecimento mediante compatibilização das linguagens, obtivemos um conjunto de áreas de conhecimento, conforme Quadro 8:

Quadro 8 - Divisão por áreas do conhecimento após desenvolvimento das estruturas hierárquicas

Grandes áreas de conhecimento (7) (CNPq/CAPEs)	Áreas de conhecimento da taxonomia (64)
Engenharias – ENG (13)	Engenharia Aeronáutica e Aeroespacial, Engenharia Civil, Engenharia de Biotecnologia, Engenharia de Materiais, Engenharia de Minas e Petróleo, Engenharia de Produção, Engenharia de Telecomunicações, Engenharia Elétrica, Engenharia Industrial, Engenharia Mecânica, Engenharia Metalúrgica, Engenharia Naval e Oceânica, Engenharia Química.
Ciências Humanas – CHU (11)	Artes, Ciências Sociais, Educação Musical e Música, Filosofia, Geografia, História, Línguas e Literatura, Linguística, Educação, Psicologia, Relações Internacionais.
Ciências Sociais Aplicadas – CSA (14)	Administração, Arquitetura, Planejamento Urbano, Arquivologia, Biblioteconomia, Economia, Design, Direito, Jornalismo, Museologia, Relações Públicas, Serviço Social, Turismo, Comunicações.
Ciências Biológicas – CBA (2)	Biociências, Engenharia Ambiental

Ciências Exatas e da Terra – CET (7)	Ciência da Computação, Estatística, Física, Geologia, Matemática, Meteorologia, Química.
Ciências Agrárias – CAG (7)	Agronomia, Engenharia Agrônômica, Engenharia de Alimentos, Engenharia de Pesca, Engenharia Florestal, Medicina Veterinária, Zootecnia.
Ciências da Saúde – CSU (10)	Ciências Biomédicas, Educação Física, Enfermagem, Farmácia, Fisioterapia, Fonoaudiologia, Medicina, Nutrição, Odontologia, Terapia Ocupacional.

Fonte: Elaborado pelos autores.

O resultado apresentado no Quadro 8 revela algumas mudanças necessárias em função do processo de compatibilização das linguagens de indexação. Observamos durante a construção de estruturas hierárquicas que áreas de conhecimento (Vide Quadro 1) como “Pedagogia” existe com o nome de “Educação” ou, no caso da área de “Teatro e Artes Cênicas” pertence à área de conhecimento mais ampla de “Artes”, portanto, no Quadro 8 as denominações das áreas de “Pedagogia” e “Teatro e Artes Cênicas” aparece com suas respectivas substituições de denominações. O primeiro caso, de mudança de nomenclatura, aconteceu também com “Ciências Econômicas” para “Economia”. O segundo caso, da área ser subordinada à uma área genérica superior, foi o mais frequente e o mesmo aconteceu com as denominações das áreas de conhecimento de “Administração Pública” para “Administração”, “Radialismo” para “Comunicações”, “Sistemas de Informação” para “Ciência da Computação”, “Física Médica” para “Física”, “Farmácia Bioquímica” para “Farmácia”, “Engenharia Cartográfica” para “Geografia” e etc.

A criação de novas áreas não existentes aconteceu com “Línguas e Literatura” e “Linguística” que substituíram “Letras”, “Planejamento Urbano” que substituiu “Arquitetura e Urbanismo”, “Arquitetura” que substituiu “Arquitetura e Urbanismo”, “Comunicações” que incluiu “Radialismo”, “Biotecnologias” que incluiu “Ciências Biológicas” e “Ecologia”, “Agronomia” que não existia na grande área de “Ciências Agrárias”, “Engenharia de Biotecnologia” que incluiu e substituiu “Engenharia Biotecnológica” e “Engenharia de Bioprocessos e Biotecnologia”, “Engenharia Industrial” que substituiu “Engenharia Industrial Madeireira”.

Por outro lado, a grande área de Engenharia precisou apresentar mais áreas subordinadas, porém, muito especializadas, inovadoras e representativas tais como: Engenharia de Minas e Petróleo, Engenharia Industrial, Engenharia Metalúrgica e Engenharia Naval e Oceânica. A área de “Engenharia de Energia” foi eliminada por falta de nomenclatura nas estruturas hierárquicas.

Foram excluídas do Quadro 8 as seguintes denominações de áreas: “Engenharia de Controle e Automação”, “Artes Visuais”, “Teatro e Artes Cênicas”, “Letras”, “Pedagogia”, “Administração Pública”, “Arquitetura e Urbanismo”, “Ciências Econômicas”, “Radialismo”, “Ciências Biológicas”, “Ecologia”, “Engenharia Biotecnológica”, “Engenharia de Bioprocessos e Biotecnologia”, “Engenharia Cartográfica”, “Sistemas de Informação”, “Engenharia Industrial Madeireira”, “Farmácia Bioquímica”, “Física Médica” e “Engenharia de Energia”.

Entretanto, a quantidade de áreas de conhecimento no Quadro 8 (65 áreas) está ligeiramente maior que no Quadro 1 (60 áreas) que tomou como parâmetro os cursos de graduação. Ocorre que a exclusão se refere à substituição de denominação ou inclusão em área mais ampla, portanto, não houve grande mudança quantitativa. As áreas incluídas sem que houvesse uma substituição foram as cinco áreas de “Comunicação”, “Agronomia”, “Engenharia de Minas e Petróleo”, “Engenharia Metalúrgica” e “Engenharia Naval e Oceânica”. por isso a diferença.

5.3 APLICAÇÃO DAS ESTRUTURAS HIERÁRQUICAS DE ÁREAS DE CONHECIMENTO PARA A CLASSIFICAÇÃO DOS TERMOS DO TESAURO

A aplicação das estruturas hierárquicas se realizou pelo software TemaTres utilizado para a exportação dos registros de autoridade em formato de intercâmbio interoperável MARC21 e que atualmente apresenta o Tesouro Unesp para consulta junto às interfaces de busca dos sistemas de recuperação em bases de dados da Unesp, quais sejam, o catálogo online das bibliotecas universitárias, biblioteca digital da Unesp e Repositório Institucional da Unesp.

Para aplicação da estrutura hierárquica no software TemaTres foram utilizadas as sete categorias que representam as grandes áreas de

conhecimento conforme Quadro 2. Essas categorias podem ser vistas no Quadro 8 com os seguintes códigos que as precedem:

Quadro 9 - Código e nome da categoria de áreas do conhecimento

CÓDIGO	NOME DA CATEGORIA
ENG	ENGENHARIA
CHU	CIÊNCIAS HUMANAS
CSA	CIÊNCIAS SOCIAIS APLICADAS
CBA	CIÊNCIAS BIOLÓGICAS
CET	CIÊNCIAS EXATAS E DA TERRA
CAG	CIÊNCIAS AGRÁRIAS
CSU	CIÊNCIAS DA SAÚDE

Fonte: Elaboração dos autores.

Cada uma das categorias contém suas áreas de conhecimento relacionadas às 7 grandes áreas de conhecimento. Dessa forma, o nome das áreas de conhecimento ficou precedido do código da categoria, conforme exemplo da categoria de “Engenharias”:

1. Engenharias

ENG - Engenharia Aeronáutica e Aeroespacial

ENG - Engenharia Civil

ENG – Engenharia de Biotecnologia

ENG - Engenharia de Materiais

ENG – Engenharia de Minas e Petróleo

ENG - Engenharia de Produção

ENG - Engenharia de Telecomunicações

ENG - Engenharia Elétrica

ENG – Engenharia Industrial

ENG - Engenharia Mecânica

ENG – Engenharia Metalúrgica

ENG – Engenharia Naval e Oceânica

ENG - Engenharia Química

Os metatermos são precedidos pelos códigos, conforme Quadro 10, e sua função é a de classificar os descritores e não descritores do Tesouro Unesp nas áreas de conhecimento.

Quadro 10 - Metatermos: códigos de categorias e subcategorias após desenvolvimento das estruturas hierárquicas

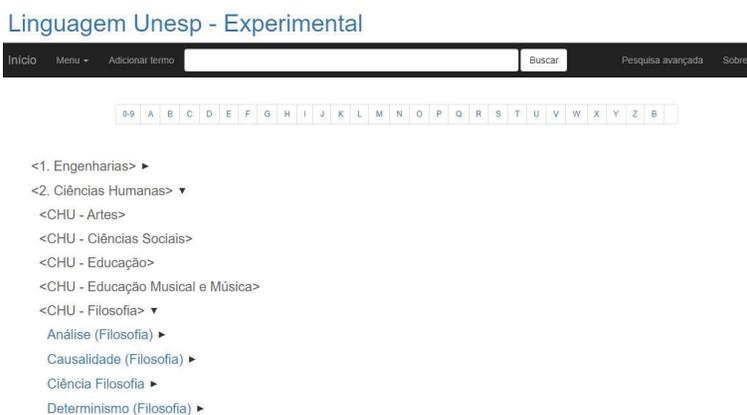
1. Engenharias (13)	2. Ciências Humanas (11)	3. Ciências Sociais aplicadas (14)	4. Ciências Biológicas (2)	5. Ciências Exatas e da Terra (7)
ENG - Engenharia Aeronáutica e Aeroespacial	CHU - Artes	CSA - Administração	CBA - Biociências	CET - Ciências da Computação
ENG - Engenharia Civil	CHU - Ciências Sociais	CSA - Arquitetura	CBA - Engenharia Ambiental	CET - Estatística
ENG – Engenharia de Biotecnologia	CHU - Educação Musical e Música	CSA - Planejamento Urbano		CET - Física
ENG - Engenharia de Materiais	CHU - Filosofia	CSA - Arquivologia		CET - Geologia
ENG – Engenharia de Minas e Petróleo	CHU - Geografia	CSA - Biblioteconomia		CET - Matemática
ENG - Engenharia de Produção	CHU - História	CSA - Economia		CET - Meteorologia
ENG - Engenharia de Telecomunicações	CHU - Línguas e Literatura	CSA - Design		CET - Química
ENG - Engenharia Elétrica	CHU - Linguística	CSA - Direito		
ENG – Engenharia Industrial	CHU - Educação	CSA - Jornalismo		
ENG - Engenharia Mecânica	CHU - Psicologia	CSA - Museologia		

ENG – Engenharia Metalúrgica	CHU - Relações Internacionais	CSA - Relações Públicas		
ENG – Engenharia Naval e Oceânica		CSA - Serviço Social		
ENG - Engenharia Química		CSA - Turismo		
		CSA - Comunicações		

Fonte: Elaboração dos autores.

Essas categorias e subcategorias da Taxonomia de áreas de conhecimento da Unesp foram consideradas no software TemaTres como “metatermos” conforme demonstrado na Figura 1:

Figura 1 – Implementação dos metatermos no TemaTres: representação das grandes áreas e respectivas áreas de conhecimento



Fonte: Interface TemaTres.

Considerando-se que os registros de autoridades de assunto dos descritores do Tesauro Unesp precisarão, necessariamente, ser classificados nos metatermos das áreas de conhecimento, denominaremos essa operação de classificação como Retrospectiva. A classificação dos descritores foi realizada em duas etapas: a primeira consistiu da classificação automática

retrospectiva composta da aplicação de seis métodos e a segunda consistirá da classificação intelectual dos demais descritores que não puderam ser classificados automaticamente. Após a operação retrospectiva poderá ser realizada a classificação intelectual dos descritores em metatermos durante a elaboração do registro de autoridade de termos novo bem como durante a adequação de registros de autoridades existentes de acordo com recomendações finais (item 5.4).

5.3.1 CLASSIFICAÇÃO AUTOMÁTICA DOS DESCRITORES E NÃO DESCRITORES RETROSPECTIVOS EM CATEGORIAS E SUBCATEGORIAS (METATERMOS)

A operação de classificação automática dos descritores nos metatermos do software TemaTres foi realizada em instalação experimental do Tesouro Unesp onde realizaram-se diferentes procedimentos de compatibilização de descritores que denominamos de tentativas para classificação automática. Nesse intuito realizaram-se 5 tentativas de compatibilização de descritores mediante os seguintes procedimentos:

Primeira tentativa: compatibilização de descritores por semelhanças e igualdades entre o conjunto de caracteres dos metatermos e dos termos gerais

Consistiu da compatibilização dos termos das áreas de conhecimento (vide Quadro 8) com as áreas da Taxonomia, subindo para a instalação experimental de testes do Tematres todos os termos da linguagem disponibilizados até o momento, utilizando os termos mais gerais da hierarquia (cerca de 3800 termos) conseguindo vincular aproximadamente 300 termos, cujas vinculações automáticas com os metatermos foram verificadas conforme orientações do Apêndice A.

Segunda tentativa: compatibilização de descritores sem plurais, palavras vazias e sufixos.

Foram executados os seguintes procedimentos de compatibilização: idêntica, idêntica retirando as palavras vazias, idêntica retirando os plurais e idêntica retirando os sufixos. Por questões de codificação dos caracteres, alguns termos não foram compatíveis na primeira tentativa nas quais são idênticas como demonstrado no Quadro 11:

Quadro 11 - Termos compatibilizados

CSU - Educação Física	Educação física
ENG - Engenharia de Produção	Engenharia de produção
ENG - Engenharia Elétrica	Engenharia elétrica
CAG - Engenharia Florestal	Engenharia florestal
CHU - Relações Internacionais	Relações internacionais
CSA - Relações Públicas	Relações publicas
CSA - Serviço Social	Serviço social

Fonte: Elaborado pelos autores.

No processo de compatibilização idêntica sem palavras vazias e sem plurais não tiveram nenhuma correspondência de compatibilização. No processo de compatibilização idêntica sem sufixo, entretanto, foram possíveis as seguintes compatibilizações (Quadro 12):

Quadro 12 - Termos compatibilizados

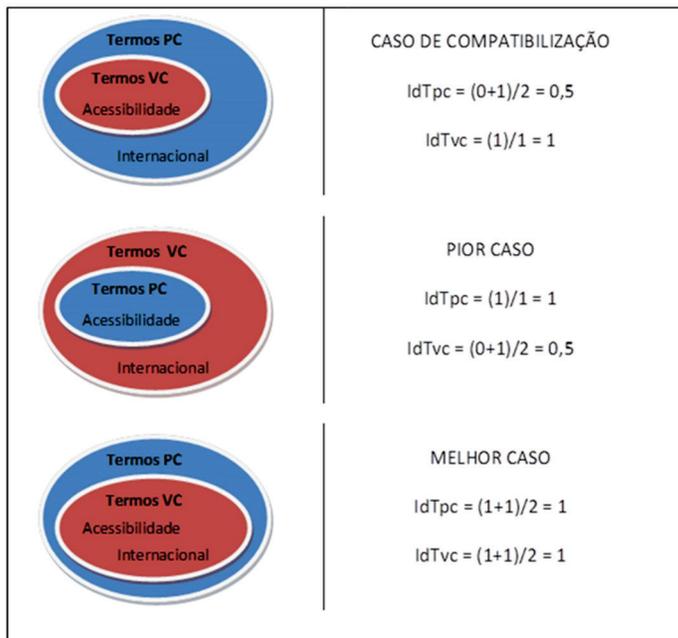
CSU - Enfermagem	Enfermeiros
CHU - Filosofia	Filósofos
CET - Geologia	Geólogos
CET - Química	Químicos
CSA - Comunicações	Radiação

Fonte: Elaborado pelos autores.

Terceira tentativa: compatibilização de descritores pela aplicação da teoria dos conjuntos por meio de índice para verificar quais termos estão contidos

Nesta etapa, os processos de compatibilização estão centrados no cálculo de índice contido, a partir da teoria de conjuntos (SANTOS, J., 2015). Na Figura 2 está demonstrado um exemplo de cálculo, conforme aplicado por Santos, J. (2015) em compatibilização de palavras-chave. Para efeitos de aplicação nesta compatibilização as correspondências de <IdPC> x <DGs> e <IdVC> x <metatermo>. No estudo do autor supracitado o <IdPC> que é o cálculo do índice contido da palavra-chave no termo do vocabulário controlado, nesta aplicação é considerado o cálculo do índice contido do descritor geral com o temo do metatermo. Logo no <IdVC> é o cálculo do índice contido do termo do vocabulário controlado na palavra-chave e nesta aplicação será o índice contido calculado a partir do metatermo com o descritor geral.

Figura 2 – Exemplo do cálculo do índice contido



Fonte: Santos, J. 2015, p.99

Foram calculados os índices para termos em seu formato original, posteriormente com a retirada de palavras vazias, retirada de plurais e sufixo.

No quadro 13 foram identificados os Metatermos da taxonomia e os termos gerais candidatos à compatibilização. São identificados na coluna Modo de compatibilização as variações de compatibilização: <índice_identica> termos originais sem alteração, <índice_pv> termos sem as palavras vazias, <índice_plural> termos sem os plurais e <índice_suf> termos sem o sufixo. Observa-se que em ambos os termos, metatermo e termo geral, são realizados os processos de redução para realizar a compatibilização. Totalizando compatibilização de termos na seguinte ordem: 26 por termos originais, 39 sem palavras vazias, 41 sem plurais e 55 sem sufixos. Nos casos de compatibilização pelos termos originais eles também serão compatíveis pelos demais processos, razão pela qual existe a repetição dos termos nas primeira e segunda colunas.

Quadro 13 – Descritores das subcategorias de Física e Matemática compatibilizados com aplicação da teoria dos conjuntos

Metatermo (Área do conhecimento)	Termos Gerais (TG)	Modo de Compatibilização
CET - Física	Metalurgia física	índice_identica, índice_pv, índice_plural, índice_suf
CET - Física	Óptica física	índice_identica, índice_pv, índice_plural, índice_suf
CET - Matemática	Análise matemática	índice_identica, índice_pv, índice_plural, índice_suf
CET - Matemática	Capacidade matemática	índice_identica, índice_pv, índice_plural, índice_suf
CET - Matemática	Engenharia matemática	índice_identica, índice_pv, índice_plural, índice_suf

Fonte: Elaborado pelos autores.

Quarta tentativa: compatibilização de descritores por meio dos termos específicos com os metatermos

Foi realizada a coleta de todos os descritores específicos a partir dos termos gerais, por meio de localização do identificador de cada termo geral e consulta por meio da API do TemaTres, neste caso a requisição <suggestDetails>, e para a consulta dos termos específicos a requisição <fetchDown>. Posteriormente, efetuados os procedimentos de compatibilização dos termos específicos com os metatermos, o resultado obtido (Quadro 14) foi uma compatibilização de maneira idêntica, duas retirando os plurais e quatro retirando os sufixos.

Quadro 14 - Termos compatibilizados por meio dos termos específicos

Metatermo (Área do conhecimento)	Termos Gerais (TG)	Modo de Compatibilização
CET - Matemática	Cientistas	IDENTICO_TE_SUF
CET - Física	Físicos	IDENTICO_TE_PLURAL
CSU - Nutrição	Fisiologistas	IDENTICO_TE_SUF
CET - Estatística	Mecânica analítica	IDENTICO_TE
CET - Física	Mulheres cientistas físicas	IDENTICO_TE_PLURAL
CBA – Engenharia Ambiental	Naturalistas	IDENTICO_TE_SUF
CHU – Língua e Literatura	Poesia visual	IDENTICO_TE_SUF

Fonte: Elaborado pelos autores.

Quinta tentativa: compatibilização de descritores por meio dos termos gerais das estruturas hierárquicas das áreas de conhecimento (Taxonomia)

Para formar o campo semântico das áreas de conhecimentos presente nas subáreas dos metatermos serão considerados os termos presentes em cada área de conhecimento da Taxonomia de Áreas do Conhecimento da Unesp que colocados em lista alfabética serão comparados com as listas alfabéticas dos descritores gerais do Tesouro Unesp. Será verificado se os compatíveis podem ser incluídos nos metatermos correspondentes.

A compatibilização a partir da estrutura hierárquica em seus níveis 1 e 2 da área de conhecimento <CSA - Arquitetura e Urbanismo> constaram de 63 termos do nível 1 e 174 termos do nível 2, totalizando 237. A compatibilização foi executada por meio da comparação dos conjuntos de caracteres do termo geral com o termo da estrutura hierárquica que resultaram em 11 compatibilizações (Quadro 15), sendo uma compatibilização com a comparação dos termos sem as palavras vazias (Quadro 16) e três compatibilizações sem os plurais (Quadro 17). O total de termos da Taxonomia de Áreas do Conhecimento da Unesp são 16.477.

Quadro 15 - Descritores compatibilizados a partir da comparação com a estrutura hierárquica de área do conhecimento “Arquitetura”, níveis 1 e 2

Metatermo (Área do conhecimento)	Termos Gerais (TG)	Modo de Compatibilização
CSA - Arquitetura	Acústica arquitetônica	IDENTICO_EH
CSA - Arquitetura	Arquitetura colonial	IDENTICO_EH
CSA - Arquitetura	Arquitetura moderna	IDENTICO_EH
CSA - Arquitetura	Arquitetura paisagística	IDENTICO_EH
CSA - Arquitetura	Casas de correção	IDENTICO_EH
CSA - Arquitetura	Desenho arquitetônico	IDENTICO_EH
CSA - Arquitetura	Edifícios	IDENTICO_EH
CSA - Arquitetura	Edifícios industriais	IDENTICO_EH
CSA - Arquitetura	Edifícios públicos	IDENTICO_EH
CSA - Arquitetura	Monumentos funerários	IDENTICO_EH
CSA - Arquitetura	Túmulos	IDENTICO_EH

Fonte: Elaborado pelos autores.

Quadro 16 - Descritores compatibilizados a partir da comparação sem palavras vazias com a estrutura hierárquica nível 1 e 2

Metatermo (Área do conhecimento)	Termos Gerais (TG)	Modo de Compatibilização
CSA - Arquitetura	Escadas e escadarias	IDENTICO_EH_PV

Fonte: Elaborado pelos autores.

Quadro 17 - Descritores compatibilizados a partir da comparação sem plurais com a estrutura hierárquica nível 1 e 2

Metatermo (Área do conhecimento)	Termos Gerais (TG)	Modo de Compatibilização
CSA - Arquitetura	Estruturas hidráulicas	IDENTICO_EH_Plural
CSA - Arquitetura	Monumentos	IDENTICO_EH_Plural
CSA - Arquitetura	Pórticos estruturais	IDENTICO_EH_Plural

Fonte: Elaborado pelos autores.

A compatibilização estendeu-se para as demais áreas do conhecimento, resultando em 1249 compatibilizações a partir do campo semântico instanciado dos termos da Taxonomia de Áreas do Conhecimento da Unesp.

Sexta tentativa: compatibilização de descritores específicos com os termos das estruturas hierárquicas das áreas de conhecimento (Taxonomia)

Após a quinta tentativa, a compatibilização dos descritores gerais com os termos da Taxonomia de Áreas do Conhecimento da Unesp foi executada a sexta tentativa com o objetivo de compatibilizar maior quantidade de descritores para realizar a classificação nos metatermos. Nesse processo considerou-se os descritores específicos do Tesouro Unesp

para comparar com os termos da Taxonomia de Áreas do Conhecimento da Unesp e identificar a classificação dos descritores gerais nos metatermos, resultando em 1085 compatibilizações por meio da igualdade dos termos. No Quadro 18 apresenta-se um extrato das compatibilizações, nesse extrato estão as compatibilizações da área do conhecimento Ciência da Computação.

Quadro 18 – Extrato dos descritores compatibilizados a partir da comparação dos descritores específicos com a Taxonomia

Metatermo	TG	TE	HE
CET - Ciência da Computação	Sistemas autoorganizadores	Inteligência artificial	Inteligência Artificial
CET - Ciência da Computação	Computadores	Processamento de listas (Computadores)	Processamento De Listas (Computadores)
CET - Ciência da Computação	Prática de escritório Automação	Processamento de palavras	Processamento De Palavras
CET - Ciência da Computação	Computadores digitais Confiabilidade	Tolerância a falha (Computadores)	Tolerância A Falha (Computadores)
CET - Ciência da Computação	Algoritmos	Algoritmos genéticos	Algoritmos Genéticos
CET - Ciência da Computação	Linguagens de programação funcional	COMMON LISP (Linguagem de programação de computador)	Common Lisp (Linguagem De Programação De Computador)
CET - Ciência da Computação	Computadores digitais Programação	Compressão de dados (Computação)	Compressão de dados (Computação)
CET - Ciência da Computação	Quadros de aviso eletrônico	Grupos de bate-papo pela Internet	Grupos de bate-papo pela Internet
CET - Ciência da Computação	Cabeça	Indexação automática	Indexação automática
CET - Ciência da Computação	Integrated software	Integração semântica (Sistemas de computação)	Integração semântica (Sistemas de computação)
CET - Ciência da Computação	Negócios Processamento de dados	Intercâmbio eletrônico de dados	Intercâmbio eletrônico de dados

CET - Ciência da Computação	Computadores digitais Programação	Multiprogramação (Computadores)	Multiprogramação (Computadores)
CET - Ciência da Computação	Computadores	Processamento de listas (Computadores)	Processamento de listas (Computadores)
CET - Ciência da Computação	Automação	Sistemas de coleta automática de dados	Sistemas de coleta automática de dados
CET - Ciência da Computação	Solução de problemas	Sistemas de consultas e respostas	Sistemas de consultas e respostas
CET - Ciência da Computação	Automato celular	Inteligência coletiva	Inteligência Coletiva
CET - Ciência da Computação	System theory	Sistemas híbridos	Sistemas Híbridos

Fonte: Elaborado pelos autores.

O resultado das seis tentativas com diferentes métodos permitiu compatibilizar 2658 descritores do total de 3892 e inclui-los nos metatermos.

Quadro 19 – Resultados das tentativas de compatibilizações

Tentativas	Quantidades de Compatibilizações	% do Total
Primeira	300	7,71
Segunda	12	0,31
Terceira	5	0,13
Quarta	7	0,18
Quinta	1249	32,09
Sexta	1085	27,88
Total de compatibilizações	2658	68,29
Total de descritores	3892	100

Fonte: Elaborado pelos autores.

Os demais descritores não compatibilizados e classificados deverão ser submetidos à classificação intelectual cuja orientação está no próximo item.

5.3.2 CLASSIFICAÇÃO INTELCTUAL DOS DESCRITORES E NÃO DESCRITORES RETROSPECTIVOS EM CATEGORIAS E SUBCATEGORIAS (METATERMOS)

Os demais termos não compatilizados automaticamente terão que, necessariamente, ser vinculados manualmente aos metatermos seguindo os mesmos procedimentos do Apêndice D abaixo indicados:

- entrar no termo a ser vinculado;
- clicar em “opções” e subordinar termo;
- realizar a busca do meta-termo a ser vinculado;
- selecionar o metatermo e adicionar o vínculo.
- Identificaram-se, porém, vínculos de termos inconsistentes com a estrutura hierárquica dos metatermos. Para modificar a compatibilização serão adotados os seguintes procedimentos:
- entrar no termo a ser vinculado;
- excluir o vínculo do termo geral indevido conforme estrutura hierárquica do tesauro da Unesp;
- clicar em “opções” e subordinar termo;
- realizar a busca do metatermo a ser vinculado;
- selecionar o metatermo e adicionar o vínculo;
- modificar o vínculo no registro de autoridade.

Em casos específicos, verificar orientações abaixo:

Caso 1: Termo vinculado a um metatermo que não corresponde à sua área de assunto de origem

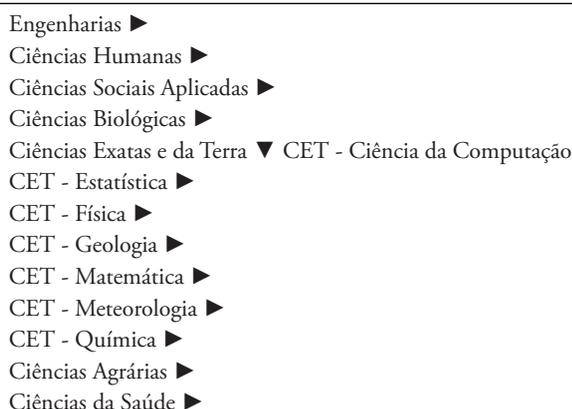
ARQUITETURA DE COMPUTADOR era um termo vinculado ao termo genérico (TG) ARQUITETURA, com a eliminação do TG de outra área é preciso que seja feito o estudo para inserção do termo superior. Para isso, o primeiro passo é a consulta à taxonomia hierárquica do Tesauro Unesp da área correspondente. Após essa verificação, o termo ARQUITETURA DE COMPUTADOR está vinculado à área de Ciência da Computação.

Visto isso, primeiro fazer a classificação do termo ARQUITETURA DE COMPUTADOR no metatermo de Ciência da Computação e depois abrir o registro de autoridade e fazer a compatibilização conforme orientações anteriormente definidas, ou seja, consulta à Terminologia da BN e LCSH, e incluir como termo genérico (TG) a mesma denominação do metatermo Ciência da Computação.

Caso 2: termo vinculado com um termo genérico sem correspondência com a área de origem e sem vínculo com metatermo:

CIÊNCIA DA COMPUTAÇÃO tem como termo genérico (TG) no registro de autoridade o termo CIÊNCIA que é um termo maior que o metatermo CIÊNCIAS EXATAS E DA TERRA e de sua subordinada área de conhecimento CIÊNCIA DA COMPUTAÇÃO, como se vê na representação abaixo. Nesse caso, não é possível mudar as relações hierárquicas estabelecidas no registro de autoridade do termo CIÊNCIA DA COMPUTAÇÃO. A relação hierárquica com o termo genérico CIÊNCIA deve ser mantida. Mas, o termo CIÊNCIA DA COMPUTAÇÃO será vinculado ao metatermo CIÊNCIA DA COMPUTAÇÃO como se observa na Figura 3:

Figura 3 - METATERMOS: Hierarquias de áreas de conhecimento



Engenharias ►
Ciências Humanas ►
Ciências Sociais Aplicadas ►
Ciências Biológicas ►
Ciências Exatas e da Terra ▼ CET - Ciência da Computação
CET - Estatística ►
CET - Física ►
CET - Geologia ►
CET - Matemática ►
CET - Meteorologia ►
CET - Química ►
Ciências Agrárias ►
Ciências da Saúde ►

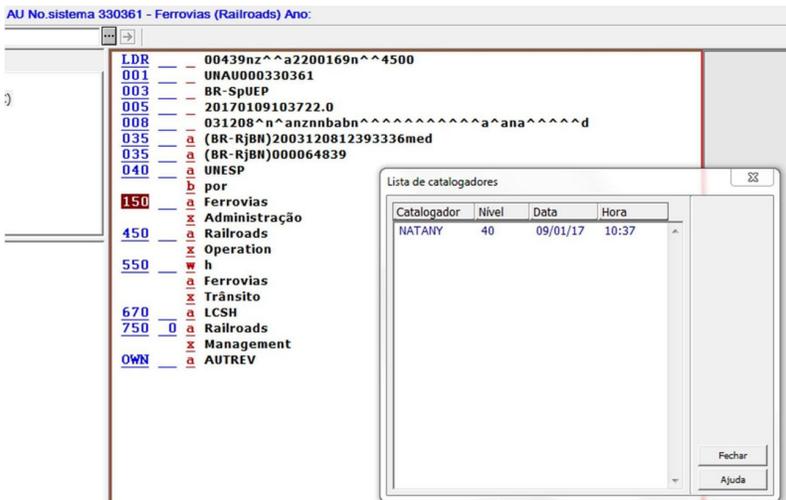
Fonte: Elaborado pelos autores.

Caso 3: Termos órfãos: sem registro de autoridade e sem vínculo com registro bibliográfico

Somente termos tópicos, que constam do campo 150 do registro de autoridade são tratados como descritores com vínculos em registros bibliográficos.

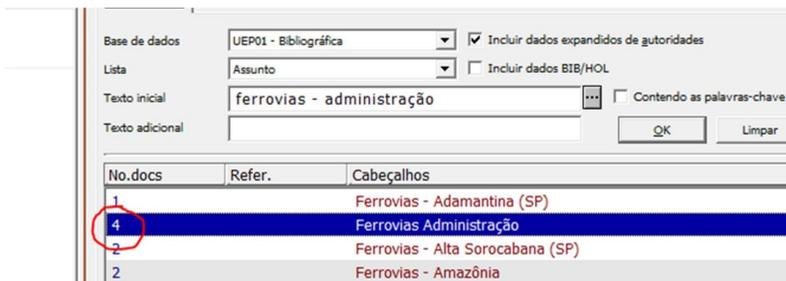
Como exemplo, o termo tópico Ferrovias – Administração foi incorporado conforme figura 4 pela catalogadora e na Figura 5 se constata que 4 registros bibliográficos estão vinculados a esse descritor.

Figura 4 – Registro de autoridade de incorporação do descritor



Fonte: Software Aleph – Módulo de Catalogação

Figura 5 – Vinculação de registros bibliográficos ao descritor



Fonte: Software Aleph – Módulo de Catalogação

Entretanto, quando um termo foi gerado a partir do campo 550 para termo específico (TE) ou relacionado (TR) e não possui um registro de autoridade próprio é porque não possui vínculo com registros bibliográficos. Trata-se de um termo “órfão” cuja orientação solicita eliminação. Descritores sem registro de autoridade não são vinculados aos metatermos e tem tratamento específico.

Como exemplo demonstrativo de termo órfão é possível observar na Figura 6 o termo: motoristas de automóveis.

Figura 6 - Registro de autoridade com termo órfão

UJ No sistema 354666 - Motoristas (Drivers, Motor vehicle) Ano:

LDL	---	00647nz^^a2200241n^^4500
001	---	UNAU000354666
003	---	BR-SpUEP
005	---	20180328081545.0
008	---	090330dn^anznbnbn^^^^^^^^^^^ana^^^^^^d
035	---	a (BR-RJBN)2009033010030178med
035	---	a (BR-RJBN)000094296
040	---	a UNESP
	---	b por
150	---	a Motoristas
450	---	a Drivers, Motor vehicle
450	---	a Motor vehicle operators
450	---	a Motorists
450	---	a Operadores de veículos a motor
450	---	a Operators, Motor vehicle
550	---	w h
	---	a Motociclistas
550	---	w h
	---	a Motoristas de automóveis
550	---	w h
	---	a Motoristas de caminhão
550	---	w h
	---	a Motoristas de ônibus
550	---	w g
	---	a Pessoas
667	---	a Pode ser subdividido geograficamente
670	---	a LCSH
750	---	0 a Motor vehicle drivers
OWN	---	a AUTREV

1. Mensagens | 2. Ajuda do Campo | 3. Ver no OPAC WEB (Ctrl+O) | 4. Registros HOL | 5. Objetos |

Erro validação doc

Campo 550 com texto "\$wh\$aMotoristas de automóveis" não é compatível com cabeçalho GEN.

Fonte: Software Aleph – Módulo de Catalogação

Quando o catalogador importa ou compatibiliza um registro de autoridade, se aparecer a mensagem (destacada em azul no final da Figura 4) de que o campo 550 com texto “\$\$wh\$aatermo” não é compatível com cabeçalho GEN, significa que ele não tem registro de autoridade (RA) próprio e não possui registro bibliográfico ligado a ele pois, o Aleph combina o campo 150 do RA com o campo de assunto do registro bibliográfico e, nesse caso, a mensagem se refere a um campo 550. Essas

orientações são as mesmas para o campo 550 com texto “\$\$\$atermo” (termos relacionados por associação e não hierarquia).

Portanto, recomenda-se não realizar a classificação intelectual do termo que tiver as características acima citadas e, na sequência encaminhá-lo para eliminação.

5.4 RECOMENDAÇÕES PARA USO DA TAXONOMIA DE ÁREAS DE CONHECIMENTO DA UNESP AOS DESENVOLVEDORES DO TESAURO UNESP

As 62 estruturas hierárquicas pertencem à Taxonomia de áreas de conhecimento da Unesp disponíveis no GDrive¹ e no endereço <https://www.biblioteca.Unesp.br/taxonomia>. Na Figura 7, a seguir, apresenta-se o layout da Taxonomia.

Figura 7 – Taxonomia de Áreas do Conhecimento da Unesp



Fonte: Elaborado pelos autores.

A Taxonomia de Áreas de Conhecimento deverá ser consultada para elaboração de registros de autoridade de novos descritores, bem como para adequação de registros de autoridades de descritores existentes na base do catálogo online com a finalidade de inclusão do vínculo com os metatermos. Após consulta às estruturas hierárquicas da Taxonomia e

¹ https://drive.google.com/drive/u/0/folders/0B3SVt_Fdq_SHbkdyWHIMVG8yWlk

verificação à qual área de conhecimento o termo em questão pertence, será possível definir o código do Metatermo conforme Quadro 20. O registro de autoridade deverá ter um campo específico para isso denominado “MET” de Metatermo na seguinte ordem dos campos do formato MARC21:

Quadro 20 - Campo do Metatermo na ordem de campos do registro de autoridade em MARC21

Campo	Conteúdo
150	termo autorizado
450	termo remissiva ver
550	termo remissiva ver também (relação hierárquica ou associativa)
670	Fonte positiva de dados
680	Nota de escopo/histórica
750	Ligação de cabeçalhos autorizados (Termo da LC)
MET	Metatermo (Exemplo: ENG - Engenharia Aeronáutica)
OWN	indicativo de atualização por catalogador do Grupo de Linguagem da Rede de Bibliotecas da Unesp

Fonte: Elaboração dos autores.

Como exemplo a ser fornecido, suponhamos que o registro de autoridade seja do termo “Engenharia Mecânica”, adotado a partir da Figura 4, incluída no item 6.2. Com a inclusão do Campo Metatermo na ordem descrita acima o registro de autoridade seria composto da seguinte forma começando pelo campo 150 do formato MARC21:

Quadro 21- Exemplo de registro de autoridade do termo Engenharia Mecânica com o campo do Metatermo na ordem de campos do registro de autoridade em MARC21

Campo	Engenharia Mecânica	Conteúdo
150	a Engenharia Mecânica	termo autorizado
450	a Engineering, Mechanical	termo remissiva ver
550	w h a Dispositivos eletromecânicos	termo remissiva ver também (relação hierárquica ou associativa)
550	w h a Engenharia de produção	termo remissiva ver também (relação hierárquica ou associativa)
550	w h a Engenharia térmica	termo remissiva ver também (relação hierárquica ou associativa)

550	wh a Mecatrônica	termo remissiva ver também (relação hierárquica ou associativa)
550	wh a Movimentos mecânicos	termo remissiva ver também (relação hierárquica ou associativa)
550	wh g Engenharia	termo remissiva ver também (relação hierárquica ou associativa)
550	a Máquinas	
667	a Pode ser subdividido geograficamente	
670	a LCSH	Fonte positiva de dados
680	I Usado para obras que tratam da aplicação dos princípios da mecânica para o projeto, construção e operação de máquinas. Obras relativas a aplicação dos princípios da mecânica para a engenharia de estruturas entram em Mecânica aplicada	Nota de escopo/histórica
750	a Mechanical Engineering	Ligação de cabeçalhos autorizados (Termo da LC)
MET	ENG – Engenharia Mecânica	Metatermo
OWN	a AUTREV	indicativo de atualização por catalogador do Grupo de Linguagem da Rede de Bibliotecas da Unesp

Fonte: Elaboração dos autores.

Esse procedimento fará diferença na organização do conhecimento na medida em que será possível visualizar a área de conhecimento a que pertence cada termo do tesauro. A estratégia de busca será aprimorada porque será permitido que sejam recuperados todos os documentos representados pela área do conhecimento. É uma união perfeita da indexação de assuntos com a classificação de assuntos e esse é o maior objetivo da organização e representação do conhecimento.

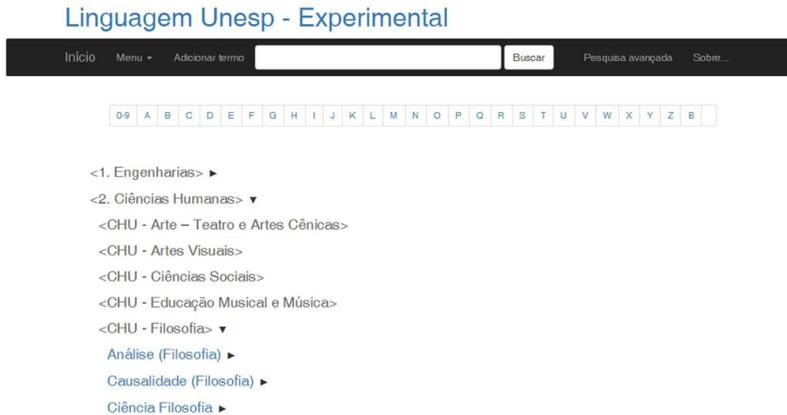
REFERÊNCIAS

- CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e classificação: o princípio de categorização. *DataGramaZero*, Rio de Janeiro, v. 9, n. 4, 2008. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/6615>. Acesso em: 05 jul. 2020.
- FUJITA, M. S. L.; MOREIRA, W.; SANTOS, L. B. P. dos; CRUZ, M. C. A. e; RIBAS, R. R. de 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.220-31, 2019.
- GRINGS, Luciana. Controle de autoridades na Biblioteca Nacional do Brasil: breve histórico e práticas atuais. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v.11, n.2, p.139-54, 2015. Disponível em: <https://rbbd.febab.org.br/rbbd/issue/view/71>. Acesso em: 27 out. 2019.
- NATIONAL INFORMATION STANDARDS ORGANIZATION. ANSI/NISO Z39.19: *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, MD: National Information Standards Organization. Disponível em: https://ils.unc.edu/courses/2015_fall/inls151_002/Readings/NISO.pdf. Acesso em: 27 out. 2019.
- SANTOS, J. C. F. *Vocabulário controlado em periódicos científicos eletrônicos: uma proposta de controle de termos*. 144 f. 2015. Dissertação (Mestrado em Ciência da Informação) – Centro de Educação, Comunicação e Artes, Universidade Estadual de Londrina, Londrina, 2015.
- SHINTAKU, M., FUJITA, M. S. L.; SCHIESSEL, M. Aspectos conceituais. In: SHINTAKU, M. (org.). *Guia do usuário do TemaTres*. Brasília: Ibict, 2019. p. 21-50.

APÊNDICE C

MANUAL DE COMPATIBILIZAÇÃO DE TERMOS COM A TAXONOMIA

No sistema da Linguagem Unesp - Experimental (<https://www.biblioteca.Unesp.br/tematres-teste>) acessar o termo a ser vinculado à estrutura hierárquica, por meio de um clique.



Verificar as seguintes alternativas:

- corretamente subordinado: colocar “OK” na lista do termo
- está incorretamente subordinado: marcar as alterações realizadas conforme orientações abaixo na lista de Termos classificados nos metatermos.

Excluir a relação “TG” do termo relacionado ao metatermo;

Clicar em “opções” e subordinar termo ao metatermo;

Análise (Filosofia)



Realizar a busca do metatermo a ser vinculado; Selecionar o metatermo e adicionar o vínculo;

[Voltar](#)

Editor de termo

Subordinar [Análise \(Filosofia\)](#) a um termo superior

[Buscar](#) [Lista de términos libres](#) [Términos top](#) [Cancelar](#)

118 termo/s encontrados na busca *filosofia*.

Tippee para filtrar términos

Termo	Data de criação
<input type="radio"/> Absurdo (Filosofia)	2018-12-19 08:46:27
<input type="radio"/> Absurdo (Filosofia) na literatura	2018-12-19 09:03:34
<input type="radio"/> Animais (Filosofia)	2018-12-19 08:53:23
<input type="radio"/> Aparência (Filosofia)	2018-12-19 08:47:39
<input type="radio"/> Voz (Filosofia)	2018-12-19 08:57:48
<input type="radio"/> Absurdo (Filosofia) na literatura	2018-12-19 09:09:14

[Adicionar](#)

Criar um metatermo (se necessário), por exemplo:

[Voltar](#)

Editor de termo

Novo termo subordinado

Termo [Get for recommendations](#)

Para adicionar vários termos de uma vez encarna uma palavra por linha.

[Enviar](#) [Cancelar](#)

Termo candidato

Meta-termo

Um meta-termo é um termo que NÃO deve ser utilizado para indexação. É um termo que descreve outros termos.
Ex: Termos guia, Facetas, Categorias, etc.

Subordinar o metatermo criado a uma pertinente grande área de conhecimento, após pesquisa a taxonomia do vocabulário controlado USP.
Modificar o vínculo no registro de autoridade.