

Manual do planejamento, construção e manutenção do tesouro Unesp para bibliotecas:

do conceitual a práxis

Mariângela Spotti Lopes Fujita
Walter Moreira
(Organizadores)



**CULTURA
ACADÊMICA**
Editora

MANUAL DO PLANEJAMENTO,
CONSTRUÇÃO E MANUTENÇÃO
DO TESAURO UNESP PARA
BIBLIOTECAS: DO CONCEITUAL A
PRÁXIS

MARIÂNGELA SPOTTI LOPES FUJITA

WALTER MOREIRA
(ORGANIZADORES)

MANUAL DO PLANEJAMENTO,
CONSTRUÇÃO E MANUTENÇÃO
DO TESAURO UNESP PARA
BIBLIOTECAS: DO CONCEITUAL A
PRÁTICA

Marília/Oficina Universitária
São Paulo/Cultura Acadêmica

2021



**CULTURA
ACADÊMICA**
Editora

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS - FFC
UNESP - campus de Marília

Diretor

Prof. Dr. Marcelo Tavella Navega

Vice-Diretor

Dr. Pedro Geraldo Aparecido Novelli

Conselho Editorial

Mariângela Spotti Lopes Fujita (Presidente)

Adrián Oscar Dongo Montoya

Célia Maria Giacheti

Cláudia Regina Mosca Giroto

Marcelo Fernandes de Oliveira

Marcos Antonio Alves

Neusa Maria Dal Ri

Renato Geraldi (Assessor Técnico)

Rosane Michelli de Castro

Parecerista

Prof.^a Dr.^a Brígida Maria Nogueira Cervantes

Professora Associada do Departamento de Ciência da Informação da Universidade Estadual de Londrina (UEL).

Ficha catalográfica

Serviço de Biblioteca e Documentação - FFC

M294 Manual do planejamento, construção e manutenção do Tesouro Unesp para bibliotecas : do conceitual a práxis / Mariângela Spotti Lopes Fujita, Walter Moreira (organizadores). – Marília : Oficina Universitária ; São Paulo : Cultura Acadêmica, 2021.
226 p. : il.
Inclui bibliografia
ISBN 978-65-5954-068-6 (Impresso)
ISBN 978-65-5954-069-3 (Digital)
DOI <https://doi.org/10.36311/2021.978-65-5954-069-3>

1. Tesouros - Desenvolvimento. 2. Organização da informação - Metodologia.
3. Linguagem documentária. 4. Indexação. 5. Registros de autoridade (Recuperação da informação). I. Fujita, Mariângela Spotti Lopes. II. Moreira, Walter.

CDD 025.47

Copyright © 2021, Faculdade de Filosofia e Ciências

Editora afiliada:



Associação Brasileira de
Editoras Universitárias

Cultura Acadêmica é selo editorial da Editora UNESP
Oficina Universitária é selo editorial da UNESP - campus de Marília

SUMÁRIO

Agradecimento -----	7
Apresentação	
<i>Mariângela Spotti Lopes Fujita, Walter Moreira</i> -----	9

PARTE I

FUNDAMENTOS TEÓRICOS E METODOLÓGICOS SOBRE TESAURO

1. LINGUAGEM DE INDEXAÇÃO PARA BIBLIOTECAS NA PERSPECTIVA DA POLÍTICA DE INDEXAÇÃO	
<i>Mariângela Spotti Lopes Fujita</i> -----	17
2. CONSTRUÇÃO E MANUTENÇÃO DE TESAuros: RELATO DE EXPERIÊNCIA SOBRE A CAPACITAÇÃO DO GRUPO DE LINGUAGEM DA REDE DE BIBLIOTECAS DA UNESP	
<i>Walter Moreira</i> -----	51
3. TESAuros: PRINCIPAIS ASPECTOS TEÓRICOS E METODOLÓGICOS PARA A SUA ELABORAÇÃO	
<i>Maria Luiza de Almeida Campos</i> -----	79

PARTE II
DIRETRIZES PARA O PLANEJAMENTO, CONSTRUÇÃO E
MANUTENÇÃO DO TESAURO UNESP

4. TERMOS DO TESAURO UNESP: ASPECTOS CONCEITUAIS E FORMAIS DOS TERMOS, RELAÇÕES ENTRE TERMOS E RELAÇÕES ENTRE CONCEITOS <i>Lucia Silva Parra; Milena Maria Rodrigues Pedrozo</i> <i>Rosane Rodrigues de Barros Ribas; Rosimara Lofiego</i> <i>Telma Jaqueline Dias Silveira</i> -----	97
4.1 ASPECTOS CONCEITUAIS DO TESAURO UNESP <i>Mariângela Spotti Lopes Fujita, Walter Moreira,</i> <i>Maria Luiza de Almeida Campos</i> -----	100
5. TAXONOMIA DE ÁREAS DO CONHECIMENTO DA UNESP: ORGANIZAÇÃO, DESENVOLVIMENTO E APLICAÇÃO DE ESTRUTURAS HIERÁRQUICAS <i>Mariângela Spotti Lopes Fujita; Luciana Beatriz Piovezan Rio Branco</i> <i>José Carlos Francisco dos Santos; Maria Marlene Zaniboni</i> <i>Minervina Teixeira Lopes</i> -----	149
6. PLANEJAMENTO DA ESTRUTURA E OPERACIONALIZAÇÃO DO TESAURO UNESP NO TEMATRES <i>Walter Moreira; Deise Maria Antonio Sabbag;</i> <i>Cibele Araújo C. M. dos Santos; José Carlos Francisco dos Santos</i> <i>Luciana Beatriz Piovezan Rio Branco; Érica Fernanda Vitorini</i> -----	193
Sobre Autores -----	219

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento do projeto de pesquisa “Política de Indexação para bibliotecas” (Processo 305648/2009-8) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelas bolsas de mestrado e doutorado que possibilitaram publicações conjuntas, bem como, dissertações e monografias de conclusão de curso que integraram este trabalho, possibilitando a construção de conhecimento contribuinte para o desenvolvimento da ciência e para os serviços bibliotecários.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento do projeto de pesquisa “Linguagem de indexação para bibliotecas na perspectiva da política de indexação” (Processo 2015/13410-8) com recursos para bolsa de treinamento técnico-3, benefícios complementares e compra de equipamentos que possibilitou o avanço da elaboração da Linguagem Unesp, instrumento imprescindível para a política de indexação em bibliotecas.

À Coordenadoria Geral de Bibliotecas da Unesp, aos Diretores de Bibliotecas da Unesp e, em especial, aos catalogadores pela participação intensa no desenvolvimento dos estudos levados a cabo pelo Grupo de Linguagem da Rede de Bibliotecas da Unesp, fundamental para a obtenção

dos resultados exitosos de planejamento, construção e manutenção do tesouro Unesp para bibliotecas.

Aos autores, catalogadores e pesquisadores do Grupo de Linguagem da Rede de Bibliotecas da Unesp pela convivência que nos possibilitou compartilhar experiências, conhecimentos e dúvidas para avaliar de forma crítica e construtiva o desenvolvimento da pesquisa que propiciou a construção do Tesouro Unesp mediante processo de compatibilização terminológica entre linguagens de indexação de mesma fonte dentro do espírito de troca de conhecimentos teórico-práticos.

APRESENTAÇÃO

Não é usual um manual de construção e manutenção de tesouros para bibliotecas. Não porque não seja possível que bibliotecas possam construir e manter um tesouro, mas porque essa não é a prática usual de bibliotecas - terem um tesouro que construíram e que cuidam para mantê-lo sempre atualizado. Porém, acreditamos que essa deve ser a prática tendo em vista a necessidade de constante atualização de vocabulário especializado e multidisciplinar exigido pela divulgação científica.

Em muitos casos a linguagem de indexação utilizada é de uma rede cooperativa de bibliotecas, ou do uso de um mesmo software integrado para a gestão dos diversos serviços de bibliotecas, ou também de uma linguagem traduzida. Nessas situações, nem sempre é possível que os catalogadores possam modificar a linguagem de indexação ou incluir termos, mesmo que o vocabulário não seja o mais adequado para a representação dos conteúdos dos documentos em processo de análise e representação. Essa situação gera frustrações na indexação e resultados ruins na recuperação.

Além disso, o ideal é que a linguagem de indexação, como ferramenta de conversão, deve ser utilizada, tanto pelo indexador após a atribuição de conceitos identificados no documento, quanto pelo usuário antes da atribuição de conceitos extraídos da necessidade de informação que orientará a estratégia de busca no catálogo. Disso depende a simples

disponibilização da mesma linguagem de indexação utilizada pelos catalogadores na indexação para o usuário junto à ferramenta de busca. A partir dessa providência relativamente simples, ficará mais fácil para o usuário entender que existe o trabalho de um catalogador por trás do resultado recuperado que satisfaz sua necessidade de informação e, por consequência, gera a valorização do trabalho do catalogador.

Entretanto, a realidade das bibliotecas demonstra ser diferente dessa conduta aparentemente simples, talvez porque a cultura do compartilhamento de registros bibliográficos prontos tenha adotado condutas de aceitação do que foi compartilhado da forma como está, sem questionamento, não porque foi imposto, mas por uma questão de respeito ao trabalho de outro profissional. Tal conduta profissional, embora eticamente responsável, esbarra nos princípios de atendimento às necessidades de informação de usuários que são sempre contextuais. O compartilhamento é importante e necessário e não obriga a aceitação irrestrita sem modificações.

Em pesquisa sobre uso de linguagens de indexação por bibliotecas da região sudeste no Brasil com questionário aplicado em amostra de 60 bibliotecas universitárias, as principais conclusões revelam o desconhecimento de profissionais sobre a função mediadora da linguagem de indexação, uso de várias linguagens de indexação por 24 bibliotecas sem representatividade na indexação e na estratégia de busca; nenhuma dessas bibliotecas disponibiliza a linguagem de indexação para consulta dos usuários, um terço das bibliotecas desenvolve linguagem própria e 10 pretendem construir, revelam ainda desconhecimento de softwares de gestão de tesouros e de construção compartilhada pela interoperabilidade (CRUZ; SANTOS; FUJITA, 2016; FUJITA; CRUZ; PATRÍCIO, 2019).

Entendemos pelos últimos resultados que essa realidade é propícia para mudanças. A prática profissional em bibliotecas tem evoluído consideravelmente com as tecnologias de informação e ambiente web facilitando o uso e manutenção de linguagens de indexação com possibilidade de interoperabilidade. Por isso, esperamos que o esforço da mudança de prática profissional realizado pelos catalogadores da rede de bibliotecas da Unesp possa servir de inspiração à outras bibliotecas.

O livro foi idealizado nos moldes do desenvolvimento da criação do Tesouro Unesp de modo a espelhar o planejamento e organização do

plano de ação proposto pelo projeto de pesquisa “Linguagem de indexação para bibliotecas na perspectiva da política de indexação”. Por isso, dividimos a organização dos textos em duas partes: Parte 1 – Fundamentos teóricos e metodológicos sobre tesouro, composta de três capítulos iniciais que tratam da gênese (capítulo 1) e dos princípios norteadores da construção e gestão do Tesouro Unesp (capítulos 2 e 3); e Parte 2 – Diretrizes para o planejamento, construção e manutenção do Tesouro Unesp, que contém três capítulos dedicados ao “Manual do planejamento, construção e manutenção do tesouro Unesp para bibliotecas: do conceitual a práxis”.

Tudo começou com o desenvolvimento do projeto de pesquisa “Linguagem de indexação para bibliotecas na perspectiva da política de indexação”, relatado no capítulo 1, que consistiu de: a) dimensão teórica sobre linguagem de indexação, sua elaboração e manutenção com enfoque em aspectos de configuração hierárquicas, semânticas e léxicas de principais linguagens de indexação utilizadas por bibliotecas; b) dimensão metodológica sobre pesquisa participante e pesquisa-ação para desenvolver, com o Grupo de Linguagem da Rede de Bibliotecas da Unesp composto de catalogadores de bibliotecas universitárias da Unesp e pesquisadores, o plano de ação para elaboração, manutenção e uso da linguagem de indexação; c) dimensão aplicada propiciou a construção e manutenção do Tesouro Unesp mediante domínio de sua gestão com o software TemaTres, a disponibilização do Tesouro Unesp junto à ferramenta de estratégia de busca integrada, a construção da Taxonomia de áreas do conhecimento da Unesp e a elaboração do Manual do Tesouro Unesp para uso dos catalogadores.

A dimensão metodológica desse projeto de pesquisa foi essencial para o desenvolvimento do plano de ação para elaboração, manutenção e uso da linguagem de indexação. A pesquisa participante e pesquisa-ação tornaram possível a oportunidade única de envolvimento cooperativo entre pesquisadora e participantes em torno da situação de elaboração, implantação, manutenção e uso da linguagem de indexação. Para isso, destacamos o apoio institucional da Coordenadoria Geral de Bibliotecas da Unesp no desenvolvimento do projeto junto à rede de bibliotecas universitárias com a participação dos catalogadores e do Grupo de Linguagem da Rede de Bibliotecas da Unesp.

Nos capítulos 2 e 3 estão os fundamentos teóricos e metodológicos inicialmente transmitidos pelos pesquisadores convidados, Prof. Dr. Walter Moreira e Profa. Dra. Maria Luiza de Almeida Campos, em dois cursos de capacitação cujos conteúdos devem ser precursores das discussões sobre a criação ou adequação do tesouro como linguagem de indexação para bibliotecas. O curso de capacitação “Construção e manutenção de tesouro”, convertido no texto do capítulo 2 pelo Prof. Dr. Walter Moreira, esclarece os elementos constituintes de um tesouro, termos, conceitos e relações conceituais, no contexto da interdisciplinaridade com a área de Terminologia e os elementos de categoria, faceta e taxonomia da área de Classificação aplicados aos tesouros com enfoque em relações terminológicas e relações conceituais. O capítulo 3 “Tesouros: principal aspecto teórico e metodológico para a sua elaboração” tem conteúdo compatível com o curso de capacitação de mesmo título ministrado pela Profa. Dra. Maria Luiza de Almeida Campos aos catalogadores da Unesp. Descreve os principais tipos de vocabulários com enfoque no Tesouro documentário e seus principais elementos, os termos, relações conceituais entre termos e a categorização de um domínio.

Os dois primeiros capítulos da Parte 2 foram pensados com base nos citados elementos essenciais de um tesouro, conceitos, termos e suas relações conceituais (capítulo 4) e a categorização (capítulo 5).

O último capítulo (6) é dedicado à operacionalização do tesouro com base no software TemaTres e tendo como referência o registro de autoridades em formato de intercâmbio MARC para interoperabilidade com outras linguagens de indexação.

Nestes três capítulos foi demonstrado como se desenvolveu um sofisticado esquema de compatibilização de linguagens de indexação de mesma fonte, a Library of Congress Subject Headings (LCSH), para a produção de um tesouro a partir de registros de autoridades em MARC já existentes e criados. A criação de novos termos sustenta a veracidade e indiscutível necessidade de bibliotecas construírem suas linguagens de indexação ainda mais neste momento de grandes avanços da catalogação. As ações de cooperação e de promoção de interoperabilidade, vale ressaltar, não lograrão sucesso sem que as bibliotecas envolvidas nesse processo estejam comprometidas com a padronização e a qualificação dos procedimentos de tratamento da informação que realizam.

O contexto profissional e a infraestrutura de informação da Unesp foram levados em consideração para o desenvolvimento de soluções que pudessem viabilizar a harmonização de toda uma política de indexação dentro da cultura de serviços e sistemas de informação da rede de bibliotecas. Provavelmente, outros contextos deverão ser estudados para que essa harmonização possa coexistir. Linguagens de indexação para bibliotecas são necessárias, mas devem, também, ser úteis aos propósitos do processo de indexação e de recuperação dentro do desenho da cultura e da política de gestão.

Mariângela Spotti Lopes Fujita

Walter Moreira

CRUZ, M.C. A.; SANTOS, L. P.; FUJITA, M. S. L. Linguagens de indexação em bibliotecas universitárias brasileiras: diagnóstico preliminar das regiões sul e sudeste. In: ENCONTRO DA ASSOCIAÇÃO DE EDUCAÇÃO E PESQUISA EM CIÊNCIA DA INFORMAÇÃO DA IBEROAMÉRICA E CARIBE - EDICIC, 10., 2016, Belo Horizonte. *Anais do Encontro da Associação de Educação e Pesquisa Em Ciência da Informação da Ibero-América e Caribe*. Belo Horizonte: ECI/UFMG, 2016. v. 1. p. 1821-1835. Disponível em: <https://pt.scribd.com/document/372647015/Anais-Edicic2016>. Acesso em: 10 jul. 2020.

FUJITA, M. S. L.; CRUZ, M. C. A.; PATRÍCIO, B.O.M. Linguagens de indexação em bibliotecas universitárias: estudo analítico. *Informação & Informação*, Londrina, v.24, p.190 -225, 2019. <http://dx.doi.org/10.5433/1981-8920.2019v24n1p190>.

PARTE 1

FUNDAMENTOS TEÓRICOS E METODOLÓGICOS SOBRE TESAURO

1. LINGUAGEM DE INDEXAÇÃO PARA BIBLIOTECAS NA PERSPECTIVA DA POLÍTICA DE INDEXAÇÃO¹

Mariângela Spotti Lopes Fujita

1.1 Introdução

1.2 Linguagem de indexação em Biblioteconomia e Ciência da Informação

1.3 A construção do tesouro Unesp: plano de ação e resultados

1.3.1 Plano de ação para a construção do tesouro Unesp

1.3.1.1 Compatibilização de registros de autoridade

1.3.1.2 Construção da Taxonomia das áreas de conhecimento na Unesp: a macroestrutura

1.3.1.3 Software para gestão e manutenção do Tesouro Unesp

1.3.1.4 Disponibilização do Tesouro Unesp para consulta e uso na indexação e recuperação da informação

1.3.1.5 Manual do Tesouro Unesp

1.4 Considerações finais

Referências

Apêndice A – Manual do Tesouro Unesp – Estrutura Preliminar

Apêndice B - Lista das publicações que demonstram o desenvolvimento das etapas do plano de ação

¹ Projeto de pesquisa desenvolvido no período de novembro de 2014 a nov. 2017 com apoio da FAPESP e do CNPq

1.1 INTRODUÇÃO

Este capítulo traz o relato do desenvolvimento do projeto de pesquisa sobre linguagem de indexação para bibliotecas que se originou da evolução da política de indexação na rede de bibliotecas da Unesp. Um dos pontos principais da política é que o processo de indexação necessita de uma linguagem dotada de vocabulário continuamente atualizado para realizar a adequada representação dos conteúdos científicos cuja característica é justamente a evolução constante. Bibliotecas sempre utilizaram listas de cabeçalhos de assuntos, cuja pré-coordenação de termos, não acompanhava a rapidez de atualização do vocabulário científico. Esta situação mudou com a transição de importantes listas de cabeçalhos de assunto para tesouros tal como a Library Congress Subject Headings, considerada linguagem de referência mundial por bibliotecas. A decisão de ter uma linguagem de indexação que pudesse ser continuamente atualizada a partir de suas fontes originais é uma proposta que se tornou realidade na política de indexação adotada pelas bibliotecas da Unesp. O objetivo final alcançado proporcionou a redação deste manual de elaboração, implantação, uso e avaliação de linguagem de indexação para bibliotecas universitárias.

A investigação sobre o tema linguagem de indexação tem como objeto de análise a elaboração, manutenção e uso de linguagem de indexação em catálogos de bibliotecas universitárias que não dispõem de instrumental teórico e metodológico para essas finalidades.

A linguagem de indexação, como ferramenta de conversão, deve ser utilizada, tanto pelo indexador após a atribuição de conceitos extraídos do documento, quanto pelo usuário antes da atribuição de conceitos extraídos da necessidade de informação que orientará a estratégia de busca no catálogo.

O uso de uma linguagem de indexação para escolha do descritor ou cabeçalho de assunto reduz a diversidade e a ambiguidade do vocabulário e estabelece uma uniformidade de representação dos termos selecionados pelo indexador para descrever o assunto dos documentos, já que vários autores podem utilizar diferentes palavras para expressar uma mesma ideia, assim como os usuários podem apresentar diversidade de vocabulário quando da expressão de uma estratégia de busca.

A linguagem de indexação faz-se necessária uma vez que a qualidade dos serviços de disseminação da informação depende da linguagem utilizada pelo sistema para a indexação e recuperação dos documentos. Isso se dá quando a linguagem do sistema permite que se traduza a linguagem natural do autor sem que se perca a ideia principal e quando permite que se traduza a linguagem do usuário de modo que satisfaça suas necessidades de informação. Esse é o foco da representação da informação dentro do processo de indexação. Para isso, depende de vários componentes para completar o conjunto de procedimentos e atingir seus objetivos de recuperação, quais sejam, o indexador, a linguagem de indexação, elementos e variáveis da política de indexação da biblioteca ou serviço de informação.

A indexação, manual, semiautomática ou automática, tem por objetivo o armazenamento das representações conceituais dos documentos para atender as necessidades de informação (GIL LEIVA, 2008, p. 70), compreendendo, fundamentalmente, duas etapas: 1) Análise de assunto; 2) Representação de conceitos.

As bibliotecas, tendo em vista o modo como realizam a catalogação cooperativa no sistema de bibliotecas universitárias, tem a necessidade de sistematizar processos e condutas de indexação e para isso dependem de uma política elaborada em consenso por catalogadores para o planejamento, elaboração e implantação de normas, procedimentos, técnicas e manual da política de indexação com orientações gerais e específicas.

Considerando-se que a investigação sobre política de indexação em bibliotecas iniciou-se pelos aspectos da indexação, a ordenação do desenvolvimento da pesquisa foi realizada tendo em vista Lancaster (2004) e Gil Leiva (2008), examinando-se:

- prática de indexação: processo de indexação, tematicidade, normas sobre indexação, manuais de indexação, indexação automática;
- qualidades da indexação: exaustividade, especificidade, correção e consistência
- ferramentas para a indexação: listas de cabeçalhos de assunto e tesauros
- avaliação da indexação

A proposta de política de indexação desenvolvida e avaliada tornou-se o “Manual de política de indexação de bibliotecas universitárias da Unesp” (FUJITA, 2014), um conjunto de condutas, procedimentos e instrumentos imersos em um contexto de gestão e envolvidos pela cultura organizacional da biblioteca.

Pelas características de desenvolvimento da pesquisa e da articulação de ambos os grupos, de pesquisadores e de catalogadores, significou que a política de indexação emanou de uma negociação entre as pessoas que realizaram um consenso, pautados por fundamentação teórica e metodológica e respaldados pelo conhecimento prático, para promoverem um serviço estratégico com resultados visíveis para a comunidade. Por isso, a proposta foi subdividida em três partes principais: os elementos da política de indexação, o processo de indexação e o instrumento da indexação.

Neste sentido, é importante observar que a pesquisa desenvolveu um produto altamente qualificado de política de indexação cujos resultados são, principalmente, a modificação das condutas de catalogação de assuntos, pois o desenvolvimento do Manual de indexação pautou-se por sua implantação e avaliação, simultaneamente. Isso significa que a política de indexação foi aceita, implantada, avaliada e determinante para a mudança de realidade de atuação profissional dos catalogadores e do catálogo online ATHENA. Dentre os itens realizados, iniciou-se, em separado, o trabalho de sistematização da Linguagem Unesp.

A linguagem de indexação da rede BIBLIODATA, anteriormente adotada para a representação de assuntos no catálogo online da rede de bibliotecas da Unesp, por não estar disponibilizada na web junto à ferramenta de busca, nunca foi conhecida ou utilizada por usuários durante a estratégia de busca. Portanto, nunca existiu a mediação da linguagem no catálogo ATHENA e, em realidade, não precisaria ser utilizada pelos catalogadores, pois a linguagem natural era a mais utilizada pelos usuários, como demonstraram os resultados obtidos do estudo de avaliação realizado por Boccato, Fujita, Gil Leiva (2011). O estudo de avaliação verificou o desempenho pelos índices de precisão e revocação de recuperação da informação de áreas científicas especializadas em 3 áreas de conhecimento no catálogo Athena (GIL LEIVA, 2008; LANCASTER, 2004).

Em consequência desse problema, foi observado que, os catalogadores na tentativa de resolver o problema de recuperação para os

seus usuários passaram a utilizar o campo 690, além do campo 650 do MARC autorizado para os cabeçalhos de assuntos da Linguagem da Rede BIBLIODATA (LCARB)². Nesse campo 690, cada biblioteca resolveu colocar descritores de outras linguagens mais especializadas e conhecidas por seus usuários e até termos de linguagem livre por falta de procedimentos sistematizados de análise documental de conteúdos e, em consequência, a linguagem não era utilizada em toda a sua capacidade de vocabulário e de estrutura sintático-semântica. Por esse motivo, o “Manual da Política de Indexação das bibliotecas universitárias da Unesp” (FUJITA, 2014) se compõe de três partes: a definição dos elementos da política de indexação que causam influência na recuperação dos assuntos no catálogo ATHENA, o processo de indexação para a catalogação de assuntos dos documentos das bibliotecas e a linguagem de indexação como instrumento de controle de vocabulário para mediar a representação temática na indexação e na estratégia de busca durante a recuperação por assuntos.

Na parte de instrumento de indexação, o texto refere-se ao uso de somente uma linguagem de indexação que o Grupo decidiu por denominar, inicialmente, de “Linguagem Unesp” considerando resultados de avaliação da Linguagem da Rede BIBLIODATA (LCARB) em comparação com a Linguagem da Biblioteca Nacional. A decisão de usar uma linguagem construída pela rede de bibliotecas da Unesp foi tomada em função de dois motivos: a LCARB estava desatualizada e não teria mais manutenção tendo em vista a desativação do sistema BIBLIODATA pela Fundação Getúlio Vargas³; foi necessário realizar a compatibilização de vocabulário de uma mesma linguagem a ser utilizada na indexação e na estratégia de busca para obter resultados de recuperação mais precisa, correta e específica.

Nessa ocasião, em que foi constatado que a LCARB não atendia às necessidades das bibliotecas universitárias da Unesp, foi necessário realizar o processo de substituição dos termos de uma linguagem pela outra à medida do processo de indexação. Em consulta ao sistema BIBLIODATA, foi assegurada a importação de registros bibliográficos sem a obrigatoriedade de adoção de sua Lista de Cabeçalhos de Assuntos para realizar, durante a indexação, a substituição da LCARB pela linguagem

² A Linguagem da Rede BIBLIODATA (LCARB) foi utilizada até 2013 pelo sistema de bibliotecas da Unesp.

³ O sistema BIBLIODATA e sua linguagem de indexação está atualmente sob responsabilidade do IBICT.

da Biblioteca Nacional, Library of Congress Subject Headings (LCSH) e Medical Subject Headings (MeSH).

As linguagens citadas para compatibilização são atualizadas e utilizadas por várias bibliotecas no Brasil e no mundo, cujos catálogos são acessíveis pela web. Além disso, é expressivo o número de descritores que possuem. A Terminologia de assuntos da Biblioteca Nacional, no Brasil, adota a LCSH como fonte de vocabulário.

A compatibilidade de vocabulários entre linguagens de mesma raiz e estrutura formaram a estrutura léxica e lógica da Linguagem Unesp: Terminologia de assuntos da Biblioteca Nacional⁴, Library of Congress Subject Headings (LCSH)⁵ e Medical Subject Headings (MeSH)⁶. Dessa forma, a linguagem UNESP foi disponibilizada com o nome de “Tesauro Unesp”⁷ ao usuário para consulta com a garantia de compatibilização de vocabulário utilizado na indexação.

Quando um registro bibliográfico é criado, o catalogador faz a descrição bibliográfica do item documentário nos campos devidos e quando se depara com os campos de assunto, recorre à Base de Autoridades de Assuntos de Registros de autoridade que formam, atualmente, o Tesauro Unesp em formato MARC21.

Os registros de autoridade destinam-se a conter informação autorizada para padronizar nomes e assuntos a serem usados como pontos de acesso em registros bibliográficos, bem como permitir os relacionamentos entre eles. Compõem uma base de dados que é consultada, tanto pelo catalogador ao atribuir os termos que representam assuntos, quanto automaticamente pelo próprio sistema para correção de termos e inclusão de remissivas dentro dos registros bibliográficos. Isso significa que os registros de autoridade têm o objetivo de assegurar a constante padronização da linguagem de indexação de forma automática e podem ser transferidos de uma base de dados para outra na formação dos catálogos de bibliotecas e rede de bibliotecas. Abaixo se demonstra exemplo de registro de autoridade de termo tópico da Biblioteca Nacional na versão de ficha completa para o usuário e no formato MARC21:

⁴ http://acervo.bn.gov.br/sophia_web/busca/autoridades

⁵ <http://id.loc.gov/authorities/subjects.html>

⁶ <https://www.nlm.nih.gov/mesh/meshhome.html>

⁷ <https://www.biblioteca.Unesp.br/tesauro/vocab/index.php>

Figura 1- Ficha completa do termo “Contratos de afretamento”

Descrição	Contratos de afretamento (subdividido geograficamente)
Library of Congress	 Charter-parties
Remissiva Ver (US/UF)	Affreightment Afretamento - Contratos Chartering of ships Contratos de frete marítimo Fretes marítimos - ContratosShip chartering
Remissiva Ver Também (TR)	TG: Direito marítimo TG: Fretes
Fonte positiva dos dados (R)	LCSH

Fonte: Catálogo de Autoridades da Biblioteca Nacional

Figura 2 - Ficha no formato MARC21 do termo “Contratos de afretamento”

000 00563cz a2200217o 4500
001 000085744
003 Br
005 20070227114454.6
008 070208d anznnbabn a ana d
035 __ a 2007020811064084med
040 __ a Br c Br f Br
150 __ a Contratos de afretamento
450 1_ a Affreightment
450 __ a Afretamento x Contratos
450 __ a Chartering of ships
450 __ a Contratos de frete marítimo
450 __ a Fretes marítimos x Contratos
450 __ a Ship chartering
550 1_ w g a Direito marítimo
550 __ w g a Fretes
670 1_ a LCSH
750 _0 a Charter-parties

Fonte: Catálogo de Autoridades da Biblioteca Nacional

Obedecendo ao formato e à Linguagem de indexação adotada, o catalogador pode formatar o cabeçalho, caso não esteja pronto, extraindo os termos autorizados da Base de Autoridades de Assuntos e os distribuindo nos campos e subcampos destinados ao assunto do registro bibliográfico.

Outro aspecto a ser colocado é referente à atualização do vocabulário controlado do Tesouro Unesp com termos novos. Considerando-se a constante evolução científica registrada pelos recursos de informação presentes nas coleções impressas e digitais das bibliotecas universitárias, surgiram termos que não são encontrados em nenhuma das linguagens de compatibilização. Foi necessário investigar procedimentos da criação de termos novos. Observamos pelos procedimentos que termos novos precisam estar vinculados à uma ou mais áreas de conhecimento para identificação, principalmente, de termos genéricos e específicos e isso depende de um conjunto de estruturas hierárquicas de áreas do conhecimento que oriente a inclusão e classificação dos termos. Nos Quadros 1 e 2 é possível verificar a estrutura hierárquica dos termos superordenados e subordinados que compõe a “Macroestrutura do Vocabulário USP”⁸:

Quadro 1- Macroestrutura com as relações Lógico-Semânticas com as áreas e terminologia em seus diferentes níveis

CA100 CIÊNCIAS AGRÁRIAS
AGRONOMIA
CA120 ENGENHARIA DE PESCA
CB200 BIOCÊNCIAS
CB210 BIOLOGIA
CB220 BOTÂNICA
CB230 IMUNOLOGIA
CB240 MICROBIOLOGIA CB250 ZOOLOGIA

Fonte: Vocabulário Controlado da USP: <http://vocabusp.sibi.usp.br/Vocab/Sibix652.dll/Mac>.

⁸ Verificar que este relato se remete à nossa experiência de participação inicial do desenvolvimento pioneiro do Vocabulário USP pela equipe de pesquisadores e profissionais da USP em “Histórico do Vocabulário Controlado do SIBI/USP”. <http://143.107.154.62/Vocab/imagens/Historico.htm>.

Quadro 2 - Vocabulário Controlado do SIBi/USP - Ordem Hierárquica dos Assuntos

CÓDIGO DA MACROESTRUTURA	DESCRITORES
CH712	ECONOMIA
CH712.1	CENSOS
CH712.1.1	CENSO AGRÍCOLA
CH712.1.2	CENSO AGROPECUÁRIO

Fonte: Vocabulário Controlado da USP: <http://vocabusip.sibi.usp.br/Vocab/Sibix652.dll/ARV?Hier=CH712>

Essa macroestrutura é uma ferramenta de classificação de organização intelectual e ordem hierárquica dos assuntos do Vocabulário SIBi/USP que pode ser denominada de taxonomia. A Linguagem de indexação Unesp não possuía uma taxonomia que precisava ser criada para orientar a alteração de termos importados da Linguagem da Rede BIBLIODATA (LCARB) e a inclusão de termos novos pela organização dos registros de autoridade. Aparentemente, a Terminologia de Assuntos da Biblioteca Nacional (BN) e a Library of Congress Subject Headings (LCSH) não possuem uma taxonomia que organiza a ordem hierárquica dos assuntos dentro de seus registros de autoridade, porém, todos os termos possuem uma organização hierárquica evidente nas fichas dos termos da Biblioteca Nacional que segue a Classification Web da Library of Congress⁹ como se observa nas Figuras 3 e 4:

Figura 3 - Ficha da autoridade – Termo tópico “Engenharia de Transportes” na Linguagem da Biblioteca Nacional

Descrição	Engenharia de transportes (subdividido geograficamente)
Library of Congress	 Transportation engineering
Remissiva Ver Também (TR)	TE: Engenharia de tráfego TE: Engenharia ferroviária TE: Engenharia rodoviária TG: Engenharia TG: Engenharia civil
Fonte positiva dos dados (R)	LCSH

Fonte: Terminologia de assuntos da Biblioteca Nacional .

⁹ <https://classweb.org/>

Figura 3 - Ficha da autoridade – Termo Tópico “Banks and banking, Cooperative” da LCSH

Banks and banking, Cooperative

From Library of Congress Subject Headings URI(s)
<http://id.loc.gov/authorities/subjects/sh85011651> info:lc/authorities/sh85011651 <http://id.loc.gov/authorities/sh85011651#concept>

Instance Of

MADS/RDF Topic

MADS/RDF Authority

SKOS Concept [↗](#)

Scheme Membership(s)

Library of Congress Subject Headings

Collection Membership(s)

LCSH Collection - Authorized Headings

LCSH Collection - General Collection

LCSH Collection - May Subdivide Geographically

Variants

-  Cooperative banks
-  Credit cooperatives
-  People's banks

Broader Terms

-  Banks and banking
-  Cooperative societies

Narrower Terms

-  Agricultural cooperative credit associations
-  Credit unions
-  Rotating credit associations

Closely Matching Concepts from Other Schemes

-  Banche cooperative [↗](#)
-  Banques coopératives [↗](#)
- 

Genossenschaftlicher Bankensektor

Fonte: Library of Congress Subject Headings.

O processo de substituição de uma linguagem pelas outras foi iniciado em junho de 2012 e está em andamento a partir da construção em formato MARC21 de registros de autoridade de assuntos tópicos e geográficos para descritores da linguagem Unesp. Esses procedimentos, descritos em Ribas e Parra (2016), são etapas da construção da Linguagem Unesp que altera termos importados da Linguagem LCARB ou inclui termos novos conforme procedimentos da terminografia.

Considerando-se a necessidade de construção da linguagem de indexação bem como a observação do problema de estruturação hierárquica de áreas de conhecimento da Unesp, foi realizada a investigação sobre o processo de construção do Tesauro Unesp a partir da compatibilização de registros de autoridade de outras linguagens de indexação de mesma fonte e elaboração da Taxonomia contendo as estruturas hierárquicas de áreas de conhecimento da Unesp na perspectiva de sua política de indexação.

Para isso, foi realizada sistematização teórica e metodológica de elaboração de linguagem de indexação com e para bibliotecas com os objetivos de: elaboração de linguagens para bibliotecas obterem boa recuperação por assunto de seus catálogos online; fornecer subsídios teóricos e metodológicos para elaboração, implantação e avaliação de linguagens de indexação em ambientes de bibliotecas; e avaliar estratégias de intervenção metodológica mediante formação de equipes de elaboração, implantação e avaliação de linguagens de indexação em bibliotecas.

1.2 LINGUAGEM DE INDEXAÇÃO EM BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO

A linguagem de indexação é o instrumento de conversão da linguagem natural do documento durante o processo de indexação e, por isso, exerce importante influência na recuperação da informação cujos efeitos são controlados e avaliados pela política de indexação adotada pela biblioteca ou serviço de informação (FUJITA, 2012). No contexto de sistemas de recuperação da informação dotados de ferramentas e padrões de representação, as linguagens de indexação são fundamentais e aceitas no âmbito teórico e metodológico tanto da Biblioteconomia e Ciência da Informação quanto da Organização do Conhecimento (MAZZOCHI, 2018)

O uso de uma Linguagem de Indexação¹⁰, pelo indexador e pelo usuário do sistema de recuperação da informação, destina-se, de forma pragmática, à escolha do termo correto para descrever a fim de reduzir a diversidade e a ambiguidade de vocabulário. Desse modo, realiza uma mediação entre a linguagem do texto e a linguagem do usuário e estabelece uma representação precisa por meio de termos que correspondem aos termos utilizados pelo autor para expressar o assunto do texto.

A linguagem de indexação é formada de vocabulário e sintaxe. O vocabulário se refere à relação dos descritores usados para a identificação do conteúdo de um documento e a sintaxe se refere as regras utilizadas para a combinação dos descritores usados para a identificação do conteúdo de um documento.

Analizados seus aspectos de estrutura e funcionalidade, a linguagem de indexação como ferramenta de controle de vocabulário apresenta-se por meio de tipologias diferenciadas pela forma e objetivos diversos.

Nos sistemas pré-coordenados os termos são previamente combinados de acordo com regras sintáticas estabelecidas pela linguagem. Por outro lado, nos pós-coordenados os termos serão combinados de acordo com a sintaxe da lógica de busca. Gil Urdiciain (2004, p.24) conclui que “[...] nas linguagens pré-coordenadas a relação entre os termos é gramatical e que na linguagem pós-coordenada a relação é lógica.” As linguagens de estrutura combinatória, como os tesouros, permitem a combinação de seus termos a partir da necessidade de indexação.

Os tesouros constituem-se de descritores que podem ser combinados ou não entre si na indexação, relacionados entre si semanticamente e genericamente. Representam primeiramente os conceitos para depois especificar suas relações e são desenvolvidos para sistemas pós-coordenados. Constituem-se como Linguagem de indexação alfabética, controlada, pós-coordenada e combinatória.

Listas de cabeçalhos de Assunto são linguagens de indexação concebidas, anteriormente, a partir de cabeçalhos de assunto cuja composição prévia de dois ou mais termos realizava uma pré-coordenação.

¹⁰ Linguagem de indexação é denominação utilizada por Lancaster (2004) e Gil Leiva (2008), autores representativos da linha “Indexing” e Linguagem de indexação ou Documental é utilizada por autores como Guimarães (2008), Gil Urdiciain (2004), autores representativos da linha de “Análise Documentária”.

Apresentam sinônimos, quase sinônimos e homógrafos. Preveem regras específicas tais como forma de entrada do descritor, uso de singular/plural, uso de abreviaturas e de termos traduzidos. Geralmente são listas gerais que arrolam termos de todas as áreas do conhecimento, tendo em vista o contexto de bibliotecas que incluem várias áreas do conhecimento.

A linguagem de indexação criada por bibliotecas é a Lista de Cabeçalhos de Assuntos. Existem, portanto, muitas listas de cabeçalhos de assunto, algumas desenvolvidas por bibliotecas individuais que não tem acesso às redes de cooperação. As principais listas de cabeçalhos de assunto, atualmente existentes e atualizadas, são:

Quadro 1 - Principais Listas de cabeçalhos de assunto

Lista de cabeçalho de assunto	Breve descrição
Library Congress of Subject Headings (LCSH) (http://id.loc.gov/authorities/subjects.html)	“Os cabeçalhos de assuntos da Library of Congress (LCSH) foram desenvolvidos e são mantidos pela Biblioteca do Congresso dos EUA, inicialmente para as coleções da Biblioteca. LCSH, um vocabulário controlado usado para indexar, catalogar e procurar registros bibliográficos em catálogos de bibliotecas e bancos de dados eletrônicos, tornou-se um padrão de fato para bibliotecas que usam registros de catalogação divulgados pela Library of Congress ou que catalogam de acordo com os padrões estabelecidos colaborativamente pela Library of Congress e pela comunidade de bibliotecas. LCSH é atualizada diariamente na Classification Web” (AMERICAN LIBRARY ASSOCIATION, 2019).
Medical Subject Headings (MeSH) https://www.nlm.nih.gov/mesh/meshhome.html	“O tesauro Medical Subject Headings (MeSH) é um vocabulário controlado e hierarquicamente organizado, produzido pela Biblioteca Nacional de Medicina dos EUA. É usado para indexação, catalogação e pesquisa de informações biomédicas e relacionadas à saúde. O MeSH inclui os títulos dos assuntos que aparecem no MEDLINE / PubMed, no Catálogo NLM e em outros bancos de dados NLM.” (NATIONAL LIBRARY OF MEDICINE, 2019)

<p>Sears List of Subject Headings https://searslistofsubjectheadings.com/page/frontmatter</p>	<p>A lista de Sears, desde 1923 até a atualidade, foi desenvolvida para interpretar as necessidades das bibliotecas de médio e pequeno porte mantendo, ao mesmo tempo, as possibilidades de expansão. Com edições sucessivas, cada vez mais ampliadas, é destinada, também, para grandes bibliotecas. Uma adaptação da lista de Sears para a língua portuguesa, de autoria de Wanda Ferraz, publicada pelo Instituto Nacional do Livro em 1944 e atualizada em 1972 e 1977, foi uma contribuição às pequenas bibliotecas brasileiras. Possui versão impressa em formato livro, disponível em livrarias, e em formato eletrônico mediante subscrição em base de dados online.</p>
<p>Terminologia de assuntos da Biblioteca Nacional do Brasil http://acervo.bn.gov.br/sophia_web/busca/autoridades</p>	<p>A Biblioteca Nacional do Brasil (www.bn.br) disponibiliza online a Terminologia de Assuntos como linguagem de indexação atualizada pela equipe de bibliotecários em português e inglês com base na Library of Congress Subject Headings e coordena um consórcio formado por bibliotecas públicas, universitárias e governamentais.</p>
<p>Lista de cabeçalhos de Assuntos da rede BIBLIODATA (LCARB)</p>	<p>A uniformização e padronização das entradas de assunto são garantidas pela FGV pela manutenção de uma Lista de Cabeçalhos de Assunto para uso de bibliotecas brasileiras participantes da Rede. Esta</p>
	<p>Lista, representada como a Bases de Autoridades Assuntos, é fundamentada na Library of Congress Subject Headings (LCSH), mantendo a estrutura de linguagem pré-coordenada, respeitando as características da língua portuguesa (FUNDAÇÃO GETÚLIO VARGAS, 1995, p. 11-12). Atualmente, indisponível está sob responsabilidade do IBICT.¹¹</p>
<p>Descritores em Ciências da Saúde (DeCS) http://decs.bvs.br/</p>	<p>Utilizado para a indexação e recuperação da informação em fontes de informação disponíveis na Biblioteca Virtual em Saúde (BVS), foi desenvolvido a partir do Medical Subject Headings MeSH.</p>
<p>Vocabulário Controlado USP (VocaUSP) http://vocabusp.sibi.usp.br/vocab/</p>	<p>“[...] é uma linguagem documentária construída, a partir de procedimentos terminológicos e documentários, pelos bibliotecários do SIBi/USP, com a participação de especialistas de todas as áreas do conhecimento abrangidas pelos seus descritores.” (LIMA, BOCCATO, 2009). Surgiu com base no vocabulário de listas de assuntos compiladas pelas bibliotecas mais especializadas na elaboração dos catálogos de assuntos que tem uma mistura de princípios de cabeçalhos de assuntos e de tesouros.</p>

Fonte: Sistematização com base em informações coletadas na web e texto de Agustín-Lacruz, Fujita e Terra (2014).

¹¹ Conforme esclarecimentos de Cecília Leite Oliveira, Diretora do IBICT, em 01 de julho de 2020, o BIBLIODATA está passando por processo de revisão e atualização tecnológica e, por essa razão, o acesso ao serviço tem sofrido instabilidade.

A sintaxe dos cabeçalhos de assuntos se manifesta no universo dos cabeçalhos compostos de que são exemplos típicos os cabeçalhos-frase ou as expressões compostas. Os cabeçalhos de assuntos servem para expressar assuntos ou o conteúdo e/ou forma da informação registrada e entende-se, como sintaxe de cabeçalhos de assunto, a combinação dos elementos ou palavras que formam os cabeçalhos compostos.

Atualmente, em sistemas de busca online a pré-coordenação perde o efeito de sequência e de prioridade porque o programa de busca não se importa em como as palavras são combinadas durante a indexação, mas sim na combinação de termos durante a estratégia de busca avançada que os sistemas de descoberta aprimoraram. Esta evolução tecnológica dos catálogos online proporcionou às listas de cabeçalhos de assunto aproveitarem seus recursos de controle de vocabulário acumulados por muitos anos e adotarem padrões de organização do conhecimento desenvolvidas por tesouros. Nessa transição, foi necessária a inclusão de estruturas hierárquicas para se obter as ligações semânticas entre termos de uma mesma área de assunto.

A lista de cabeçalhos de assuntos é a forma mais antiga de linguagem de indexação baseado no sistema alfabético de assunto. Sua prática, desenvolvida na Library of Congress (LC) a partir da elaboração dos catálogos, sedimentou o uso do cabeçalho de assunto. Entretanto, sua atualização para uso nos catálogos online transformou sua configuração e visualização com uma aparência de linguagem de indexação póscoordenada tal como a dos tesouros que possuem estrutura lógica e hierarquizada entre os termos.

Ressalte-se que a Library Congress of Subject Headings – LCSH, a mais importante lista de cabeçalhos de assunto e a linguagem de indexação alfabética mais completa e específica da atualidade, evoluiu em suas últimas edições com a introdução e uso da nomenclatura de relações semânticas adotadas por tesouros, incluindo uma estrutura lógica de relações hierárquicas anteriormente atrofiada na linguagem.

Partindo da ideia de catalogação cooperativa e uso de recursos tecnológicos a Library of Congress (LC) criou a maior Central de Catalogação Cooperativa. A introdução da tecnologia computacional na década de 1960 e o desenvolvimento do formato MARC para padronização de registros bibliográficos e formatos de intercâmbio reflete-se até os dias

atuais nas bibliotecas e centros de informação. Atualmente os registros da LC são disponibilizados em outros formatos de intercâmbio e possibilitam acesso a outros vocabulários controlados que utilizam descritores idênticos e suas variações terminológicas.

O desenvolvimento de tecnologias de software e hardware juntamente com a Internet, proporcionaram às centrais de catalogação cooperativa disponibilizarem seus catálogos em meio digital, acelerando a troca de informações.

Para propiciar o intercâmbio de informações, foram criados, junto ao formato MARC, os protocolos Z39.50 e ISO-2709 que orientam sobre a estrutura de um registro e a transferência de dados bibliográficos, o que “[...] permite que um registro criado para uma máquina por um determinado software seja lido por qualquer outro software em qualquer tipo de máquina” (MARASCO; MATTES, 1995, p. 42). Essa evolução da LCSH, os formatos de intercâmbio de registros bibliográficos e os catálogos online demonstram que são importantes para a finalidade de recuperação da informação.

Na perspectiva da gestão e planejamento, a compatibilização dos registros de autoridade da LCARB feita com a adoção da LCSH, da Terminologia de Assuntos da Biblioteca Nacional e do MeSH, linguagens de indexação de mesma fonte, reduziu problemas estruturais e proporcionou resultados quantitativos e qualitativos para a criação do Tesouro Unesp, sobretudo no que se refere à atualização contínua do vocabulário especializado de modo confiável.

1.3 A CONSTRUÇÃO DO TESAURO UNESP: PLANO DE AÇÃO E RESULTADOS

O planejamento e a gestão de tesouros devem ser realizados por equipes multidisciplinares em que diferentes profissionais e pesquisadores especializados precisam realizar uma negociação para obter consenso de suas propostas e decisões. A aplicação da metodologia de pesquisa-ação permitiu que as ações fossem negociadas durante o desenvolvimento da pesquisa para elaboração de um manual da linguagem de indexação com vistas à sua constante atualização e manutenção.

Considerando que o desenvolvimento desta pesquisa contou com o apoio da Coordenadoria Geral de Bibliotecas e dos catalogadores das bibliotecas da Unesp em função dos resultados obtidos na pesquisa “Política de indexação para bibliotecas”¹² que os levou a decidir sobre a elaboração, manutenção e uso da Linguagem de indexação para suas bibliotecas, foi possível aproveitar a oportunidade única de envolvimento cooperativo entre pesquisadora, grupo de pesquisa e participantes em torno da situação de elaboração, implantação, manutenção e uso da linguagem de indexação.

Nesse sentido, esclarecemos que o Grupo de Política de indexação da Unesp que elaborou, implantou e avaliou a “Política de Indexação para as bibliotecas universitárias da Unesp” (FUJITA et al., 2016) discutiu e decidiu, no bojo da pesquisa-ação, pela elaboração da Linguagem de indexação da Unesp e para isso foi criado, em 2013, o Grupo de estudos da elaboração da Linguagem de Indexação que atuou na colaboração efetiva desta investigação.

A pesquisa-ação propiciou a comunicação de diferentes perspectivas do contexto de indexação pelos participantes, usuários, catalogadores, dirigentes de bibliotecas e bibliotecários, assim como revelaram o que pensavam sobre os problemas e restrições desse contexto, como seria possível resolvê-los e quais os instrumentos e métodos necessários, o que ajudou no aprimoramento da política de indexação. Mais do que isso, deu aos catalogadores a possibilidade de compreensão e, conseqüente domínio, de como desenvolver a organização e representação do conhecimento no contexto de bibliotecas universitárias.

A intervenção da pesquisa-ação foi possível mediante reuniões quinzenais com equipe formada por catalogadores dedicados à elaboração de registros de autoridades e pesquisadores da área de Ciência da Informação de modo a promover a reflexão crítica com a finalidade de identificar problemas ou avaliar mudanças. À cada dúvida, problema ou necessidade de implantar novos procedimentos foram realizadas duas modalidades de seminários: de vivências profissionais e acadêmicas ou compartilhamento de experiências com pesquisadores convidados para transferência de conhecimento quando o desenvolvimento dependia de metodologias ou ferramentas a serem implantadas.

¹² Projeto de pesquisa desenvolvido no período de 2010 a 2018 (Processo CNPq: 305648/2009-8).

Os pesquisadores do Grupo de Linguagem da Rede de Bibliotecas da Unesp apresentaram a sistematização teórica e metodológica sobre linguagens de indexação e experiências de elaboração de linguagens de indexação com uso de softwares a partir das considerações analíticas e comparativas do desenvolvimento da pesquisa.

Nos seminários de vivências haviam, após as apresentações dos pesquisadores, discussões sobre o contexto profissional de modo a levantar as dúvidas, problemas e restrições para os quais eram apresentadas propostas que, dependendo da situação eram destacadas para possível ação dentro do plano. Em alguns casos, outros pesquisadores eram convidados para transferência de conhecimentos durante as reuniões quinzenais ou mediante cursos de capacitação que nos ajudassem, de modo viável, a ter outros pontos de vista para solucionar as questões a serem ajustadas para o bom desenvolvimento de uma determinada ação. Os pesquisadores convidados e respectivas ações foram:

Ações	Pesquisador convidado
Escolha do software	José Carlos Francisco dos Santos (PPGCI/UNESP); Cibele Araújo C. M. dos Santos (USP-São Paulo); Tiago Rodrigo Marçal Murakami (SIBI/USP)
Taxonomia de áreas especializadas	Maria Luiza de Almeida Campos (UFF) Deise Maria Antonio Sabbag (USP-Ribeirão Preto)

Os cursos de capacitação para o Grupo de Linguagem da Rede de Bibliotecas da Unesp ministrados pelos pesquisadores do grupo e convidados foram:

“A terminologia na elaboração de linguagens de indexação” - 28 de abril de 2014 -

Profa. Dra. Mariângela Spotti Lopes Fujita

“Procedimentos para autorização de termos novos considerando a metodologia terminográfica” - 28 de abril de 2014

Profa. Dra. Mariângela Spotti Lopes Fujita

“Construção e manutenção de tesauro” 31 ago. 2015

Prof. Dr. Walter Moreira

Profª. Dra. Mariângela Spotti Lopes Fujita

“Elaboração da macroestrutura da linguagem Unesp: análise de modelos de categorias” -

23 de outubro de 2015

Dr. Walter Moreira

“Vocabulário Controlado: principais aspectos para o tratamento e a recuperação de informações” - 12 de Abril de 2018

Dra. Maria Luiza de Almeida Campos

O resultado das discussões e intervenções possibilitou a construção de um plano de ação para elaboração, manutenção e uso da linguagem de indexação. Em resumo, o uso da linguagem viabilizou-se a partir da escolha de um software implementado em 2017 e avaliado pelas bibliotecas durante o ano de 2018. Com o software TemaTres foi possível a importação dos registros de autoridade de assunto em formato MARC21 cuja compatibilização havia sido iniciada em 2012. Essa importação marcou a criação do Tesauro Unesp para consulta de todos os usuários profissionais da rede de bibliotecas e os usuários da comunidade externa. O vocabulário, antes invisível para todos por estar em internalizado em linguagem de máquina, tornou-se aparente, acessível e disponível. Tudo isso foi possível mediante a elaboração e desenvolvimento do plano de ação no contexto da pesquisa-ação.

1.3.1 PLANO DE AÇÃO PARA A CONSTRUÇÃO DO TESAURO UNESP

O plano de ação, elaborado nas reuniões do Grupo de Linguagem da Rede de Bibliotecas da Unesp, definiu quatro metas principais:

- a) Compatibilização dos registros de autoridade da linguagem BIBLIODATA com as linguagens de mesmo vocabulário e estrutura: Library of Congress Subject Headings (LCSH), Terminologia da Biblioteca Nacional (TBN) e Medical Subject Headings;

- b) Criação e desenvolvimento de uma Taxonomia de áreas do conhecimento da Unesp formada por categorias e subcategorias para classificação dos termos contidos nos registros de autoridade compatibilizados;
- c) Escolha de um software para possibilitar a gestão e manutenção do Tesouro Unesp com a importação dos registros de autoridade compatibilizados;
- d) Disponibilização do Tesouro Unesp junto à interface de busca integrada que dá acesso ao catálogo.

1.3.1.1 COMPATIBILIZAÇÃO DE REGISTROS DE AUTORIDADE

O contexto de contínua geração de novos conhecimentos e de inovação da universidade foi determinante para a construção da Linguagem Unesp. Entretanto, foi necessário desenvolver uma metodologia de compatibilização para uso colaborativo que consiste na combinação dos registros de autoridade de assuntos em formato MARC21 de outras linguagens documentais para linguagem Unesp. O estudo de Fujita; Santos; Cruz; Moreira e Ribas (2018) desenvolveu processo inovador de compatibilização de vocabulários de três linguagens de indexação de mesma matriz terminológica utilizadas por bibliotecas: Terminologia de assuntos da Biblioteca Nacional, Library of Congress Subject Headings (LCSH) e Medical Subject Headings (MeSH).

Entretanto, a compatibilização demonstrou que é necessário incluir termos novos a partir de nova metodologia, criada especificamente para o Tesouro Unesp. Nesse sentido, foi realizado em abril de 2014, o curso de formação em serviço “A terminologia na elaboração de linguagens de indexação” para o grupo de catalogadores, participantes do Grupo de Política de indexação da Unesp com a finalidade de discutir a elaboração da Linguagem de indexação da Unesp. Este curso formou o grupo inicial de estudos da elaboração da Linguagem de Indexação Unesp, capacitando-os para a inclusão de termos novos e alteração de termos conforme Método Terminográfico.

A proposta teórica e metodológica para elaboração, manutenção e uso de linguagem de indexação em bibliotecas consiste na combinação

dos registros de autoridade de assuntos em formato MARC21 de outras linguagens documentais para linguagem Unesp descrita em Ribas e Parra (2016) e em Fujita (2017).

Esta meta, implementada pelos catalogadores da Unesp, possibilitou a compatibilização atual de 89.581 termos¹³ e está em continuidade.

1.3.1.2 CONSTRUÇÃO DA TAXONOMIA DAS ÁREAS DE CONHECIMENTO NA UNESP: A MACROESTRUTURA

A segunda meta, desenvolvida e descrita no Capítulo 5 deste livro, foi a elaboração da Taxonomia de áreas do conhecimento da Unesp utilizando os procedimentos para a construção da estrutura hierárquica do Tesauro Unesp com uso da ferramenta Classification Web da Library of Congress Subject Headings (LCSH), macroestrutura do Vocabulário USP e estrutura hierárquica da Terminologia de Assuntos da Biblioteca Nacional.

As principais considerações referem-se à proposta de uma metodologia de compatibilização de uso colaborativo para elaboração, manutenção e uso de linguagem de indexação para catálogos online de bibliotecas que consiste na construção de macroestrutura a partir da análise das estruturas hierárquicas das linguagens LCSH, Terminologia de Assuntos da Biblioteca Nacional (TBN), LCARB e Vocabulário USP.

A metodologia proposta utilizou o método de comitê ascendente da Diretriz ANSI/NISO2005 com uso do Modelo metodológico integrado, proposto por Cervantes (2009). A aplicação do método comparou as hierarquias da LCSH, TBN, LCARB e Vocabulário USP nas áreas de Física e Matemática, demonstrando ser apropriado para a construção de macroestrutura de linguagens de indexação que permita a classificação dos termos e seus relacionamentos.

Verificou-se que a escolha das linguagens LCSH, TBN, BIBLIODATA e Vocabulário USP foi adequada para a análise das estruturas de categorias e das hierarquias lógicas. A linguagem LCARB não possui completa hierarquização de termos e sua utilização é inviável.

¹³ <https://www.biblioteca.Unesp.br/tesauro/vocab/sobre.php>.

As linguagens de indexação destinadas à análise de macroestruturas devem estar constantemente atualizadas e direcionadas a contextos de uso semelhantes. A LCSH destaca-se por sua exaustividade e diversidade terminológica. A TBN, embora seja uma versão da LCSH, não apresenta diversidade terminológica. É necessária a combinação das duas linguagens. O Vocabulário USP apresenta estrutura diferenciada com características das áreas de domínio da universidade. As três linguagens demonstram ser importantes para a construção de estruturas hierárquicas com profundidade adequada ao âmbito das bibliotecas universitárias.

Com o avanço das discussões sobre a elaboração da estrutura hierárquica contamos com a colaboração da Profa. Dra. Maria Luiza de Almeida Campos da Universidade Federal Fluminense que passou a integrar institucionalmente o Grupo de Linguagem da Rede de Bibliotecas da Unesp. A participação da Profa. Maria Luiza tem sido muito importante pela sua experiência na construção de vocabulários controlado, sobretudo na modelização de conceitos. Muitas informações, conhecimentos e decisões das reuniões do Grupo foram registradas e forneceram os principais contributos do trabalho de pesquisa sobre construção e avaliação da estrutura lógica e hierárquica de linguagens de indexação para catálogos online.

A proposta teórica e metodológica para elaboração, manutenção e uso de linguagem de indexação em bibliotecas consiste na combinação dos registros de autoridade de assuntos em formato MARC21 de outras linguagens documentais para linguagem Unesp e a elaboração de estruturas lógico-hierárquicas das áreas de conhecimento existentes na universidade que permitam a classificação dos termos que representam assuntos e seus relacionamentos em registro de autoridade de assuntos. A elaboração de estruturas será, portanto, a de incluir as semelhanças e as diferenças entre as três estruturas para que a exaustividade e a especificidade sejam contempladas de modo a garantir profundidade no permanente desenvolvimento da linguagem de indexação que se pretende para bibliotecas universitárias.

1.3.1.3 SOFTWARE PARA GESTÃO E MANUTENÇÃO DO TESAURO UNESP

O estudo de softwares, desenvolvido e descrito no capítulo 6, foi realizado, inicialmente, por meio de revisão de literatura com enfoque nos

softwares TemaTres e MultiTes (FUJITA et al. 2017a). O enfoque dado aos dois softwares foi definido em função das características e funcionalidades que ambos apresentam para uso na interface de busca do catálogo online ATHENA e a existência de literatura específica sobre a avaliação dos dois softwares. O Grupo de Linguagem da Rede de Bibliotecas da Unesp escolheu o TemaTres com base em duas variáveis influentes: a) ser um software livre, gratuito e com atualizações constantes fundamentadas em normas internacionais que preconizam a interoperabilidade semântica e tecnológica; b) ter sido implementado pelo Sistema de Bibliotecas da USP para gestão e manutenção do Vocabulário USP e servir como interface de busca por assuntos pela linguagem de indexação.

Portanto, para realizar a implementação do software TemaTres realizamos reuniões com a participação da Profa. Dra. Cibele Araújo Camargo Marques dos Santos da ECA/USP, do servidor técnico e pesquisador do SIBI/USP, Tiago Murakami, do aluno de doutorado do PPGCI, José Carlos Francisco dos Santos, especialista em Informática, e do servidor técnico Oberdan Luiz May da Coordenadoria Geral de Bibliotecas, responsável pelo catálogo online ATHENA e especialista da área de informática. Atualmente, estamos realizando testes com a importação dos registros de autoridade compatibilizados e, também, está em desenvolvimento o aplicativo para fazer a interface do TemaTres com o Aleph.

1.3.1.4 DISPONIBILIZAÇÃO DO TESAURO UNESP¹⁴ PARA CONSULTA E USO NA INDEXAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO

Tesauro é uma linguagem artificial utilizada em bibliotecas e em outros sistemas de informação para organizar e representar o conteúdo de diferentes publicações científicas de modo que seja possível, a qualquer usuário, recuperar a informação necessária acerca do conhecimento científico armazenado em suas bases de dados tais como catálogos, repositórios e bibliotecas digitais.

O tesauro é uma linguagem com a função específica de representação do conhecimento durante a armazenagem e a busca em bases de dados. Para isso, é composto de palavras-chave (descritores)

¹⁴ <https://www.biblioteca.Unesp.br/tesauro/vocab/index.php>.

relacionadas semântica e genericamente com as áreas de conhecimento especializado. Dessa forma, a escolha de uma palavra-chave do tesauro garantirá, ao usuário que realiza uma determinada busca, a coincidência com a palavra-chave utilizada para a representação conteúdo de um ou mais documentos que serão recuperados.

O Tesauro da Unesp tem vocabulário com termos especializados das áreas de conhecimento de suas atividades de ensino, pesquisa e extensão. Cada termo compatibilizado e operacionalizado por um registro de autoridade em formato de intercâmbio MARC21 (vide Figuras 5 e 6 do item 6.2 da Seção 6, p.). A compatibilização do termo mediante registro de autoridade é realizada, conforme resumido no item 1.3.1.1 e detalhadamente esclarecido na Seção 5 para que se possa ter atualização do vocabulário do Tesauro Unesp. O registro de autoridade compatibilizado, possibilita tanto o controle e correção internos do tesauro Unesp nas bases de dados em que atua, quanto sua exportação para qualquer outro sistema externo como é o caso do software TemaTres onde assume o formato customizado de um tesauro pronto para ser compreendido e usado por qualquer usuário que necessite empregar seus termos.

É utilizado pelos bibliotecários catalogadores para representar o conteúdo da informação mais significativa de livros, dissertações, teses, monografias, trabalhos de conclusão de curso, artigos de periódicos, documentos, legislação e etc.

O mais importante é que foi desenvolvido para ser utilizado nas interfaces de busca integrada das bases de dados das bibliotecas durante a consulta de assunto por alunos de graduação e de pós-graduação, bem como por pesquisadores para recuperar informações sobre livros, artigos, teses, dissertações, monografias, ou qualquer outro documento que tenha sido indexado em qualquer base de dados da Unesp. Para isso, basta acessar o link do tesauro Unesp, abaixo da interface de busca, digitar uma palavra, parte da palavra ou frase na caixa de pesquisa e escolher a palavra-chave mais representativa da necessidade de informação. Ao lado direito da palavra-chave tem o link de acesso para a escolha dos registros bibliográficos.

A construção e manutenção do Tesauro da Unesp é realizada desde 2013 pelo Grupo de Linguagem da Rede de Bibliotecas da Unesp composto por catalogadores das bibliotecas e pesquisadores da área de Organização

do Conhecimento, supervisionados pela Profa. Dra. Mariângela Spotti Lopes Fujita (Programa de Pós-Graduação em Ciência da Informação da Unesp – Campus de Marília).

A disponibilização do Tesauro Unesp para consulta de suas palavras-chave é parte fundamental da Política de Indexação das bibliotecas Universitárias da Unesp (<https://www.biblioteca.Unesp.br/portal/arquivos/manual-politica-indexacao-2017.pdf>) com finalidade de aprimoramento da recuperação da informação mais precisa por seus usuários durante a busca por assuntos.

O processo de construção e manutenção sempre dará continuidade à correção das palavras-chave e, principalmente da consistência dos relacionamentos hierárquicos entre os termos.

1.3.1.5 MANUAL DO TESAURO UNESP

A investigação sobre os manuais fundamentou-se na Norma ISO 25964-1, Parte I- tesauro para recuperação da informação com enfoque na seção 13 sobre gestão de tesauros que propiciou discussão de proposta de sistematização de elaboração, implantação, uso e avaliação de linguagem de indexação para bibliotecas no contexto da política de indexação a partir das orientações definidas no item 3, “O instrumento de indexação – linguagem Unesp”, do Manual de Política de Indexação da Unesp (FUJITA, 2014). Essa discussão resultou na recomendação de uma estrutura preliminar para elaboração do “Manual do Tesauro Unesp” (Apêndice A) desenvolvida pelo Grupo de Linguagem da Rede de Bibliotecas da Unesp cuja estrutura, finalizada, compõe este livro.

1.4 CONSIDERAÇÕES FINAIS

A indexação na catalogação é a finalidade desta pesquisa de cunho aplicado para que os catálogos possibilitem a recuperação por assuntos mais precisa e específica. A principal contribuição para a área de Ciência da Informação, em especial da Biblioteconomia, é este manual de elaboração, implantação e avaliação de linguagem de indexação que faça parte do Manual de política de indexação para assegurar a qualidade da recuperação por assuntos na atual conjuntura de catalogação cooperativa.

Recomenda-se aos demais pesquisadores e profissionais a necessidade de: - Elaboração de linguagens de indexação para bibliotecas obterem boa recuperação por assunto de seus catálogos on-line;

- Contribuir com mais subsídios teóricos e metodológicos para elaboração, implantação e avaliação de linguagens de indexação em ambientes de bibliotecas;
- Aplicar e avaliar estratégias de intervenção metodológica mediante formação de equipes de elaboração, implantação e avaliação de linguagens de indexação em bibliotecas;
- Propiciar à formação do indexador o conhecimento e prática suficiente de uma metodologia consistente elaboração, avaliação e controle de linguagens de indexação capacitando-o em metodologias da Ciência da Informação e das áreas interdisciplinares; - Contribuir para o desenvolvimento curricular de conteúdos programáticos das disciplinas da área de Organização da Informação de Cursos de Graduação em Biblioteconomia e de cursos de capacitação em indexação para formação em serviço de profissionais da informação;
- Disseminar a importância e a influência que a linguagem de indexação exerce sobre o desempenho da recuperação da informação por assunto em biblioteca;
- Conscientizar e capacitar os responsáveis pela política de indexação: dirigentes e catalogadores em bibliotecas e sistemas de bibliotecas.

Considera-se concluído o plano de trabalho do projeto de pesquisa em apreço com diversas contribuições científicas em forma de publicações, porém, as principais contribuições são as experiências de compartilhamento de conhecimentos entre os membros do Grupo de Linguagem da Rede de Bibliotecas da Unesp, sobretudo para os alunos de pós-graduação e de graduação envolvidos na pesquisa que tiveram suas pesquisas fortalecidas com a riqueza de conhecimentos teóricos, metodológicos e tecnológicos. Destaca-se, também, as soluções importantes para a construção de uma Linguagem Unesp com o catálogo em funcionamento para uso de muitos

usuários internos e externo. Esse projeto de pesquisa teve essa importante repercussão de conscientizar os catalogadores para a importância de um controle de vocabulário sempre atualizado e disponível para busca e recuperação da informação. Com certeza o Grupo de Linguagem da Rede de Bibliotecas da Unesp continuará atuante por muito tempo pois a linguagem de indexação necessita de atualização constante e de respaldo científico e tecnológico.

Conforme previsto no plano de trabalho, enfatiza-se a continuidade das reuniões do grupo de catalogadores agora na composição da Comissão Permanente do Tesauro Unesp cuja atuação segue os procedimentos de pesquisa participante com metodologia de pesquisa-ação. Essas reuniões propiciaram várias orientações, levantamento de dúvidas, decisões e relatos registrados em atas e anotações que conduziram vários dos trabalhos publicados, em manuais, artigos, livros e capítulos de livros com a colaboração dos integrantes do Grupo de Linguagem da Rede de Bibliotecas da Unesp em um trabalho coletivo de muita participação de todos os envolvidos. A produção científica realizada pelo Grupo (Apêndice B) atesta de forma inequívoca o acerto da escolha da pesquisa ação como indutora das decisões inovadoras que proporcionaram os relevantes resultados à Unesp e à toda comunidade científica e profissional.

REFERÊNCIAS

AGUSTÍN LACRUZ, M. del C.; FUJITA, M.S.L.; TERRA, A.L.S. Linguagens documentais para bibliotecas escolares: o caso da Espanha, Portugal e Brasil. *Informação & Sociedade*, João Pessoa, v.24, n.3, p.83-97, set./dez. 2014.

AMERICAN LIBRARY ASSOCIATION. *Cataloging tools and resources: subjectheadings*. 2019. Disponível em: <https://libguides.ala.org/catalogingtools>. Acesso em: 01 dez. 2020.

BOCCATO, V. R. C.; FUJITA, M. S. L.; GIL LEIVA, I. Avaliação comparada do uso de linguagens de indexação em catálogos de bibliotecas universitárias para recuperação por assunto. *Scire*, Zaragoza, v.17, p.55 – 64, 2011.

CERVANTES, B. M. N. *A construção de tesouros e a integração de procedimentos terminográficos*. 2009. 209 f. Tese (Doutorado em Ciência da Informação) –Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.

FUJITA, M. S. L. A política de indexação para representação e recuperação da informação. In: GIL LEIVA, I.; FUJITA, M.S.L. (ed.). *Política de indexação*. São Paulo:

- Cultura Acadêmica, 2012. p.17-28. Disponível em: http://www.culturaacademica.com.br/_img/arquivos/Livro%20Politica-deindexacao_ebook.pdf. Acesso em: 01 dez. 2020.
- FUJITA, M. S. L. *Manual da política de indexação para as bibliotecas universitárias da Unesp*. Marília: Faculdade de Filosofia e Ciências, Unesp; São Paulo: Coordenadoria Geral de Bibliotecas, Reitoria da Unesp, 2014.
- FUJITA, M.S.L (org.). *Política de indexação para bibliotecas: elaboração, avaliação e implantação*. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2016. Disponível em: http://www.marilia.Unesp.br/Home/Publicacoes/politicas-de-indexacaopara-bibliotecas_ebook.pdf. Acesso em: 01 dez. 2020.
- FUJITA, M.S.L *et al.* Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias. *Scire*, Zaragoza, v.23, p.63-73, 2017a. Disponível em: <http://www.ibersid.eu/ojs/index.php/scire>. Acesso em: 01 dez. 2020.
- FUJITA, M.S.L.; SANTOS, L.B.P; CRUZ, M.C.A.; MOREIRA, W.; 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 (UNESP). *Knowledge Organization*, Baden-Baden, v.45, p. 220 - 231, 2018.
- FUNDAÇÃO GETÚLIO VARGAS. *Manual de cabeçalhos de assunto: normas e procedimentos*. Rio de Janeiro: FGV, 1995.
- GIL LEIVA, I. *Manual de indización: teoría y práctica*. Gijón: Trea, 2008.
- GIL URDICIAIN, B. *Manual de lenguajes documentales*. 2. ed. rev. y ampl. Madrid: Noesis, 2004.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Tradução de Antonio Agenor Briquet de Lemos. 2. ed. Brasília: Briquet de Lemos, 2004.
- LIMA, V. M. A.; BOCCATO, V. R. C. O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semi-automática. *Perspect. Ciênc. Inf.*, Belo Horizonte, v. 14, n. 1, Apr. 2009. <http://dx.doi.org/10.1590/S1413-99362009000100010>. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362009000100010&lng=en&nrm=iso. Acesso em: 14 jun. 2014.
- MARASCO, L.C.; MATTES, R.N. Avaliação e seleção de software para automação de centros de documentação e bibliotecas. *Cadernos da FFC*, Marília, v. 4, n. 1, p. 40-49, 1995.
- MAZZOCCHI, F. Knowledge Organization Systems (KOS). *Knowledge Organization*, Baden-Baden, v.45, p.54-78, 2018. doi:10.5771/0943-7444-2018-1-54 .
- NATIONAL LIBRARY OF MEDICINE. *Medical Subject Headings*. 2019. Disponível em: <https://www.nlm.nih.gov/mesh/meshhome.html>. Acesso em: 01 dez. 2020.

RIBAS, R.R.deB.; PARRA, L.S. A construção e atualização da linguagem Unesp. *In*: FUJITA, M.S.L (org.). *Política de indexação para bibliotecas*: elaboração, avaliação e implantação. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2016. p.6787. Disponível em: http://www.marilia.Unesp.br/Home/Publicacoes/politicas-deindexacao-para-bibliotecas_ebook.pdf. Acesso em: 01 dez. 2020.

APÊNDICE A

MANUAL DO TESAURO UNESP¹⁵

ESTRUTURA ERELIMINAR

Introdução [Capítulos, 1, 2 e 3]

Parte 1: Fundamentos conceituais e normativos da elaboração do tesauro

1. Visão geral e objetivos do tesauro [Capítulo 4]
 - 1.1 Símbolos, termos abreviados e outras convenções
 - 1.2 Aspectos conceituais do tesauro (diferenciar o que é conceito e o que é termo, explicitar o significado da terminologia utilizada em tesauros, tais como termo específico, tesauro multilíngüe e etc)
2. Abrangência temática do tesauro [Capítulo 5]
3. Termos do tesauro [Capítulo 4]
 - 1.1 Forma dos termos (termos simples e termos compostos)
 - 1.2 Forma gramatical de termos (substantivos, adjetivos, advérbios, gênero, número, maiúsculas e minúsculas, abreviaturas e siglas, hífen, parênteses, elementos numéricos)
 - 3.3 Esclarecimento e desambiguação dos termos do dicionário de sinônimos ou controle terminológico para seleção do termo preferido (Homônímia, antonímia, sinonímia, termos estrangeiros, gírias ou jargões, nomes comuns e comerciais, nomes populares e nomes científicos e etc)
 - 3.4 Relação de equivalência
 - 3.4.1 A relação de equivalência, num contexto monolíngüe (Sinônimos, Quasesinônimos, Termos específicos incluídos num conceito mais amplo)
 - 3.4.2 Equivalência entre idiomas
 - 3.5 Relações entre conceitos e termos
 - 3.5.1 A relação hierárquica
 - 3.5.2 A relação associativa
 - 3.6 Notas explicativas
 - 3.7 Definição
 - 3.8 Análise de faceta

Parte 2: Planejamento da estrutura e operacionalização do tesauro [Capítulo 6]

5. Estrutura do banco de dados
 6. Protocolos do banco de dados usados com o tesauro
 7. Apresentação e layout
 7. A escolha do software de gerenciamento do tesauro
 8. Formato de troca: registros de autoridade em formato MARC
 9. Interoperabilidade com outras linguagens de indexação
 10. Integração com aplicações de indexação e pesquisa
- Parte 3: Construção do tesauro [Capítulo 5]
10. A estrutura lógico-hierárquica
 11. Atualização do vocabulário
 - 11.1 Correção de termos existentes no banco de dados
 - 11.2 Compilação dedutiva de termos novos

¹⁵ Capítulos correspondentes da organização atual neste livro.

APÊNDICE B

LISTA DAS PUBLICAÇÕES QUE DEMONSTRAM O DESENVOLVIMENTO DAS ETAPAS DO PLANO DE AÇÃO

ARTIGOS EM REVISTAS CIENTÍFICAS INDEXADAS

FUJITA, M.S.L.; SANTOS, L. B. P. Política de indexação em bibliotecas universitárias: estudo diagnóstico e analítico com pesquisa participante. *Transinformação*, Campinas, v. 28, p. 5976, 2016. Disponível em: www.scielo.br/pdf/tinf/v28n1/0103-3786-tinf-28-01-00059.pdf. Acesso em: 21 nov. 2020.

FUJITA, M. S. L.; SANTOS, L. B. P. A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias. *Scire*, Zaragoza, v.22, n.2, p.37-46, jul./dic. 2016. Disponível em: <http://www.iberid.eu/ojs/index.php/scire/article/view/4336/3866>. Acesso em: 21 nov. 2020.

FUJITA, M.S.L.; SANTOS, L.B.P; CRUZ, M.C.A.; MOREIRA, W. Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias. *Scire*, Zaragoza, v.23, p.63-73, 2017. Disponível em: <http://www.iberid.eu/ojs/index.php/scire>. Acesso em: 21 nov. 2020.

FUJITA, M.S.L.; SANTOS, L.B.P; CRUZ, M.C.A.; MOREIRA, W.; 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 (UNESP). *Knowledge Organization*, Baden-Baden, v.45, p.220 - 231, 2018.

FUJITA, M.S.L.; CRUZ, M.C.A.; PATRÍCIO, B.O.M. Linguagens de indexação em bibliotecas universitárias: estudo analítico. *Informação & Informação*, Londrina, v.24, p.190 -225, 2019. <http://dx.doi.org/10.5433/1981-8920.2019v24n1p190>.

ARTIGOS EM REVISTAS CIENTÍFICAS NÃO INDEXADAS

FUJITA, M. S. L. A linguagem documentária na negociação de uma política de indexação para bibliotecas universitárias: procedimentos e estratégias da pesquisa-ação integral. *Revista Conhecimento em Ação*, Rio de Janeiro, v.1, n.1, p.2-17, jan./jun.2016. Disponível em: <https://revistas.ufjf.br/index.php/rca/article/view/3555/2782>

TRABALHOS APRESENTADOS EM CONFERÊNCIAS NACIONAIS

CRUZ, M.C.A.; SANTOS, L.P.; FUJITA, M.S.L. Linguagens de indexação em bibliotecas universitárias brasileiras: diagnóstico preliminar das regiões sul e sudeste. *In: ENCONTRO DA ASSOCIAÇÃO DE EDUCAÇÃO E PESQUISA EM CIÊNCIA DA INFORMAÇÃO DA IBEROAMÉRICA E CARIBE - EDICIC*, 10., 2016, Belo Horizonte. *Anais do Encontro da Associação de Educação e Pesquisa Em Ciência da Informação da Ibero-América e Caribe*. Belo Horizonte: ECI/UFMG, 2016. v. 1. p. 1821-1835. Disponível em: <https://pt.scribd.com/document/372647015/Anais-Edicic-2016>. Acesso em: 10 jul. 2020.

CRUZ, M.C.A.; FUJITA, M.S.L.; SANTOS, L.B.P. Linguagem de indexação no contexto da política de indexação: estudo em bibliotecas universitárias. *In: PINTO, F.A.; GUIMARÃES, J.A.C. (org.) Memória, tecnologia e cultura na organização do conhecimento*. Recife: Ed.UFPE, 2017. v.4. p. 217-224. (Trabalho apresentado no IV Congresso Brasileiro em Organização e Representação do Conhecimento – IV ISKO Brasil, realizado em setembro de 2017 em Recife). Disponível em: <http://isko-brasil.org.br/>. Acesso em: 21 nov. 2020.

FUJITA, M.S.L.; CRUZ, M.C.A.; PATRÍCIO, B.O.M. A construção de tesouros na perspectiva dos manuais de indexação. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 18., Marília, 2017. Disponível em: <http://enancib.marilia.Unesp.br/index.php/xviiienancib/ENANCIB>. Acesso em: 21 nov. 2020.

SANTOS, J.C.F. dos; CERVANTES, B.M.N.; FUJITA, M.S.L. Tesouro eletrônico: importação no TemaTres e disponibilização na web. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 19., 2018, Londrina. *Anais do XIX Encontro Nacional de Pesquisa em Ciência da Informação*. Londrina: UEL, 2018. v. 1. p. 1-19.

SANTOS, J.C.F. dos; FUJITA, M.S.L. Tesouro Unesp: Integração do registro de autoridade para o TemaTres. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 19., 2018, Londrina. *Anais do XIX Encontro Nacional de Pesquisa em Ciência da Informação*. Londrina: UEL, 2018. v. 1. p. 1-20.

FUJITA, M.S.L. Construção de linguagem de indexação para bibliotecas universitárias: uma experiência da Unesp com a pesquisa-ação integral. *In: SEMINÁRIO DO GRUPO DE PESQUISA MHTX*, 3., 2018, Belo Horizonte. *Anais do III Seminário do Grupo de Pesquisa MHTX: perspectivas em representação e organização do conhecimento: atualidades e tendências na relação universidade-empresa*. Belo Horizonte: ECI/UFMG, 2018. v. 3. p. 36-40.

FUJITA, M. S. L. Construção e manutenção de tesouro em bibliotecas universitárias: uma experiência de interoperabilidade semântica. *In: CAMPOS, M.L. de A. et al. (org.). Produção, tratamento, disseminação e uso de recursos informacionais heterogêneos*. Niterói: IACS/UFF, 2018, v. 5, p. 110-116. Disponível em: <https://app.uff.br/riuff/handle/1/9195>. Acesso em: 10 jul. 2020.

TRABALHOS APRESENTADOS EM CONFERÊNCIAS INTERNACIONAIS

FUJITA, M.S.L ; RUBI, M.P. Compatibilidade entre linguagens documentais para construção, atualização e adequação de vocabulário de bibliotecas universitárias. *In: RODRÍGUEZ MUÑOZ, J.V.; GIL-LEIVA, I; DÍAZ ORTUÑO, P; MARTÍNEZ MÉNDEZ, F.J. (ed.). Organización del conocimiento: sistemas de información abiertos. Actas del XII Congreso ISKO España y II Congreso ISKO España y Portugal, 19 y 20 de noviembre.* Murcia, 2015. p. 345-356. Disponível em: <http://www.iskoiberico.org/congresos/murcia-2015/>. Acesso em: 21 nov. 2020.

FUJITA, M.S.L.; SANTOS, L.B.P.; CRUZ, M.C.A.; MOREIRA, W. La gestión del language de indización por software: evaluación para uso en control de los registros de autoridad de catálogos en línea de bibliotecas. *In: ENCUENTROS INTERNACIONALES SOBRE SISTEMAS DE INFORMACIÓN Y DOCUMENTACIÓN (IBERSID 2017)*, 22., Zaragoza, 2 a 4 out. 2017. Disponível em: <http://www.ibersid.es/wpcontent/uploads/2017/06/Programa.pdf>. Acesso em: 21 nov. 2020.

FUJITA, M.S.L.; SANTOS, L.B.P.; CRUZ, M.C.A.; MOREIRA, W. Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias. *Scire*, Zaragoza, v.23, p.63-73, 2017. Disponível em: <http://www.ibersid.eu/ojs/index.php/scire/article/view/4430/3901>. Acesso em: 21 nov. 2020.

FUJITA, M.S.L.; SANTOS, L.B.P.; CRUZ, M.C.A.; MOREIRA, W.; RIBAS, R.R. de B. Construction and evaluation of logical hierarchical structures of indexing languages for online catalogs of libraries *In: ISKO UK Conference 2017: Knowledge Organization: what's the story?*, 2017, Londres. [Trabalho submetido como artigo ao periódico Knowledge Organization]. Disponível em: <http://www.iskouk.org/content/iskomedia>. Acesso em: 21 nov. 2020.

FUJITA, M.S.L.; SANTOS, L.B.P. dos; SANTOS, N. de S. dos. A função do registro de autoridade de assunto na construção e uso de linguagens de indexação para catálogos online. *In: CONGRESO ISKO ESPANA Y III CONGRESO ISKO ESPANA-PORTUGAL*, 13., 2017, Coimbra. Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO. Coimbra: Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20, 2017. v. 1. p. 577-586. Disponível em: <https://purl.org/sci/atas/isko2017>. Acesso em: 21 nov. 2020.

LIVRO PUBLICADO COM MEMBROS DA EQUIPE COMO ORGANIZADOR

FUJITA, M.S.L (org.) *Política de indexação para bibliotecas: elaboração, avaliação e implantação.* Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2016. Disponível em: http://www.marilia.Unesp.br/Home/Publicacoes/politicas-de-indexacao-parabibliotecas_ebook.pdf. Acesso em: 21 nov. 2020.

CAPÍTULO DE LIVRO

SHINTAKU, M.; FUJITA, M.S.L; SCHIELSSL, M. Conceitos gerais. *In*: SHINTAKU, M. (org.). *Guia do usuário do TemaTres*. Brasília: IBICT, 2019. v. 1. p. 21-50. Doi: <http://DOI: 10.18225/9788570131553>.

2. CONSTRUÇÃO E MANUTENÇÃO DE TESAUROS: RELATO DE EXPERIÊNCIA SOBRE A CAPACITAÇÃO DO GRUPO DE LINGUAGEM DA REDE DE BIBLIOTECAS DA UNESP

Walter Moreira

2.1 Introdução

2.2 Linguagem documentária e tesouro

2.3 Aspectos normativos na construção do tesouro

2.3.1 Tesouros

2.3.2 Terminologia

2.4 Termos, conceito e relações conceituais

2.5 O conceito de categoria, faceta e taxonomia aplicados aos tesouros

2.6 Relações terminológicas e relações conceituais

2.7 Considerações finais

Referências

2.1 INTRODUÇÃO

O Tesouro Unesp é fruto de um longo processo, resultado do envolvimento direto e indireto de muitas pessoas, como ocorre com qualquer

outro projeto dessa magnitude e ainda está, como também se espera, em construção. Algumas experiências estão relatadas aqui e ali, neste livro, cujo objetivo também é evidenciar a importância dos tesouros como sistemas de organização do conhecimento (SOC) utilizados para organizar, representar e recuperar informações em ambientes universitários ou outros que envolvam o ensino superior e a pesquisa científica. Os relatos de experiência agrupados neste livro, assim como o que aparece neste capítulo, também visam a inspirar novos projetos destinados à elaboração e manutenção de tesouros ou outros SOC. Espera-se que tais iniciativas auxiliem a povoar as redes de bibliotecas universitárias, formal ou informalmente constituídas, com instrumentos terminológicos de controle de vocabulário, tendo em vista sua indiscutível relevância nesses e em outros ambientes de informação.

Apresenta-se neste capítulo uma espécie de síntese geral dos conteúdos expostos e discutidos em duas ocasiões pontuais com o grupo de catalogadores da rede de bibliotecas da Unesp que participaram, e ainda participam, do planejamento e da elaboração da Tesouro Unesp. Trata-se do oferecimento de dois cursos de capacitação, ambos ministrados a convite e em conjunto com a Profa. Mariângela Spotti Lopes Fujita, coordenadora do Projeto, organizados conforme as informações apresentadas no Quadro 1.

Quadro 1 – Cursos de capacitação oferecidos ao Grupo de Linguagem da Rede de Bibliotecas da Unesp

Curso 1	Construção e manutenção de tesouros
Data de realização	31 ago. 2015
Ementa	O processo de representação temática em suas relações interdisciplinares (aspectos linguísticos e lógicos) como subsídio à compreensão da estrutura e funcionalidade de linguagens documentárias. Caracterização da linguagem documentária como linguagem. As linguagens documentárias alfabéticas: tesouro e lista de cabeçalho de assunto.
Objetivos	compreender a função e importância das linguagens documentárias no processo de análise documentária; diferenciar o uso e aplicação das linguagens documentárias alfabéticas: lista de cabeçalho de assunto e tesouro; caracterizar a estrutura e funcionalidade das linguagens documentárias alfabéticas.
Conteúdo	Linguagem documentária e tesouro Normas ISO e ANSI sobre vocabulários controlados Taxonomia de linguagens documentárias Relações terminológicas e relações conceituais 5 Exercícios

Curso 2	Elaboração da taxonomia da Linguagem Unesp
Data de realização	02 out. 2015
Ementa	Categorias e campo semântico. Definição de parâmetros para a criação de categorias. O processo de elaboração da estrutura de categorias de vocabulários controlados.
Objetivos	Compreender o conceito de categoria e sua aplicação na definição de taxonomia de linguagens documentárias
Conteúdo	O conceito de categoria Distinções entre categorização e classificação As categorias fundamentais de Ranganathan e do CRG 4 Exercícios

Fonte: Elaborado pelo autor (2020).

O que se apresenta a seguir não se refere exatamente aos materiais ou recursos utilizados nos cursos, tendo-se em vista que a organização deste texto lhes é posterior e é sumarizadora, mas ao esforço de síntese das ideias então discutidas como mais um subsídio para compreensão do processo de construção do Tesauro Unesp.

2.2 LINGUAGEM DOCUMENTÁRIA E TESAURO

Uma linguagem documentária é uma linguagem construída e utilizada para auxiliar na representação e na descrição dos assuntos dos documentos em ambientes de informação, algo que é normalmente referido na literatura como “representação temática”. A forma de adjetivação pelo uso da expressão “temática”, nesse caso, é utilizada para diferenciar o escopo desse modelo de representação da “representação descritiva”, ainda que esta aparente dicotomia atenda mais a questões didáticas do que epistemológicas propriamente ditas. Enquanto a representação temática vincula-se à produção de condensações dos textos, gerando índices e resumos, a representação descritiva vincula-se à identificação e descrição de aspectos que possibilitam a identificação material dos documentos. Neste caso, a identificação dos elementos ocorre por “apreensão instantânea” (KOBASHI, 2008, p. 52).

A distinção entre representação temática e representação descritiva, reafirma-se, não deve ser tomada em aspecto absoluto, pois a identificação do autor de uma obra, por exemplo, normalmente referida

como representação descritiva, configura-se, também como uma indicação de assunto, isto é, a partir da identificação do nome do autor é possível, muitas vezes, saber, com maior ou menor precisão, qual o assunto do documento.

Dentre as linguagens documentárias, destacam-se, por sua maior robustez, os sistemas de classificação – tais como a Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU) ou o sistema de classificação utilizado pela Biblioteca do Congresso, nos Estados Unidos, a Library of Congress Classification (LCC) – e os tesouros, tais como o Thesaurus Agrícola Nacional, mantido pela Biblioteca Nacional de Agricultura (BINAGRI) ou o Art & Architecture Thesaurus, desenvolvido pelo Getty Institute Research.

Linguagens documentárias como os sistemas de classificação são chamadas de linguagens documentárias notacionais, os tesouros, por sua vez, são denominados linguagens documentárias alfabéticas. As primeiras utilizam notações (de modo prioritário) e expressões verbais para designar os assuntos e ordenam os conceitos hierarquicamente. As últimas usam expressões verbais para designar os assuntos e ordenam os conceitos de forma prioritariamente alfabética. Enquanto os sistemas de classificação instrumentalizam o processo de classificação e contribuem para a inserção dos documentos em classes, em processos de generalização, os tesouros são aplicados em processos de indexação, contribuindo para buscas mais precisas em processos de especialização (individualização) dos documentos. Observem-se, por fim, os destaques dados na frase às prioridades, pois ao final ambos os modelos de linguagem documentária possuem partes alfabéticas e sistemáticas em relações de complementaridade.

Os tesouros são sistemas de organização do conhecimento que possibilitam a identificação e a descrição dos sistemas conceituais que constituem os domínios, juntamente com as teorias e ontologias que lhes sustentam. Funcionam também como instrumentos de padronização das representações de modo a oferecer condições para que pesquisadores (ou usuários, de modo geral), sistemas de informação documentária e profissionais da informação responsáveis pela construção de representações documentárias (indexadores) utilizem a mesma forma verbal ou seus equivalentes em outros idiomas para se referir a um determinado conceito. Obviamente que aproximar discussões sobre linguagem, univocidade e

padronização nunca é algo tão pacífico e que seria uma ideia ingênua, para dizer o mínimo, imaginar que a complexidade dessas relações pode ser traduzida e resolvida por meio de qualquer algoritmo.

Um destes aspectos complexos envolve discussões sobre padronização. Um tesauro, considerando-se os aspectos pragmáticos que circundam sua construção e seu uso e os aspectos dinâmicos das culturas, pode promover a padronização por meio da harmonização. No âmbito da terminologia, a harmonização, conforme a compreensão desse conceito expressa em Pavel e Nolet (2002, p. 30): “combina o desejo de precisão conceitual e correção linguística, a adequação do termo à situação de comunicação e a eficácia da comunicação”. Com a finalidade de torná-lo mais claro, ao conceito de “harmonização” contrapõem-se os conceitos de “normalização” e de “recomendação” apresentados em Barros (2004, p. 87-88):

A normalização se dá com base em medidas coercitivas, adotadas por uma autoridade política ou de outra natureza e, normalmente, é fruto de um contexto sociolinguístico particular. [...] A perspectiva da recomendação é outra e significa que um termo deve ser empregado preferencialmente em relação a outros sinônimos [...] Um termo recomendado poderá eventualmente ser normalizado se ele conseguir eliminar seus concorrentes. [...]

Tálamo (1997, p. 3), por exemplo, aponta a “tensão entre a padronização da informação e a possibilidade de inseri-la em sistemas de comunicação mais potentes” como marca característica do desenvolvimento dos processos documentários. Há, evidentemente, outros aspectos semânticos e pragmáticos complexos envolvidos nessa questão e eles são amplamente discutidos na literatura da ciência da informação. Assim, as questões relativas à construção de linguagens documentárias constituem o objeto da linguística documentária. À análise documentária cabem as questões que dizem respeito ao tratamento e a recuperação da informação (TÁLAMO, 1997).

O tesauro, um sistema simbolicamente instituído (CINTRA et al., 2002) e que é utilizado como instrumento comutador, tem por finalidade facilitar a comunicação. Neste caso, refere-se à comunicação que ocorre, ou precisa ocorrer, nos contextos documentários e que é mediada

por uma linguagem construída que requer, por essa mesma condição, regras explícitas para seu uso com recurso de vocabulários controlados. Atendo-se à sua função, a finalidade precípua do vocabulário controlado é, afinal, fazer coincidir o vocabulário do indexador e o vocabulário do pesquisador.

As metodologias para construção e manutenção de tesouros evoluíram significativamente nos anos recentes e algumas orientações relativamente abrangentes sobre o assunto podem ser encontradas em livros, e.g. Broughton (2006), artigos, e.g. Mcculloch (2005) ou páginas na internet, e.g. Zeng (2005). As normas técnicas, respeitadas as limitações inerentes à sua natureza, também oferecem subsídios importantes para o tema. Por fim, a teoria da terminologia também tem sido incorporada como elemento fundamental em relação aos aportes teóricos e metodológicos que fornece.

Na sequência, apresentam-se alguns aspectos normativos e terminológicos aplicados à construção e manutenção de tesouros.

2.3 ASPECTOS NORMATIVOS NA CONSTRUÇÃO DE TESAuros

De modo geral, as normas aplicadas à documentação visam a atender a um desejo universal no universo da recuperação da informação: viabilizar a realização de apenas uma consulta simultânea e bem-sucedida em todas as bases de dados, repositórios e outras coleções relevantes, sem necessidade de qualquer reformulação. Sabe-se que proposta em si é utópica e envolve aspectos políticos, econômicos e sociais cuja discussão este capítulo não comportaria. Por ora, pode-se discutir os cuidados que se pode tomar na construção de tesouros enquanto instrumentos que podem potencialmente melhorar a qualidade das representações realizadas por indexadores e usuários de sistemas de informação documentária.

Um dos pontos críticos em relação à intenção de utilizar tesouros com a finalidade de ampliar o universo de busca, e aproximar-se da “utopia” indicada no parágrafo anterior, diz respeito às ações de interoperabilidade semântica entre os diversos tipos de SOC como condição necessária. Nesses casos, as normas internacionais podem efetivamente contribuir, pois no negócio da informação, ressalta Clarke (2010, p. 43) “o isolamento nacional está condenado”. A perspectiva internacional, globalizada, também segue

pela via errada. Espera-se que os tesouros e os outros SOC de modo geral, atendam às perspectivas da glocalização.

Apesar de ainda se configurar como um neologismo, não registrado nos dicionários de língua mais conhecidos, o conceito já tem aparecido com alguma frequência na literatura científica. Em busca rápida, sem aplicação de filtros, foram localizados 554 mil resultados no Google Acadêmico; na base de dados Scopus, já com alguma restrição na estratégia de busca por meio da expressão “KEY(glocalization) AND PUBYEAR > 2010”, foram identificados 226 documentos. A glocalização é uma reação à globalização, à ideia de homogeneização e imperialismo cultural que esse conceito carrega. A gênese do termo relaciona-se ao slogan utilizado pelo CEO da Sony, no Japão, Akio Morita: “pense globalmente, aja localmente”, em tradução livre. De modo simplificado, glocalização é “a integração de diferenças e práticas culturais locais em iniciativas, programas ou projetos baseados em uma estrutura de globalização” (FRANCOIS, 2015, p. 62, tradução livre), isto é, a definição de objetivos em escala global sem perder de vista as realidades locais.

No que se refere às normas internacionais aplicadas à construção de tesauro, é preciso ficar sempre atento aos problemas locais que ultrapassam os aspectos técnicos que podem ser alcançados por meio de normalização, aos problemas decorrentes do caráter universalista e aos limites ditados pelas tendências ou inclinações que o seu contexto de produção lhes impõe.

Embora haja muitas semelhanças de estrutura e de propósitos entre os diferentes SOC, não existem normas individuais para cada um deles e muito menos uma norma geral que os contemple a todos. Destacam-se nesse cenário as normas sobre os topic maps¹ – apresentadas brevemente neste capítulo apenas para efeitos de comparação e construção do texto – as normas sobre tesouros e, adicionalmente, as normas sobre terminologia que se aplicam à construção de tesouros.

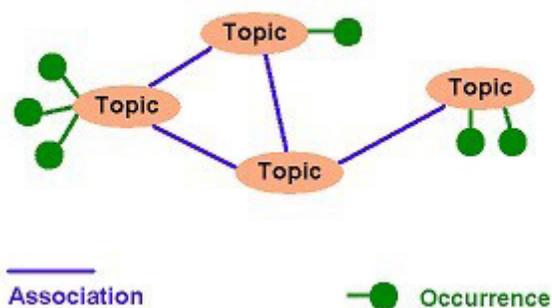
Os topic maps são tipos de SOC que surgiram no final da década de 1990, como resposta à busca por um modo mais efetivo de

¹ Apesar de alguns registros na literatura brasileira da expressão em português, “mapas de tópicos”, essa expressão parece ainda não estar consagrada, motivo pelo qual optou-se pela utilização do termo em inglês. Como recurso para mensurar a popularidade do termo, realizou-se busca no Google Acadêmico, mantendo-se as expressões entre aspas. Foram localizados 12.500 resultados para “topic maps” contra apenas 191 resultados para “mapas de tópicos”.

mesclar e organizar índices de finais de livros (COLMENERO RUIZ, 2005; GARSHOL, 2004). Nesse sentido, são inspirados nas técnicas de classificação bibliográfica e da indexação, de modo especial as técnicas de indexação aplicadas à construção de índices de final de livro.

Topic maps são mapas de conhecimento organizados a partir da identificação de tópicos (assuntos ou temas), os quais são utilizados para representar digitalmente alguma coisa. Seus elementos básicos são o tópico, a associação e a ocorrência (Figura 1).

Figura 1 – Estrutura básica de um topic map



Fonte: By Hirzel at English Wikipedia - Transferred from en.wikipedia to Commons by Econt., Public Domain, <https://commons.wikimedia.org/w/index.php?curid=3621665>.

A norma aplicada aos topic maps é a ISO/IEC 13520, publicada em 2002 pela International Organization for Standardization (ISO) em cooperação com a International Electrotechnical Commission (IEC). Essa norma apresenta elementos de arquitetura e definições dos principais termos relacionados aos topic maps (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2002).

2.3.1 TESAUROS

A norma mais completa sobre construção de tesauros é a norma ISO 25964 – Information and documentation — Thesauri and interoperability with other vocabularies (Informação e documentação – Tesauros e

interoperabilidade com outros vocabulários) (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, 2013).

Essa norma foi publicada em duas partes, respectivamente: parte 1: thesauri for information retrieval (tesauros para a recuperação da informação), publicada em 2011, com 160 páginas, e parte 2: interoperability with other vocabularies (interoperabilidade com outros vocabulários), publicada em 2013, com 106 páginas.

A parte 1 cobre: a) conteúdo e construção do tesouro, mono ou multilíngue; b) diretrizes sobre a aplicação de análise facetada em tesauros; c) diretrizes sobre a gestão do desenvolvimento e manutenção do tesouro; d) critérios para seleção de softwares para gestão de tesauros e e) modelo de dados e seu esquema XML derivado. A parte 2 inclui alguns outros tipos de SOC para efeitos das discussões sobre interoperabilidade, sem se ocupar dos aspectos de suas normatizações. Desse modo, aborda: sistemas de classificação bibliográfica, sistemas de classificação arquivística, taxonomias; listas de cabeçalhos de assunto, terminologias, ontologias, listas de autoridade, anéis de sinônimos. Relativamente a cada um deles apresenta suas principais características e antecedentes, sua tipologia, seus componentes e relações semânticas e os aspectos que dizem respeito ao mapeamento entre o tesouro e tipo de SOC em questão. A referida norma não menciona, em nenhum das suas duas partes, os topic maps.

A norma ISO 25964 representa a síntese de um processo evolutivo relativamente recente de normalização de vocabulários controlados. Assim, o documento sintetiza alguns aspectos basilares já presentes e normas anteriores e apresenta, em sintonia com as demandas dos ambientes informacionais digitais, alguns outros aspectos relevantes neste contexto.

Dentre as normas anteriores, a ISO 2788, publicada primeiramente em 1974 e revisada em segunda edição em 1986, tem como título: Guidelines for the establishment and development of monolingual thesauri (Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngues). Em 1985, a ISO publicou a norma de número 5964, complementar à ISO 2788, intitulada Guidelines for the establishment and development of multilingual thesauri (Diretrizes para o estabelecimento e desenvolvimento de tesauros multilíngues).

Essas normas foram adotadas como normas nacionais em diversos países. No Reino Unido foram publicadas como BS 5723 (em 1979 e em 1987) e BS 6723 (em 1985). Depois dessas normas, apenas os Estados Unidos continuaram a desenvolver separadamente uma norma própria para a construção e uso de tesouros, a ANSI/NISO Z39-19 – Guidelines for the construction, format, and management of monolingual controlled vocabularies (Diretrizes para a construção, formatação e gestão de vocabulários controlados monolíngues) (CLARKE, 2010), publicada em 2005 e atualizada em 2010, com uma seção contemplando a interoperabilidade e incluindo em sua cobertura, além de tesouros, outros tipos de vocabulários, como listas, anéis de sinônimos e taxonomias.

No Brasil, onde não se desenvolveu uma norma nacional, foram fundamentais as publicações de alguns manuais que derivaram das normas internacionais, tais como o “Manual de elaboração de tesouros monolíngues” (GOMES, 1990) e as “Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues” (AUSTIN; DALE, 1993). Trata-se, este último, de tradução da segunda edição do documento publicado pela Unesco, em 1981: Guidelines for the establishment and development of monolingual thesauri (Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues), revisada por Derek Austin e Peter Dale, com tradução brasileira, sob responsabilidade do IBICT, realizada por Bianca Amaro de Melo e revisada por Lígia Maria Café de Miranda.

As normas britânicas BS 5723 e BS 6723, que tratavam, respectivamente, de construção e gestão de tesouros monolíngues e de tesouros multilíngues, foram posteriormente fundidas na norma BS 8723, publicada em cinco partes entre os anos 2005 e 2008, a saber: parte 1: Definitions, symbols and abbreviations (Definições, símbolos e abreviações); parte 2: Thesauri (Tesouros); parte 3: Vocabularies other than thesauri (Vocabulários que não são tesouros); parte 4: Interoperability between vocabularies (Interoperabilidade entre vocabulários) e parte 5: Exchange formats and protocols for interoperability (Formatos de intercâmbio e protocolos para interoperabilidade).

A partir de 2007, a BS 8723 e a ANSI/NISO Z39-19 foram tomadas como base para a revisão das normas ISO 2788 e 5694, resultando na norma ISO 25964. A partir das análises das normas e de algumas

observações apontadas por Clarke e Zeng (2012), passase a indicar as principais alterações que a norma ISO 25964 apresenta em relação às normas anteriores. Na parte 1:

- a) ênfase na distinção entre termo e conceito e a prioridade ao conceito;
- b) fusão na mesma norma do tratamento de tesouros monolíngues e multilíngues, com privilégio aos primeiros;
- c) abordagem prioritária do tesauro em relação a outros tipos de SOC;
- d) ampliação do conjunto de definições dos termos utilizados na norma (glossário);
- e) respeito às diferenças culturais no tratamento das relações associativas, principalmente nos tesouros multilíngues;
- f) tratamento das relações de equivalência com observância das variações linguísticas em diferentes idiomas;
- g) inclusão de seção que trata da análise facetada e de sua importância para a construção de tesouros;
- h) possibilidade de especificação no uso das tags para indicação das relações hierárquicas, com possibilidade de distinção entre relações hierárquicas genéricas, partitivas ou de instanciação;
- i) indicações de critérios para seleção de softwares para gestão de tesouros.

Na parte 2:

- a) inclusão de modelos estruturais para mapeamento entre diferentes vocabulários;
- b) discussão sobre tipos de mapeamento: equivalência, hierárquicos, associativos;
- c) uso de mapeamentos para recuperação da informação;
- d) tratamento da pré-coordenação, gestão e exibição dos mapeamentos;

e) descrição geral das características de “outros vocabulários” (expressão presente no título da norma), incluindo para cada um: componentes semânticos e relações conceituais e relações de interoperabilidade e mapeamento com o tesouro.

A norma ISO 25964-2 torna explícita a relação entre a terminologia teórica e os tesouros, na seção de número 22 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013), ainda que, dada a natureza do documento, não a desenvolva de modo suficiente. As implicações entre a teoria e a metodologia da terminologia e os estudos sobre construção e manutenção de tesouros têm sido abordadas em diversos estudos, dentre os quais: Ruiz Pérez (1992), Barité Roqueta (2000), Cintra et al. (2001), Lara (2006, 2007, 2009), Broughton (2008) e Lima (2015), entre outros.

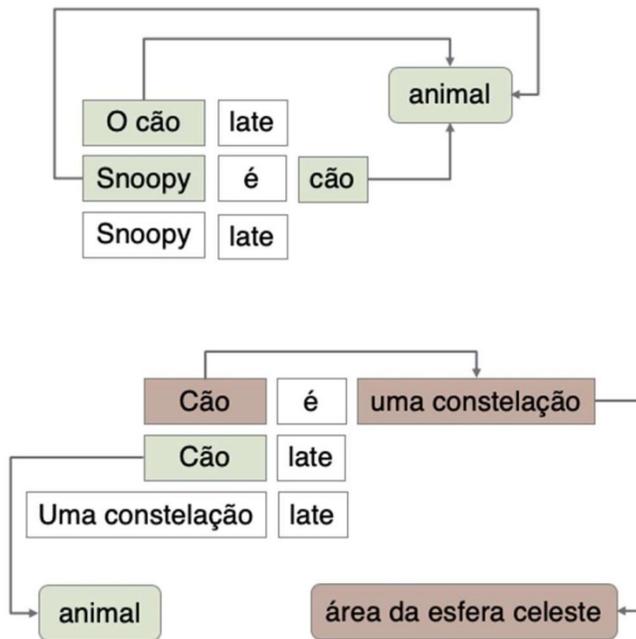
2.3.2 TERMINOLOGIA

Considerando-se as relações teóricas e metodológicas de proximidade que envolvem os tesouros e a terminologia, é fundamental conhecer as normas sobre terminologia. A norma ISO 704, publicada em 2009, e a norma ISO 1087 (partes 1 e 2), publicada em 2001, dispõem sobre o trabalho terminológico, seus princípios e métodos, esclarecendo sobre termos, conceitos e relações conceituais, bem como sobre os conceitos as funções da “definição” e da “designação”, além de descrever sobre suas funções no trabalho terminológico. Apesar da presença da palavra “métodos” no título da norma ISO 704 – Terminology work – Principles and methods (Trabalho terminológico – princípios e métodos), essa norma apresenta pouco conteúdo relativo ao sentido estrito que se atribui aos procedimentos de pesquisa terminológica. Para esse caos, é preciso consultar manuais e textos especializados como, entre outros, Arntz; Picht (1995), Aubert (2001), Pavel; Nolet (2002), Barros (2004) e Krieger; Finatto (2004).

2.4 TERMO, CONCEITO E RELAÇÕES CONCEITUAIS

Esta seção inicia-se com uma questão sutilmente capciosa: se o tesauro é um guia para ajudar o usuário a escolher o termo adequado para um dado conceito, qual é sua unidade básica: o termo ou o conceito? Para respondê-la, retoma-se, inicialmente, um exemplo simples e eloquente já utilizado em diversos manuais de introdução à lógica (Figura 2).

Figura 2 – Comparação de silogismos



Fonte: Elaborado pelo autor (2020).

Nesses casos, a falha lógica é quase óbvia ou mesmo risível para um leitor humano que mesmo sem conhecimento das sutilezas da lógica percebe alguma coisa fora do lugar. No caso apresentado na parte inferior da Figura 2, há, na verdade, dois termos: cão (animal) e cão (constelação) sendo tratados como se fossem apenas um, o que faz resultar na introdução de um quarto termo no silogismo. Nos casos de produção de inferência por raciocínio automático nos computadores, contudo, a máquina pode ser facilmente enganada se declarações sobre termos forem apresentadas

como se fossem declarações sobre conceitos ou como se as relações entre termos e conceitos fossem de denominação (MOREIRA, 2019b).

Retomando-se a questão inicialmente proposta nesta subseção, a referência ao “termo” em vez da referência aos “conceito” obedece apenas à praticidade, não se deve esquecer que a finalidade da manipulação de termos é subsidiar a recuperação dos conceitos que lhes estão subjacentes.

A ISO 25964 é mais rigorosa do que as normas sobre tesauro que lhe antecederam na distinção entre termo e conceito. A norma faz questão de enfatizar, por exemplo, que apesar de manter as tags já tradicionalmente conhecidas e amplamente utilizadas (TG, TE, TR), as relações que elas indicam ocorrem entre conceitos.

A indexação, enquanto operação de construção de representações do conteúdo temático do documento por meio de descritores selecionados com o recurso de tesouros (ou outros vocabulários controlados), apresenta dois pontos críticos: a fase de análise, quando se procede, mediante a adoção de políticas de indexação específicas, à identificação do assunto do documento, e a fase de tradução que, seguindo as mesmas políticas, auxilia o indexador a escolher os termos que melhor representam os conceitos identificados. Considerando-se a finalidade última de indexação, isto é, a recuperação da informação pelo usuário, é preciso acrescentar que o tesauro lhe será igualmente útil para a seleção dos termos que melhor representam sua questão de busca. Destaque-se, portanto, a função mediadora do tesauro atuando como código comutador entre os universos linguísticos envolvidos.

Isso pode fazer parecer que, ideal e hipoteticamente, seria mais adequado um tesauro para cada usuário. Ocorre, contudo, que ainda que a informação possua caráter subjetivo, essa subjetividade não ocorre em sentido fundamentalmente individual. Assim, como, já destacaram Capurro e Hjørland (2007, p. 192), “[o]s critérios sobre o que conta como informação são formulados por processos socioculturais e científicos”, não sendo possível por isso que os sistemas de informação consigam mapear todos os possíveis valores da informação.

Em vista disso, é preciso destacar que a seleção dos termos/conceitos que irão compor o tesauro e orientar sua política de manutenção

e desenvolvimento, bem como a definição da estrutura que irá revelar suas relações conceituais deve, necessariamente, considerar:

- a) o domínio ou espaço de informação em que o tesauro será aplicado, em busca da precisão terminológica;
- b) a garantia literária, que pressupõe a consulta a fontes de referência tais como dicionários, textos da literatura especializada e vocabulários de referência. A garantia literária visa a assegurar, tanto quanto possível, a correspondência dos conceitos ao uso predominante na literatura do domínio;
- c) a garantia institucional, para que seja possível identificar a forma preferida ou predominante do termo conforme seu uso pela instituição que irá abrigar e se responsabilizar pelo tesauro. No caso discutido neste capítulo, é desejável que sejam tomadas como bases, por exemplo, a legislação relativa à Unesp e o modo de distribuição e designação dos seus diversos cursos de graduação e de pós-graduação para os fins de construção das macrocategorias do tesauro;
- d) a garantia do usuário, isto é, a identificação, análise e incorporação – sempre de acordo com as políticas estabelecidas – das expressões utilizadas pelos usuários nas suas solicitações de busca. A análise de logs, que registra a interação do usuário com o sistema, e a localização das expressões de busca em outros vocabulários controlados afins são alguns dos instrumentos que podem ser utilizados.

2.5 O CONCEITO DE CATEGORIA, FACETA E TAXONOMIA APLICADOS AOS TESAuros

O modelo de organização dos objetos ou ideias em categorias que é utilizado como recurso natural na relação gnosiológica com o mundo, na tentativa de interpretá-lo e compreendê-lo, inspira o modelo de estrutura categorial utilizado na organização dos sistemas de organização do conhecimento.

O conceito de categoria é complexo e, ao mesmo tempo, simples. Categorias são noções abstratas, generalizantes, construídas para e por meio da identificação e agrupamento de conjuntos de regularidades. Categorias

são sempre noções generalizantes, abstratas, mesmo quando se referem a objetos concretos.

Sendo generalizantes, as categorias carregam o perigo inerente à toda e qualquer generalização, a qual, aliás, quando acompanhada da ignorância gera grandes problemas. Isso vale tanto para os debates que são travados no dia-a-dia sobre os mais diversos temas quanto para a sofisticada estrutura conceitual que subsidia os SOC. As generalizações que as categorias promovem, nestes casos, possuem caráter institucional, atendem a interesses específicos e possuem caráter instrumental como traço implícito (BARITÉ ROQUETA, 1999; TÁLAMO; LENZI, 2006).

Nessa perspectiva, categoria é “um conceito de alta generalidade e de larga aplicação que serve de estrutura a um esquema de classificação, pois sob este esquema se pode reunir outros conceitos” (ARANALDE, 2009, p. 99).

Devem-se principalmente a S. R. Ranganathan e Henry Bliss, na primeira metade do século XX, e aos trabalhos posteriores do Classification Research Group, a proposição e o desenvolvimento inicial da análise facetada. Este procedimento alterou profundamente o modo de pensar a teoria da classificação, até então inclinada para a adoção de perspectivas universalistas e assentada na concepção hierárquica rigorosa. A moderna teoria da classificação, mais preocupada com a organização de princípios lógicos em bases científicas e com o fornecimento de estruturas sustentáveis aos SOC. A abordagem facetada encontrou terreno fértil na nova ecologia informacional construída no pós-guerra e tornou-se o pilar sobre o qual assentou-se a teoria da classificação facetada que inspirou o desenvolvimento dos tesouros.

A ISO 25964, que dedica uma seção à análise facetada, define faceta como o “agrupamento dos conceitos da mesma categoria inerente”, e apresenta os seguintes exemplos: “‘animais’, ‘ratos’, ‘narcisos’ e ‘bactérias’ poderiam ser todos os membros de uma faceta ‘organismos vivos’; ‘cavar’, ‘escrever’ e ‘cozinhar’ poderiam ser todos membros de uma faceta ‘ações’; ‘Paris’, ‘Reino Unido’ e ‘Alpes’ poderiam ser todos os membros de uma faceta ‘lugares’” (ISO 25964-1, p. 4). Há também exemplos de categorias de alto nível, utilizadas para agrupar conceitos por meio de facetas, como, “objetos”, “materiais”, “agentes”, “ações”, “lugares” e “períodos”.

Desse modo, a análise facetada pode ser compreendida como a análise dos diversos assuntos em função dos conceitos que os constituem, os quais são agrupados em facetas que permitem a subdivisão dos conceitos a partir da identificação de características específicas.

Não é a aplicação da característica de divisão o elemento que irá distinguir a abordagem facetada. Essa característica, como elemento fundamental à aplicação dos princípios da teoria da classificação, já é uma preocupação dos sistemas de classificação mais “tradicionais”, como os sistemas enumerativos. Ocorre que a análise facetada requer mais rigor na aplicação de procedimentos de análise e síntese para a composição de classes e subclasses.

A partir da análise dos graus de generalização e especialização que caracterizam as operações de delimitação de categorias e facetas, Ranganathan procurou identificar quais seriam as noções gerais ou características mais comuns aos diversos grupos de conceitos. Assim, agregou aos conceitos mais comumente empregados de tempo e espaço os conceitos de personalidade, matéria e energia. Colocados esses elementos conforme a ordem preestabelecida por Ranganathan tem-se a conhecida fórmula do PMEST (personalidade, matéria, energia, espaço e tempo). Essas categorias foram posteriormente expandidas pelo Classification Research Group: coisa, tipo, parte (órgão, constituinte), propriedade, material, processo (uma ação interna ao item), operação (uma ação realizada no item), paciente (objeto da ação, matéria-prima), produto (substância), subproduto, agente, espaço, tempo.

A motivação de Ranganathan ao fazer a opção pela abordagem facetada advém da impossibilidade inerente aos sistemas de classificação enumerativos de lidarem com a diversidade de pontos de vista no tratamento dos assuntos dos documentos. Essa “deficiência” deve-se, basicamente, ao seu caráter com pretensões universalistas de inspiração positivista. Sobre essa limitação, o bibliotecário afirma:

[u]m esquema enumerativo com uma base superficial pode ser adequado e até econômico para um sistema de conhecimento fechado. Por exemplo, tal esquema funcionará bem para a filosofia da Grécia antiga ou da Índia, ambas já cristalizadas e fixadas há um longo tempo. [...] O que distingue o universo do conhecimento atual é que ele é um continuum dinâmico. É sempre crescente;

novos ramos podem brotar de qualquer de seus infinitos pontos a qualquer tempo; eles são desconhecidos no presente. Eles não podem, portanto, ser enumerados aqui e agora; não podem ser antecipados, suas filiações só podem ser determinadas depois que aparecerem (RANGANATHAN, 1951, p. 87, tradução livre).

Com apoio na síntese elaborada por Dahlberg (1979) podem-se apontar três dentre as maiores contribuições de Ranganathan à teoria da classificação:

- a) a distinção clara entre os elementos que subsidiam o trabalho com a classificação: um plano de ideias (conceito), um plano verbal (expressão) e um plano notacional (fixação dos conceitos em formas abstratas como sinais, letras e números);
- b) a abordagem analítico-sintética na identificação de assuntos, segundo as fórmulas de facetas;
- c) dezoito princípios para o arranjo dos elementos das facetas, denominados princípios para sequência útil.

A efetiva compreensão da importância da categorização para a organização de SOC como o tesouro deve remontar, naturalmente, à filosofia do conhecimento e retomar os trabalhos de Aristóteles e Kant, entre outros, sobre o assunto. Assim como Kant inspira-se declaradamente em Aristóteles, Ranganathan também o faz. Considerando-se a temática que diz respeito a este capítulo, relativamente à discussão sobre elementos de construção e manutenção de um determinado tesouro, é preciso salientar que é com Ranganathan que os estudos sobre categorias assumem de modo mais efetivo os contornos pragmáticos na sua interface com a organização do conhecimento.

Sobre a inspiração aristotélica, é o próprio Kant, como já mencionado também em trabalho anterior (MOREIRA, 2010), quem informa sobre seu desejo de rever Aristóteles: “segundo Aristóteles, denominaremos tais conceitos [puros de entendimento] categorias na medida em que nossa intenção, em princípio, identifica-se com a de

Aristóteles, se bem que se afaste bastante dele na execução” (KANT, 1996, p. 108). E, posteriormente, ainda no mesmo texto: “[...] a procura desses conceitos fundamentais constituiu um plano digno de homem perspicaz como Aristóteles. Entretanto, por não possuir nenhum princípio catou-os como se lhe deparavam, reunindo primeiramente dez, que denominou categorias (predicamentos). A seguir, creu ter encontrado ainda mais cinco conceitos que acrescentou sob a denominação de pós-predicamentos [...]” (KANT, 1996, p. 109).

2.6 RELAÇÕES TERMINOLÓGICAS E RELAÇÕES CONCEITUAIS

As relações entre os documentos, seus assuntos e seus metadados, bem como as relações recíprocas que cada um desses elementos – documentos, seus assuntos e seus metadados – estabelecem entre si inspiram, grosso modo, a construção de índices que possibilitem ao pesquisador navegar de modo mais preciso pelo imenso conjunto de informações exponencialmente disponibilizadas. Nenhum documento é uma ilha, como reforça Smiraglia (2002), e é justamente a complexidade da inter-relação dos documentos e seu conteúdo, bem como a complexidade dessas relações que se tornaram entraves à crescente sofisticação dos sistemas de recuperação online.

Uma das funções mais importantes do tesauro é sua capacidade de fornecer, a partir de um termo apresentado pelo usuário (entenda-se aqui o pesquisador, no momento da busca, ou o indexador, no momento da construção da representação documentária), outros termos que se lhe relacionam, ampliando desse modo suas possibilidades de busca. Imagine-se, por exemplo, a situação de um usuário com pouco conhecimento de inglês e de botânica que precisa localizar informações sobre “abies” para dar prosseguimento em sua pesquisa sobre costumes. Essa busca, intermediada por um tesauro como o Agrovoc², apresentaria as seguintes opções mostradas na Figura 1.

² O tesauro Agrovoc é coordenado pela Organização das Nações Unidas para a Alimentação e a Agricultura (FAO) e mantido por uma comunidade internacional de especialistas que atuam na área da agricultura e domínios relacionados. Atualmente, é composto por aproximadamente 36 mil conceitos disponíveis em até 35 idiomas. Fonte: <http://aims.fao.org/standards/agrovoc/concept-scheme> Acesso em: 19 fev. 2020.

Figura 3 – Exemplo com utilização do tesouro Agrovoc

... > Pinophyta > Pinopsida > Pinales > Pinaceae > Abies

PREFERRED TERM	Abies																
BROADER CONCEPT	Pinaceae (en)																
NARROWER CONCEPTS	Abies alba (en) Abies amabilis (en) Abies balsamea (en)																
ENTRY TERMS	firs (en) firsz (en)																
HAS TAXONOMIC RANK	genus (en)																
IS MEMBER OF	ornamental woody plants (en)																
IS USED AS	Christmas trees (en) timber trees (en)																
IN OTHER LANGUAGES	<table border="0"> <tr> <td> شوح</td> <td>Arabic</td> </tr> <tr> <td> 冷杉属</td> <td>Chinese</td> </tr> <tr> <td> 冷杉</td> <td></td> </tr> <tr> <td> 枞树</td> <td></td> </tr> <tr> <td> Abies</td> <td>Czech</td> </tr> <tr> <td> jedle (Abies)</td> <td></td> </tr> <tr> <td> jedle</td> <td></td> </tr> <tr> <td> Abies</td> <td>French</td> </tr> </table>	شوح	Arabic	冷杉属	Chinese	冷杉		枞树		Abies	Czech	jedle (Abies)		jedle		Abies	French
شوح	Arabic																
冷杉属	Chinese																
冷杉																	
枞树																	
Abies	Czech																
jedle (Abies)																	
jedle																	
Abies	French																

Fonte: <http://aims.fao.org/standards/agrovoc/functionalities/search>.

Antes de discutir a figura é preciso esclarecer: a reprodução da tela de busca do Agrovoc foi editada para melhor atendimento dos propósitos deste trabalho. A tela de resultados original apresenta 23 termos específicos (narrower concepts) e 46 opções na lista de idiomas (in other languages). Isso posto, observa-se que o conjunto de termos enumerados como termos específicos, se não ajuda a identificar o conceito relativo ao termo “abies”, certamente o amplia pela apresentação de uma série de conceitos que lhe são subordinados. Há outras duas tags igualmente relevantes. A tag “termos de entrada” (entry terms) informa ao usuário sobre termos que o sistema admite, mas que são preferencialmente substituídos pelo “termo preferido” (preferred term) “Abies”, ou seja, os “termos de entrada” atuam como remissivas deste para aquele. As tags “is member of” (é membro de) e “is used as” (é usado como) informam, mesmo ao usuário leigo no

assunto, sobre a natureza do conceito em tela, permitindo sua imediata compreensão por associação a elementos mais conhecidos. “Abies” são, informa o Agrovoc, componentes do conjunto de “plantas lenhosas ornamentais” (ornamental woody plants) e são usadas como “árvores de Natal” (Christmas trees).

As relações conceituais são extremamente importantes para a os SOC, incluindo-se, naturalmente, os tesouros, pois é por meio dessas relações que os conceitos adquirem sentido mais preciso em relação ao contexto de sua aplicação no tesauro, isto é, a diminuição (ou eliminação, em plano ideal) da ambiguidade pela observação pragmática do contexto de ocorrência do conceito. Havendo relações entre as características dos conceitos envolvidos num determinado contexto de ocorrência, haverá relações conceituais entre eles. Dahlberg (1978), uma das precursoras na abordagem da teoria do conceito, distingue cinco modalidades possíveis de relações conceituais, nomeadamente: relações lógicas (identidade, implicação, intersecção, disjunção e negação), relações hierárquicas (o gênero e a espécie), relações partitivas (o todo e suas partes), relações de oposição (contradição e contrariedade) e relações funcionais (conceitos implicados em um processo).

O modelo de relações que compõem o tesauro inclui, de modo mais comum, as relações hierárquicas (gênero-espécie e todo-partes) e associativas, além das relações de equivalência. Algumas dos tipos de relações associativas mais comuns estão enumerados no Quadro 2. Além dessas relações, admitem-se também outras relações denominadas como “relações personalizadas” pela ISO 25.964-1 e que visam a atender situações e contextos de recuperação muito específicos. Se, por um lado, as relações extras personalizadas trazem benefícios de ordem pragmática ao sistema, por outros, trazem também complicações extras, como, por exemplo, a dificuldade de promover ações de interoperabilidade com outros tesouros.

Quadro 2 – Exemplos de relações associativas

Descrição	Exemplo
ação / destinatário ou alvo	cultivo / colheita
ação / propriedade que lhe é associada	flexão / plasticidade
atividade / produto	tecelagem / roupas

atividades complementares	ensino / aprendizagem
causa / efeito (e vice-versa)	patógenos / doenças
coisa / atribuição	economia / nível de atividade econômica
coisa ou atividade / propriedades definidoras ou agentes	venenos / toxicidade
coisa, objeto ou processo / contra-agente	insetos / inseticidas
conceito / influência	política monetária / inflação
conceito / unidade de medida	corrente elétrica / ampere
disciplina ou campo de estudo / fenômeno estudado	silvicultura / florestas
matéria-prima / produto	caulim / porcelana
objeto / aplicação (e vice-versa)	abastecimento de água / irrigação
objeto / partes (quando não se qualifica como relação hierárquica)	instrumentos óticos / lentes
operação ou processo / agente ou instrumento	termostato / controle de temperatura
opostos	emprego / desemprego
peessoas ou coisas / origem	brasileiros / Brasil
termo composto / substantivo que é seu foco (quando não há relação hierárquica)	répteis / répteis fósseis

Fonte: Moreira (2018, p. 139).

Em termos simples e considerando-se apenas o modelo de relações utilizado pelos SOC, a realidade pode ser expressa pela equação: entidade <relação> entidade. A partir dessa equação é possível verificar: a) formação de entidades complexas pela combinação de entidades simples; b) a comparação ou o agrupamento de entidades e c) a ação de uma entidade sobre a outra. No âmbito da organização do conhecimento encontram-se as relações entre registros de conhecimento (representações documentárias), relações intra e intertextuais relativas aos documentos (incluindo citações e links) e relações conceituais que compõem a base estrutural dos SOC (GREEN, 2008; MOREIRA, 2019a).

A explicação de um conceito por meio da sua relação com outro conceito que permite a realização da inferência necessária ao conhecimento é bastante usual. A definição clássica aristotélica, amplamente utilizada em diversos dicionários, segue esse modelo. Tome-se como exemplo a definição de colírio apresentada no iDicionário Aulete³ “1. Med. Medicamento que se aplica nos olhos, mais especificamente na conjuntiva, para tratamento de doenças ou para aliviar irritações. 2. Bras. Fig. Pop. Pessoa atraente, bonita, agradável de se ver: Aquela modelo é um colírio”.

A organização das relações conceituais no tesauro pode refinar e dispensar, em alguns casos, o perspectivismo e as interpretações subjetivas que caracterizam a linguagem natural, pois buscam compreender o conceito e suas relações conforme se dão nos domínios e nas linguagens de especialidade.

2.7 CONSIDERAÇÕES FINAIS

A linguagem é um fenômeno extremamente complexo, pois desempenha múltiplas funções, como apontou Jakobson (2003), referindo-se à linguagem natural, quais sejam: referencial, emotiva, conativa, fática, poética e metalinguística. Nas linguagens documentárias, que visam a estabelecer um processo de “depuração” da linguagem natural para fins de comunicação em sistemas de informação documentária, apenas duas funções são efetivamente desempenhadas: a função referencial, com controle de vocabulário, e a função metalinguística.

Longe de simplificar completamente o processo, todavia, a adoção de instrumentos de vocabulário controlado, como o tesauro, ecoa problemas de significação já conhecidos e engendram novos problemas. Não se trata, bem entendido, de apontar uma disfunção dos tesauros, pois a matéria de que são feitos é a mesma que constitui a linguagem natural, isto é, não são sistemas diferentes. Os problemas de ordem semântica, pragmática e, em menor grau, sintática, estão todos presentes.

Construir um tesauro com a grandeza e a importância do Tesauro Unesp é, por essas e pelas diversas razões já apontadas neste livro, uma

³ Disponível em: <http://www.aulete.com.br/colirio>. Acesso em: 25 mar. 2020.

tarefa ao mesmo tempo hercúlea e sisífica, mas é também, ao mesmo tempo, para se manter na cultura helênica, dionisíaca.

A consciência da relevância do Tesouro Unesp e de sua indiscutível utilidade para toda a comunidade acadêmica, seja ela Unespiana ou não, e também para a Ciência da Informação, faz crer que o trabalho, que apenas começou, ainda irá frutificar fartamente.

REFERÊNCIAS

- ARANALDE, M. M. Reflexões sobre os sistemas categoriais de Aristóteles, Kant e Ranganathan. *Ciência da Informação*, Brasília, v. 38, n. 1, p. 86-108, jan./abr. 2009.
- ARNTZ, R.; PICHT, H. *Introducción a la terminología*. Madrid: Fundación Sánchez Ruipérez, 1995.
- AUBERT, F. H. *Introdução à metodologia da pesquisa terminológica bilingue*. São Paulo: FFLCH/CITRAT, 2001.
- AUSTIN, D.; DALE, P. *Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues*. Tradução de Bianca Amaro de Melo. Brasília: IBICT, 1993.
- BARITÉ ROQUETA, M. G. Los conceptos y su representación: una perspectiva terminológica para el tratamiento temático de la información. *Scire*, Zaragoza, v. 6, n. 1, p. 31-53, ene./jun. 2000.
- BARITÉ ROQUETA, M. G. La noción de categoría y sus implicancias en la construcción y evaluación de lenguajes documentales. In: LÓPEZ-HUERTAS PÉREZ, M. J.; FERNÁNDEZ-MOLINA, J. C. La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información: *actas del IV Congreso ISKO-EspañaEOCONSID'99*, 22-24 de abril de 1999, Granada. Universidad de Granada, 1999.
- BARROS, L. A. *Curso básico de terminología*. São Paulo: Edusp, 2004.
- BROUGHTON, V. A faceted classification as the basis of a faceted terminology: conversion of a classified structure to thesaurus format in the Bliss Bibliographic Classification, 2nd edition. *Axiomathes*, Dordrecht, n. 18, p. 193-210, 2008.
- BROUGHTON, V. *Essential thesaurus construction*. London: Facet, 2006.
- CAPURRO, R.; HJØRLAND, B. O conceito de informação. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 12, n. 1, p. 148-207, jan./abr. 2007.
- CINTRA, A. M. M. et al. Linguagens documentárias e terminologia. In: ALVES, I. M. (org.). *A constituição da normalização terminológica no Brasil*. São Paulo: FFLCH/CITRAT, 2001. p. 17-22.

- CINTRA, A. M. M. *et al.* *Para entender as linguagens documentárias*. 2.ed. São Paulo: Polis/APB, 2002.
- CLARKE, S. G. D. Thesaurus standards on a converging track. *Legal information management*, n. 10, p. 43-45, 2010.
- CLARKE, S. G. D.; ZENG, M. L. From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling. *Information standards quarterly*, Cambridge, v. 24, n. 1, p. 20-26, Winter 2012.
- COLMENERO RUIZ, M. J. Introducción al modelo topic maps (ISO/IEC 13250:2003). *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 3, n. 1, p. 77-102, jul./dez. 2005.
- DAHLBERG, I. Teoria da classificação ontem e hoje. Tradução: Henry B. Cox. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA, 1972, Rio de Janeiro. *Anais...* Brasília: IBICT/ABDE, 1979. v. 1, p. 352-370. [Palestra].
- DAHLBERG, I. Teoria do conceito. *Ciência da informação*, Brasília, v. 7, n. 2, p. 101-107, 1978.
- FRANCOIS, E. J. *Building global education with a local perspective: an introduction to global higher education*. London: Palgrave MacMillan, 2015.
- GARSHOL, L. M. Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all. *Journal of Information Science*, London, v. 30, n. 4, p. 378-391, 2004.
- GOMES, H. E. *Manual de elaboração de tesouros monolíngues*. Brasília: Programa Nacional de Bibliotecas de Ensino Superior, 1990.
- GREEN, R. Relationships in knowledge organization. *Knowledge Organization*, Baden-Baden, v. 35, n. 2/3, p. 150-159, 2008.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964: *Information and documentation: thesauri and interoperability with other vocabularies - part 2: interoperability with other vocabularies*. Genebra, 2013.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964: *Information and documentation: thesauri and interoperability with other vocabularies - part 1: thesauri for information retrieval*. Genebra, 2011.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 13520: *Topic maps – information technology – document description and processing languages*. Genebra, 2002.
- JAKOBSON, Roman. *Linguística e comunicação*. 19.ed. Tradução de Izidoro Blikstein e José Paulo Paes. São Paulo: Cultrix, 2003.
- KANT, I. *Crítica da razão pura*. São Paulo: Nova Cultural, 1996. (Os Pensadores).

- KOBASHI, N. Y. Linguística textual e elaboração de informações documentárias: algumas reflexões. In: GASPAR, N. R.; ROMÃO, L. M. S. (org.). *Discurso e texto: multiplicidade de sentidos na ciência da informação*. São Carlos: EdUFSCar, 2008. p. 47-66.
- KRIEGER, M. G.; FINATTO, M. J. B. *Introdução à terminologia: teoria e prática*. São Paulo: Contexto, 2004.
- LARA, M. L. G. Ciencias del lenguaje, terminología y ciencia de la información: relaciones interdisciplinarias y transdisciplinariedad. In: RODRÍGUEZ BRAVO, B.; ALVITE DÍEZ, M. L. (ed.). *La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico: actas del VIII Congreso ISKO - España, León, 18-20 de abril de 2007*. p. 101-109.
- LARA, M. L. G. *Linguística documentária: seleção de conceitos*. Tese (Livre Docência) – Universidade de São Paulo, São Paulo, 2009.
- LARA, M. L. G. Novas relações entre terminologia e ciência da informação na perspectiva de um conceito contemporâneo de informação. *DataGramaZero*, Rio de Janeiro, v. 7, n. 4, ago. 2006.
- LIMA, V. M. A. Da análise documentária à terminologia: percurso teórico e metodológico. In: GUIMARÃES, J. A. C.; DODEBEI, V. L. D. L. M. (org.). *Organização do conhecimento e diversidade cultural*. Marília: ISKO-Brasil, 2015. p. 758-764.
- McCULLOCH, E. Thesauri: practical guidance for construction. *Library review*, Bingley, v. 54, n. 7, p. 403-409, 2005.
- MOREIRA, W. *A construção de informações documentárias: aportes da linguística documentária, da terminologia e das ontologias*. Tese (Doutorado em Ciência da Informação) – Universidade de São Paulo, São Paulo, 2010.
- MOREIRA, W. Relações conceituais como elementos constitutivos essenciais dos sistemas de organização do conhecimento. *Informação & Informação*, Londrina, v. 24, n. 2, p. 1-30, maio/ago. 2019a.
- MOREIRA, Walter. *Sistemas de organização do conhecimento: aspectos teóricos, conceituais e metodológicos*. Tese (Livre-docência em Sistemas de Organização do Conhecimento) – Universidade Estadual Paulista, Marília, 2018.
- MOREIRA, Walter. Tesouros e ontologias como modelos de sistemas de organização do conhecimento. *Brazilian Journal of Information Science*, v. 13, n. 1, p. 15-20, 2019b.
- PAVEL, S.; NOLET, D. *Manual de terminologia*. Tradução de: Enilde Faulstich. Canadá: Departamento de Tradução do Governo Canadense, 2002.
- RANGANATHAN, S. R. *Philosophy of library classification*. Copenhagen: Ejnar Munksgaard, 1951.

RUIZ PÉREZ, R. *El análisis documental: bases terminológicas, conceptualización y estructura operativa*. Granada: Universidad de Granada, 1992.

SMIRAGLIA, R. P. The progress of theory in knowledge organization. *Library Trends*, Maryland, v. 50, n. 3, p. 330-349, 2002.

TÁLAMO, M. F. G. M. *Linguagem documentária*. São Paulo: Associação Paulista de Bibliotecários, 1997.

TÁLAMO, M. F. G. M.; LENZI, L. A. F. Terminologia e documentação: a relação solidária das organizações do conhecimento e da informação no domínio da inovação tecnológica. *DataGramaZero*, Rio de Janeiro, v. 7, n. 4, 2006.

ZENG, M. L. *Construction of controlled vocabularies: a primer*. 2005. Disponível em: <https://marciazeng.slis.kent.edu/Z3919/index.htm>. Acesso em: 08 abr. 2020.

3. TESAUROS: PRINCIPAIS ASPECTOS TEÓRICOS E METODOLÓGICOS PARA A SUA ELABORAÇÃO

Maria Luiza de Almeida Campos

3.1 Questões iniciais

3.2 Principais tipos de vocabulários e suas funções

3.3 Tesouros documentários: aspectos teóricos e metodológicos

3.3.1 Elementos de um tesouro

3.3.1.1 Termos

3.3.1.2 Relações conceituais entre Termos

3.3.1.2 Categorização de um domínio

3.4 Considerações Finais

Referências

3.1 QUESTÕES INICIAIS

É possível afirmar que a qualidade de um Sistema de Recuperação de Informação (SRI) se dá na observância de duas questões, já muito discutidas por Lancaster (1979, 1993), ou seja: a linguagem utilizada pelo indexador e a linguagem utilizada pelo usuário. Quando se opta pela linguagem natural pode-se ter duas formulas distintas o que traz problemas na recuperação em um SRI. O uso de uma linguagem intermediária, que serve como um léxico conceitual que permite formular todos os assuntos

possíveis de um dado documento e assegurar correspondências entre a demanda e a oferta de informação, é o que se recomenda e que se configura como um Vocabulário Controlado.

Vocabulário Controlado é uma ferramenta terminológica que possibilita determinar o melhor termo que deve ser usado na representação de informações contidas nos documentos e também possibilita ao usuário uma maior precisão na construção de uma estratégia de busca. Tesouros são considerados vocabulários controlados.

Quanto ao uso de vocabulários em SRIs é importante ressaltar a natureza semântica da comunicação ali estabelecida, o que requer, em certa medida, levantar questões que envolvem a problemática terminológica em domínios de natureza diferentes.

Domínios Descritivos (como a Física, Biologia, Química e assemelhados) possuem uma maior utilização de termos técnicos, onde é possível estabelecer de forma mais consistente a Univocidade, ou seja, que possa ser possível a existência de uma palavra para um único significado, sempre em um dado contexto, determinando o termo apropriado. Em Domínios Normativos (como a Ciência Social, Filosofia e assemelhados) existe uma maior utilização de palavras da linguagem natural com atribuição de sentido específico. A criação de um novo termo para expressar um dado significado é pouco utilizado, na maioria das vezes, o que se dá é uso por empréstimo de uma determinada palavra com um novo significado dentro daquele contexto. Além disso, ocorre também, nesses domínios, a existência de várias escolas de pensamento, onde um mesmo termo pode ser entendido de forma diferente, conforme a escola de pensamento adotada.

Estas questões são consubstanciadas por Dalhlberg (DAHLBERG, 1978) quando ressalta que toda univocidade é relativa, ela é determinada no escopo de um domínio específico e só vale naquele dado domínio. Os ontologistas atualmente denominam esta questão apontada por Dahlberg, de comprometimento ontológico, o conteúdo conceitual de um termo é determinado a partir do que ele pode representar em um determinado domínio. Toda essa problemática semântica deve ser observada pelo profissional de informação quando do trabalho com as questões que envolvem o controle de vocabulário.

Desta forma, este capítulo tem por objetivo apresentar principais aspectos que envolvem a elaboração de vocabulários controlados, questões que passam por discutir aspectos teóricos e metodológicos. Neste sentido, visando apontar mais especificamente os Tesouros, discutiremos a seguir os principais tipos de vocabulários e suas funções e diferenças, para depois especificamente apresentarmos os aspectos teóricos e metodológicos fundamentais para a elaboração de tesouros documentários.

3.2 PRINCIPAIS TIPOS DE VOCABULÁRIOS E SUAS FUNÇÕES EM UM SRI

Abordar na atualidade o papel dos vocabulários controlados e suas funções em um SRI passa por discutirmos a função de tais vocabulários em contextos relacionados à Web e suas especificidades. Nesse sentido é importante o entendimento do que se denomina por Web sintática e Web semântica e como a tecnologia existente em cada época interferiu nos instrumentos de representação e recuperação da informação/conhecimento. Neste sentido em estudos realizados (CAMPOS, 2010, p.223) foi possível considerarmos que

O conceito de Web Semântica é cunhado no âmbito da Ciência da Computação, no contexto de viabilizar a semântica para que agentes inteligentes não humanos possam interpretar dados. [...] Este conceito, em contraposição ao conceito de Web Sintática, esta última uma rede desenvolvida para a produção de sentido realizada por humanos, traz em seu bojo a questão da produção de significado. Até o surgimento da Web Semântica a linha divisória entre documento e instrumento de tratamento estava bem definida, e podemos citar tesouros e taxonomias como exemplo disso. Os tesouros explicitam os conceitos de um domínio, assim como as taxonomias, e permitem que se tenha um quadro de conceitos e relações a partir de um dado contexto de uso. [...] Entretanto, não possuem mecanismos para a derivação do conhecimento do domínio. Nas ontologias, mais precisamente nas ontologias formais, o conhecimento pode ser obtido através de inferências sobre o conhecimento imediato disponível.

Neste sentido, na atualidade se comparamos tesouros, taxonomias e ontologias, podemos afirmar que eles possuem informações de natureza

distintas, mas que guardam semelhanças importantes. Quanto à questão terminológica os três instrumentos possuem um conjunto básico de conceitos e relações. Tanto os tesouros quanto as ontologias explicitam a representação do domínio através de taxonomias, que são as estruturas hierárquicas das relações apresentadas entre os conceitos, ou seja, as relações de gênero e espécie e partitivas. As ontologias além de possuírem relações associativas/funcionais mais explicitadas que os tesouros, apresentam todas as relações explicitadas para a leitura não só por humanos, mas também pelas máquinas. Esta possibilidade se dá através de assertivas que podem ser definidas, como um conjunto de axiomas aplicados aos conceitos e relações. Por exemplo: A definição do conceito seria escrita para “compreensão” da máquina, ou seja, $\text{solteiro}(x) = \text{homem}(x) \wedge \neg \text{casado}(x)$ em uma linguagem lógica. Neste sentido, ontologias funcionam como tesouros e taxonomias, para a representação e recuperação do conhecimento e da informação, mas além disto, possibilitam também a descoberta do conhecimento sobre o domínio devido ao uso de axiomas para explicitar as definições dos conceitos.

Assim, é possível afirmar que os três instrumentos hoje existente na Web, sendo ela sintática ou semântica, podem ser definidos como ferramentas utilizadas para organização e recuperação de informação que possuem uma base classificatória e são considerados sistemas de conceitos. Neste sentido, considera-se importante apresentarmos os principais aspectos teóricos e metodológicos para a elaboração de sistemas de conceitos., enfatizando os aspectos para a elaboração de tesouros.

3.3 TESAUROS DOCUMENTÁRIOS: ASPECTOS TEÓRICOS E METODOLÓGICOS

Tesouros são instrumentos de controle terminológico que garantem consistência no tratamento do conteúdo de documentos e permite traduzir a linguagem natural dos documentos, dos usuários e dos indexadores, numa linguagem controlada de recuperação da informação. Auxiliando assim, indexadores e usuários em suas respectivas ações. Possui termos relacionados semântica e genericamente de modo a formar uma rede de conceitos, onde tal rede se apresenta de forma sistemática através de uma taxonomia.

Na década de 70 do século passado a UNESCO (1973, p. 6) definiu tesauro sob dois aspectos:

- Segundo a Estrutura - “É um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento”.
- Segundo a Função - “É um dispositivo de controle terminológico usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários numa ‘linguagem do sistema’ (linguagem de documentação, linguagem de informação) mais restrita.”

Atualmente, a ISO 25.964-1 (INTERNATIONAL STANDARD ORGANIZATION, 2011, p.18, tradução nossa) define tesauro como

Vocabulário controlado e estruturado no qual conceitos são representados por termos, organizados para que os relacionamentos entre conceitos sejam explicitados, e os termos preferidos sejam acompanhados por termos de entrada para sinônimos ou quase sinônimos.

Tesauros com base em conceitos, também denominados de Tesauros terminológicos (CAMPOS, 2001), apresentam uma parte sistemática e uma parte alfabética. Na parte sistemática, os conceitos são como uma taxonomia, que permite ter uma visão abrangente do domínio que se está representando, e uma parte alfabética onde cada conceito é apresentado com suas relações e notas de escopo (notas de definição e de indexação).

É importante ressaltar que para o entendimento do Tesauro como uma ferramenta de tratamento e recuperação de informações não podemos deixar de citar Lancaster e o seu clássico livro *Indexação e Resumos* (LANCASTER, 1993) onde ele apresenta as principais funções desenvolvidas em muitos tipos de serviços de informação e os três subsistemas que fazem parte de um SRI, ou seja, o subsistema de indexação, o subsistema de controle de vocabulário e o subsistema de recuperação. Evidenciando a necessidade desses 3 subsistemas estarem

bem harmonizados, pois se isto não ocorre teremos diversos problemas para atingir a precisão nos processos de tratamento e recuperação de documentos.

Afirma Lancaster (1993), que o subsistema de controle de vocabulário afeta os dois subsistemas de indexação e recuperação e é afetado por eles. Neste sentido, se a análise não for consistente, os termos escolhidos para representar o conteúdo da informação de um documento, podem não se apresentarem de forma precisa e conseqüentemente afetarem o subsistema de recuperação causando recuperações falsas.

Por outro lado, muitas vezes a análise do conteúdo informativo do documento é bem realizada, mas a forma de nomear o termo causa duplo sentido o que afeta também a recuperação. Estes procedimentos estão fora da elaboração de um tesouro, mas afetam diretamente a sua consistência terminológica.

Assim, na Indexação de Documentos o tesouro tem a função de ser uma ferramenta terminológica que possibilita determinar o melhor termo que deve ser usado na representação de informações contidas nos documentos e, na preparação da estratégia de busca, o tesouro possibilita ao usuário que busca uma determinada informação estabelecer uma estratégia mais precisa através de sua parte sistemática ou alfabética. Através da parte sistemática é possível identificar o que existe de conteúdo informativo de documentos em uma base de dados e funciona como uma Taxonomia Navegacional, que pode ser definida como uma representação Gráfica de um Domínio através de uma Estrutura de Conceitos, que possui mecanismos automáticos que possibilitam percorrer pelas classes e subclasses desta Estrutura de Conceitos. Através da parte alfabética os termos são apresentados em Ordem Alfabética contendo Relações e Notas Explicativas.

Com a automatização, os SRIs hoje possuem esses 3 subsistemas integrados e os tesouros são concebidos e alimentados através de softwares que permitem a recuperação automática através dos subsistemas de entrada e recuperação.

Tesouros podem ser classificados sob diversos aspectos. Quanto à língua podemos dizer que podem ser monolíngues ou multilíngues ; quanto ao nível de especificidade de seus termos – macrotresouros e

microtesauros ; quanto ao assunto que cobrem - os tesauros podem ser voltados para uma missão ou problema, ou dedicados a um tema/área de conhecimento. Quanto a este último aspecto é importante ressaltar que Tesauros cobrem sempre um dado domínio de conhecimento. Entretanto este domínio pode por vezes não ser um domínio canônico onde o conhecimento é mais padronizado como a Engenharia, a Química, por exemplo. Mas podem cobrir um domínio de missão ou problema onde é possível reunir diversos aspectos conceituais, onde o entendimento se dá como etapas de um fazer reunindo os diversos processos e agentes para a realização de tarefas, como domínios da Telecomunicação, Exploração e Produção de Petróleo, Entretenimento, Agroecologia entre outros. Este aspecto irá influenciar bastante na organização das classes e nas relações conceituais que se irá estabelecer.

3.3.1 ELEMENTOS DE UM TESAURO

Tesauros estruturam conceitos. Aqui vamos nos apoiar em princípios estabelecidos por I. Dahlberg e em sua “Teoria do Conceito”. (DAHLBERG, 1978, 1978a, 1978b). Referentes, características e termos são elementos formadores de um conceito segundo Dahlberg. Onde conceito é definido como uma unidade do conhecimento, compreendendo afirmações verdadeiras (características) sobre um dado item de referência (referente), representado numa forma verbal (termo). Os referentes representam objetos que podem ser concretos ou abstratos, reais ou imaginários. Quando são utilizados para identificar uma instância no mundo concreto ou imaginário são denominados objetos individuais. E quando são utilizados para representar classes de objetos são chamados de objetos gerais. No processo de análise dos atributos dos objetos, em um grau de abstração, são as características do conceito. Assim, os conceitos são de dois tipos: conceito individuais, nomes próprios, também na atualidade chamados de instâncias; e conceito geral que podem se referir a objetos naturais ou artificiais, concretos ou abstratos, e para serem comunicados são designados por um termo – uma palavra ou uma expressão verbal, um símbolo ou um código.

O triângulo conceitual (referente, característica e termos) apresentado por Dahlberg (1978, 1978a, 1978b) são determinados a partir

de um dado Universo de Conhecimento e toda a seleção de características estão condicionados a este contexto específico. Devido a isto Dahlbeg afirma que toda conceituação possui uma Univocidade Relativa, depende sempre de um processo de predicação sobre o referente, onde tal predicação está refém do contexto em que ele está inserido. Desta forma, ao se selecionar característica sobre um dado referente é importante estabelecer em que Universo de Conhecimento tal referente está inserido. Assim, também os termos, como elementos de comunicação, precisam ser determinados visando representar da melhor forma possível o conteúdo conceitual em um dado contexto. Neste sentido, que Dahlberg afirma que o conceito é uma unidade de conhecimento, porque representa o compromisso assumido (acordo ontológico) entre um grupo de falantes em um dado Universo de Conhecimento. Devido a isto, a forma pela qual observamos o referente em um dado Universo de Conhecimento irá determinar as suas características, e conseqüentemente sua definição.

As definições devem ser estudadas, como um método para o entendimento do conceito e para o seu posicionamento em um sistema de conceitos de um dado contexto. Fazer uma definição equivale a estabelecer uma “equação de sentido” sendo que de um lado (à esquerda) encontramos aquilo que deve ser definido (o definiendum) e de outro (à direita) aquilo pelo qual alguma coisa é definida (o definiens) (DAHLBERG, 1978).

Assim, as definições não devem ser recolhidas e sim elaboradas para representarem a forma como os termos estão classificados em um Universo de Conhecimento.

Na elaboração de Tesouros os termos, as relações entre os termos e a categorização do domínio para apresentação dos termos e de sua estrutura conceitual são aspectos importantes a serem observados.

A seguir vamos discorrer sobre cada um desses aspectos e apresentar princípios que consideramos importantes atentar.

3.3.1.1 TERMOS

Termos são a menor unidade de representação em um tesouro, e como tal são indivisíveis na indexação. Podem ser formados por uma

palavra ou grupo de palavras. Neste sentido, na atualidade, princípios de fatoração se tornaram dispensáveis.

No âmbito de um conceito, os termos denotam um referente segundo Dahlberg (1978). Neste sentido guardam com o referente uma relação intrínseca. Os referentes representam objetos gerais e objetos individuais de um dado Universo de Conhecimento. Os objetos gerais podem ser de duas naturezas, objetos percebidos (árvore, homem, máquina, ponte) ou objetos abstratos (sentimentos, processos). Os objetos individuais são os nomes próprios, como por exemplo “Universidade Estadual Paulista”. Tais objetos individuais historicamente não faziam parte do tesauro, eram consideradas listas de autoridade, mas recentemente eles podem ser introduzidos em Tesouros a partir de relações de instâncias, como veremos adiante.

Como unidades de comunicação, os termos, estão refém de alguns fenômenos da língua que devem ser observados para não ocasionar erros na classificação desses termos. Os mais comuns são:

Sinônimo, palavras que têm significados idênticos ou muito semelhantes em um dado Universo de conhecimento. Quando este fenômeno ocorre, em tesouros, escolhe-se uma forma verbal em detrimento a outra, e estabelece-se uma relação de equivalência (USE/UP);

Quase sinonímia, identifica-se quando dois conceitos têm praticamente a mesma intensão. Entendendo-se por intensão de um conceito o conjunto de características atribuídos a um dado referente. Quando isto ocorre, pode-se tomar uma das seguintes decisões: os termos podem ser considerados sinônimos ou podem ser considerados como termos independentes e incluídos ambos os termos no Tesauro, através da relação associativa. Por exemplo: os termos Bar e Botequim possuem, em muitos aspectos características semelhantes, mas não são completamente idênticos. Assim, em dado contexto se forem considerados como quase sinônimos, devem ser mantido os dois termos e se estabelecer uma relação associativa entre Bar e Botequim.

Homonímia - palavras que possuem a mesma grafia ou a mesma pronúncia, mas com significados diferentes entre si. Neste caso, em Tesouros usa-se um qualificador para especificar o sentido. Por exemplo: Indexação (economia); Indexação (biblioteconomia);

Metáfora - palavra ou uma expressão com um sentido figurado por sua semelhança de significado. Deve-se atentar para as metáforas, pois elas podem acarretar muitos erros na classificação dos termos. Por exemplo: se em determinado Universo de Conhecimento existe o termo 'leite' sendo definido como produto de origem animal não será correto subordinarmos 'leite de soja' ao termo 'leite', pois 'leite de soja', neste contexto, é considerado uma metáfora. Usa-se a palavra leite no sentido figurado, pela aparência e textura do produto, não sendo de origem animal.

Outra questão relativa ao termo diz respeito a forma plural ou singular do termo. Para Dahlberg, os termos, como representantes de classes de conceitos, devem ser grafados no singular. Somente em casos do termo ter sido cunhado e dicionarizado com a desinência de plural, deve-se adotar a forma no plural, por exemplo: Ciências Sociais; Custas; Óculos.

3.3.1.2 RELAÇÕES CONCEITUAIS ENTRE TERMOS

Os termos, como a unidade de representação do conceito, em um Tesouro, se relacionam uns em relação aos outros, formando um sistema de conceitos. As relações entre os termos em um tesouro são de vários tipos e podem ser agrupados a partir das seguintes relações: relação hierárquica; relação partitiva; relação associativa; relação de equivalência; relação de instância

A Relação hierárquica ocorre entre termos de mesma natureza acarretando a relação de gênero/espécie. Nos tesouros, este tipo de relação se apresenta como Termo Geral/Termo Específico representado pelos símbolos TG/TE ou BT/NT.

A Relação partitiva pode referir, entre outros, a partes de um objeto, elementos de um sistema, etapas de um processo ou atividade. Nos tesouros, este tipo de relação partitiva se apresenta como Termo Geral Partitivo/Termo Específico Partitivo representado pelos símbolos TGP/TEP.

A Relação Associativa ocorre entre objetos no tempo, no espaço. As mais frequentes Causa/Efeito; Instrumento/Processo associado; Relação de descendência genealógica (pai/filho). Essas relações associativas através do símbolo TA (Termo associado) ou TR (Termo relacionado).

A Relação de equivalência ocorre no plano da língua, diferente das relações anteriores que ocorrem no plano dos conceitos. Neste sentido, as relações de equivalência ocorrem entre dois objetos com as mesmas características e que possuem formas de designação diferentes, neste sentido, os termos são considerados sinônimos. Mas a relação também pode ser usada para termos quase-sinônimos, quando a política de indexação é genérica, ou seja, a decisão é estabelecida pelos indexadores/pesquisadores. No tesauro, a relação de equivalência é representado pelos símbolos USE (relação usada do termo não preferido para o termo preferido) e UP (relação usada do termo preferido para o não preferido) (GOMES; CAMPOS, 2019).

A Relação de instância ocorre entre um objeto geral e um objeto individual (ver item 3.1), caracteriza os objetos individuais (nomes próprios, instâncias) que estão em relação a um dado objeto geral, como apresentado anteriormente, esta relação se apresenta na atualidade em Tesauros e é descrita na Norma ISO 25964-1:2011¹ e é representada pelo símbolo TGI (Termos genérico de instância) e TEI (termo específico de instância)

3.3.1.3 CATEGORIZAÇÃO DE UM DOMÍNIO

Domínios são Universos de Conhecimento. Segundo Ranganathan (1967) nos Universos de Conhecimento é onde encontra-se a totalidade dos saberes conservados pela humanidade. Assim, cada Universo de Conhecimento pode ser constituído por um campo de conhecimento disciplinar, interdisciplinar ou mesmo transdisciplinar. Este último podendo ser considerado um campo de missão ou problema. Tesauros apresentam os conceitos e suas relações no interior de Universos de Conhecimento.

No âmbito das estratégias utilizadas para a categorização de domínios, em uma primeira ação, é necessário a tomada de decisão quanto ao primeiro corte classificatório utilizado. Este corte classificatório pode ter como princípio os seguintes recortes: recorte por disciplinas, recorte

¹ Norma ISO 25964-1:2011 Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval (rev. 2017).

pela extensão das categorias de um campo de assunto e o recorte pela estrutura organizacional.

No recorte por Disciplinas, o domínio é classificado por áreas canônicas, ou seja, por disciplinas e subdisciplinas. Esta é a divisão clássica apresentada na maioria das estruturas classificadas no âmbito da biblioteconomia, fundamentalmente nos SOCs para a organização de bibliotecas e serviços de informação.

No recorte pela extensão das Categorias de um campo de assunto é realizado a partir da determinação de categorias selecionadas deste campo, no interior do qual se determina facetas e seus aspectos relacionados. As categorias propostas por Ranganathan – Personalidade, Energia (Processos), Matéria, Espaço e Tempo (RANGANATHAN, 1967) são um bom exemplo para isso. Neste sentido, estando em um dado campo de assunto, como por exemplo, o Folclore e Cultura Popular Brasileira, temos como facetas os Artefatos e Alimento que são manifestações da Categoria Personalidade (Tesouro do Folclore e Cultura Popular²). Vickery (1966) membro do CRG – Grupo de Classificação da Inglaterra, na década de 60 do século passado, apresenta um desdobramento destas categorias, visando trabalhar em Universos de Conhecimento especializados, produzindo uma lista de mais fácil compreensão, a saber: a) substâncias, produto, organismo; b) parte, órgão, estrutura; c) constituinte; d) propriedade e medida; e) objeto de ação, matéria prima; f) ação, operação, processo, comportamento; g) agente, instrumento; h) propriedade geral, processo, operação; i) espaço e j) tempo. Dahlberg (2014), apresenta uma outra sistematização para as categorias: entidades (princípios, objetos materiais e objetos imateriais); qualidades (qualidade, quantidade, relação); atividades (ação, processo, estado); dimensões (lugar, tempo, lugar no espaço). Afirma que as categorias são conceitos mais abrangentes acima das quais não se pode colocar qualquer conceito.

No Recorte pela Estrutura Organizacional, o domínio é classificado por setores ou funções de uma Organização. Neste sentido, o conhecimento é organizado pelas atividades administrativas e operacionais de uma organização. O recorte atende ao propósito da instituição. Está refém de mudanças de ordem gerencial e administrativa e na maioria das vezes não atende a um determinado princípio lógico de organização de

² Site: <http://www.cnfcp.gov.br/tesouro/>.

conceitos em dado domínio. Devido a esses fatores, muitos teóricos não consideram uma taxonomia.

No interior de tais recortes, os termos são organizados em classes de conceitos a partir de características de divisão .

Características de divisão são os princípios utilizados para a formação de classes, por exemplo: podemos classificar livros pelas seguintes características de divisão: tamanho, forma, conteúdo. Tais classes são formadas por cadeias e renques.

Cadeia “é uma sequência de classes formada por uma classe e seu universo de deslocamento 1, 2, 3 etc. até um ponto desejado” (RANGANATHAN, 1967, p. 61), ou seja, “são séries verticais de conceitos em que cada conceito tem uma característica a mais ou a menos conforme a cadeia seja descendente ou ascendente” (CAMPOS, 2001). Renque são classes derivadas de um Universo com base em uma única característica em qualquer passo de divisão até sua reunião completa em uma sequência preferida (RANGANATHAN, 1967, p.61), ou seja, “são classes formadas a partir de uma única característica de divisão, formando séries horizontais.” (CAMPOS, 2001, p. 51). Cadeias e Renques podem ser genéricas ou partitivas. Genéricas quando o princípio de sua elaboração se dá entre um conceito de gênero e outro de uma espécie e partitiva entre conceitos que representam o todo e sua parte, ou mesmo etapas de processos. Ranganathan (1967) apresenta uma série de princípios (cânones) para garantir a consistência na formação de renques e cadeias (cânones para sucessão e características, cânones para a formação de renques e cânones para a formação de cadeias).

É importante ressaltar que toda estrutura classificatória para ter consistência deve ser elaborada como um Sistema de Conceitos, o que se dá através da organização das classes de conceitos, utilizando características de divisão e reunindo as classes a partir da escolha de um princípio. Este princípio é o que se denomina por Base Classificatória (disciplina, categorias/facetas, estrutura organizacional). Deve-se sempre deixar explícito os princípios adotados para a elaboração de Tesouros, para que novos conceitos possam ser inseridos. Obedecendo ao que Ranganathan (1967) denomina por hospitalidade na classificação, possibilitando também, manter a consistência lógica.

3.4 CONSIDERAÇÕES FINAIS

Por fim é necessário ressaltar que um Tesouro, como todo instrumento terminológico de tratamento e recuperação de informação e que tem por finalidade a representar Universos de conhecimento, é dinâmico, e uma vez organizado necessitará de manutenção e atualização. Desta forma, é importante que se estabeleça sempre um cronograma de revisões para a avaliação das propostas de inclusão, retirada e modificação de termos e relações e que se estabeleça também, um padrão para que a Equipe responsável pela atualização possa apresentar tais propostas.

Princípios, normas e metodologias devem ser sempre utilizados para permitir acompanhar o desenvolvimento dos assuntos em um Universo de Conhecimento, que como nos diz Rangantahan, é dinâmico e com um crescimento constante. Nunca devem ser amarras a este desenvolvimento.

REFERÊNCIAS

- CAMPOS, M. L. A. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: EdUFF, 2001.
- CAMPOS, M. L. A. O documento e as ferramentas de tratamento e recuperação de informações na Web Semântica: um novo espaço de identidade. In: FREITAS, Lídia Silva; MARCONDES, Carlos Henrique; RODRIGUES, Ana Célia (org.). *Documento: Gênese e contextos de uso*. Niterói: Eduff, 2010. (Estudos da informação, v. 1). p. 223-234.
- DAHLBERG, I. A referent-oriented, analytical concept theory of Interconcept. *International Classification*, München, v. 5, n. 3, p. 122-151, 1978.
- DAHLBERG, I. *Ontical structures and universal classification*. Bangalore: Sarada Ranganathan Endowment for Library Science, 1978b.
- DAHLBERG, I. Teoria do conceito. *Ciência da Informação*, Brasília, v. 7, n. 2, p. 101-07, 1978a.
- DAHLBERG, I. *Wissensorganisation: entwicklung, aufgabe, anwendung, zukunft*. Würzburg: Ergon, 2014.
- GOMES, H. E.; CAMPOS, M. L. de A. *A organização do conhecimento na web: contribuições de Shiyali Ramamrita Ranganathan e Ingetraut Dahlberg*. Niterói: IACS/UFF, 2019. (Grupo de pesquisa Estudos Ônticos e Ontológicos em Contextos Informacionais: representação, recuperação e métricas) (Cadernos Acadêmicos, n. 1).

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/ FDIS25964-1: Information and documentation: *Thesauri and interoperability with other vocabularies* – Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization. 2011.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos/ Livros, 1993.

LANCASTER, F. W. The functions of information retrieval systems. *In*:LANCASTER, F. W. *Information retrieval systems*. New York: Wiley Interscience, 1979. p. 1-14.

RANGANATHAN, S. R. *Prolegomena to Library Classification*. 3. ed. New York: Asia Publishing House, 1967.

UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris, 1973.

VICKERY, B. C. *Faceted classification schemes*. New Brunswick, New Jersey: Graduate School of Library Service, Rutgers the State University, 1966.

PARTE 2

DIRETRIZES PARA O PLANEJAMENTO, CONSTRUÇÃO E MANUTENÇÃO DO TESAURO UNESP

4. TERMOS DO TESAURO UNESP: ASPECTOS CONCEITUAIS E FORMAIS DOS TERMOS, RELAÇÕES ENTRE TERMOS E RELAÇÕES ENTRE CONCEITOS

Lucia Silva Parra

Milena Maria Rodrigues Pedrozo

Rosane Rodrigues de Barros Ribas

Rosimara Lofego

Telma Jaqueline Dias Silveira

- 4.1 Aspectos conceituais do Tesouro Unesp
- 4.2 Aspectos formais do Tesouro Unesp
 - 4.2.1 Forma gramatical dos termos
 - 4.2.1.1 Substantivos e frases nominais
 - 4.2.1.2 Adjetivos
 - 4.2.1.3 Advérbios
 - 4.2.1.4 Verbos
 - 4.2.1.5 Artigos iniciais
 - 4.2.1.5.1 Omissão
 - 4.2.1.5.2 Retenção
 - 4.2.1.6 Capitalização, pontuação e caracteres especiais

- 4.2.1.6.1 Capitalização
- 4.2.1.6.2 Caracteres não alfabéticos
- 4.2.1.7 Formas singulares ou plurais
 - 4.2.1.7.1 Fatores culturais e linguísticos
 - 4.2.1.7.2 Tratamento de nomes contáveis
 - 4.2.1.7.3 Tratamento de substantivos não contáveis
 - 4.2.1.7.4 Coexistência do singular e plural
- 4.2.2 Clarificação e desambiguação dos termos do tesouro
 - 4.2.2.1 Homógrafos e qualificadores
 - 4.2.2.2 Notas
 - 4.2.2.2.1 Notas de escopo
 - 4.2.2.2.2 Definições
 - 4.2.2.2.3 Notas históricas
 - 4.2.2.3 Seleção da forma preferida
 - 4.2.2.3.1 Ortografia
 - 4.2.2.3.2 Termos de empréstimo e traduções de termos de empréstimos (estrangeirismo)
 - 4.2.2.3.3 Transliteração
 - 4.2.2.3.4 Neologismos, termos de gírias e jargões
 - 4.2.2.3.5 Nomes comuns e nomes comerciais
 - 4.2.2.3.6 Nomes populares e nomes científicos
 - 4.2.2.3.7 Abreviações e siglas
 - 4.2.2.3.8 Nomes próprios, nomes de lugares e nomes de instituições
- 4.2.3 Conceitos complexos
 - 4.2.3.1 A natureza dos termos compostos
 - 4.2.3.2 Decidir se deve ou não admitir um conceito complexo
 - 4.2.3.2.1 As opções
 - 4.2.3.2.2 Fatores a considerar
 - 4.2.3.2.3 Circunstâncias que favorecem a divisão de um conceito complexo

- 4.2.3.2.4 Circunstâncias contrárias à divisão de um conceito complexo
- 4.2.3.3 Como dividir um conceito complexo
- 4.2.3.4 Retenção de termos constituintes
 - 4.2.3.4.1 Partes e componentes
- 4.2.3.5 Consistência no tratamento de conceitos complexos
- 4.2.3.6 Ordem das palavras em termos de várias palavras
- 4.2.4 Relação de equivalência
 - 4.2.4.1 A relação de equivalência, em um contexto monolíngüe
 - 4.2.4.1.1 Sinônimos
 - 4.2.4.1.2 Quase-sinônimos
 - 4.2.4.1.3 Termos específicos incluídos em um conceito mais amplo
 - 4.2.4.2 Equivalência entre idiomas
 - 4.2.4.2.1 Graus de equivalência
 - 4.2.4.2.2 Problemas e soluções típicos
 - 4.2.4.2.2.1 Problemas causados por quase-sinônimos ou homógrafos
 - 4.2.4.2.2.2 Ausência de um termo equivalente em um ou mais idiomas
- 4.2.5 Relação entre conceitos
 - 4.2.5.1 O relacionamento hierárquico
 - 4.2.5.1.1 As relações genéricas
 - 4.2.5.1.2 O relacionamento hierárquico de todo/parte
 - 4.2.5.1.3 O relacionamento de instância
 - 4.2.5.1.4 Relações poli-hierárquicas
 - 4.2.5.2 O relacionamento associativo
 - 4.2.5.2.1 Termos e conceitos com significados sobrepostos
 - 4.2.5.2.2 Outros casos de ligação associativa Referências

4.1 ASPECTOS CONCEITUAIS DO TESAURO UNESP

Mariângela Spotti Lopes Fujita

Walter Moreira

Maria Luiza de Almeida Campos

O uso de um tesauro pressupõe a compreensão de seus conceitos básicos, por isso, tomaremos sua definição como ponto de partida para conhecermos os termos envolvidos (sublinhados no breve glossário a seguir) e seus conceitos básicos. As definições são baseadas na norma ISO 25964 – part1 (2011) e em Dahlberg, (1978).

TESAURO

Vocabulário controlado e estruturado em que os conceitos são representados por termos, organizados de modo que as relações entre conceitos sejam explicitadas e os termos preferidos sejam acompanhados por termos de entrada para sinônimos ou quasesinônimos.

Vocabulário controlado

Lista prescritiva de termos, cada um representando um conceito.

CONCEITO

Os conceitos, considerados como unidades de conhecimento, são representados por termos. Cada termo incluído em um tesauro deve representar um conceito único. Conceitos podem variar de simples (e.g., “gatos”) para muito complexos (e.g., “discriminação racial entre as minorias étnicas”). Termos ou frases compostos são geralmente necessários para expressar os conceitos mais complexos.

TERMO

A forma verbal de um conceito, o componente que, convenientemente, sintetiza e representa um conceito com o propósito de designá-lo e comunicá-lo. Assim como a palavra, utilizada na língua geral,

o termo é um signo linguístico, mas difere dela porque sua qualificação ocorre no interior dos discursos de especialidade.

RELAÇÕES ENTRE CONCEITOS

Para formar um campo conceitual de uma determinada área é necessário que os conceitos da área, representados por termos, mantenham proximidade que os relacione entre si, como por exemplo, relações de equivalência estabelecidas entre termos sinônimos, termos relacionados por associação a uma mesma ideia, ou ainda, relações hierárquicas entre um termo subordinado e outro superordenado, entre outros tipos de relações entre conceitos estabelecidas em um tesauro.

TERMOS PREFERIDOS E TERMOS NÃO-PREFERIDOS

A definição de termos preferidos e não preferidos é essencial no tesauro. Termos preferidos e não preferidos são identificados pelo estabelecimento de relações de equivalência entre os termos sinônimos, quase-sinônimos, homônimos e etc. O termo não-preferido deverá ser utilizado como remissiva para o termo preferido, sendo assim, não se configura como um descritor. O termo preferido, ao contrário, poderá ser utilizado como descritor e a partir dele são indicados os outros tipos de relações com outros conceitos componentes do tesauro.

REFERÊNCIAS:

DAHLBERG, I. Teoria do conceito. *Ciência da Informação*, Brasília, v. 7, n. 2, p. 101-107, 1978. Disponível em: <http://revista.ibict.br/ciinf/article/view/115/115>. Acesso em: 24 ago. 2018.

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

4.2 ASPECTOS FORMAIS DO TESAURO UNESP

O processo de construção das diretrizes para a inclusão de termos no Tesouro Unesp teve como base o Manual de cabeçalhos de assunto: normas e procedimentos da Fundação Getúlio Vargas (1995) e a INTERNATIONAL STANDARD. ISO 25964-1(2011).

A escolha do primeiro deu-se pelo fato da Unesp, desde a automação do seu sistema de catalogação descritiva, ter utilizado a Rede Bibliodata, que era mantida pela Fundação Getúlio Vargas (FGV), para fins de cooperação e padronização de seus registros bibliográficos. A essas regras foram acrescentadas as disposições da ISO 25964-1 por se tratar de um Padrão Internacional.

Após um trabalho de tradução e comparação entre os documentos, e com o objetivo de criar um material de trabalho e apoio criterioso para a Rede de Bibliotecas da Unesp, decidiu-se pela descrição detalhada e pelo uso de exemplos constantes do Tesouro Unesp, porém, para manter a previsão de todas as situações constantes dos documentos onde, por diferença de cultura, contexto de aplicação do termo ou nível de especificidade de indexação, não foi possível aplicar a tradução do termo existente na Norma, alguns exemplos com termos que não constam do Tesouro Unesp foram mantidos para ilustrar as explicações.

4.2.1 FORMA GRAMATICAL DOS TERMOS

Os termos selecionados para representar conceitos podem ser termos de palavra única ou termos com várias palavras (ver 4.2.3).

4.2.1.1 SUBSTANTIVOS E FRASES NOMINAIS

Um termo de tesouro deve consistir preferivelmente em um substantivo ou em uma frase nominal. Isso inclui gerúndios, que são substantivos verbais (ver 4.2.1.4). No idioma inglês (e em outros, como alemão, francês e outras línguas românicas), as frases nominais ocorrem em duas formas:

a) frases adjetivas

Exemplos:

Medicina tropical

Estudantes universitários

Produtos agrícolas Redes elétricas

b) frases preposicionais.

Exemplos:

Prisioneiros de guerra

Imposto de exportação

Indenização por acidentes

Abrigos para jovens

Como as preposições podem tornar um termo muito extenso elas devem ser evitadas, se possível, especialmente em termos preferidos.

4.2.1.2 ADJETIVOS

Adjetivos podem ser úteis como componentes de sintagmas nominais, mas se usados isoladamente, são uma causa potencial de problemas na recuperação. Uma busca por um artigo sobre “o uso de luzes vermelhas como sinais de alerta para pontes baixas”, se indexados com termos adjetivos como “vermelho” e “baixo”, poderia resultar na recuperação de informações sobre luzes baixas e/ou pontes vermelhas.

Por esta razão, o uso de adjetivos como termos do tesauro deve ser evitado. A mesma precaução se aplica mesmo se o adjetivo for convertido no substantivo correspondente, por ex. “vermelho” para “vermelhidão”.

4.2.1.3 ADVÉRBIOS

Os advérbios como “muito” ou “altamente” não devem ser usados como termos do tesouro. Uma frase que começa com um advérbio não deve ser aceita como um termo de tesouro a menos que tenha adquirido um significado especial.

Exemplos (que não devem ser usados):

Frequência muito alta

Integração em larga escala

4.2.1.4 VERBOS

Verbos expressos como infinitivos ou participios não devem ser usados sozinhos como termos do tesouro. As atividades devem ser representadas por substantivos ou substantivos verbais.

Exemplos:

Culinária (não “cozinhar”)

Destilação (não “destilar”)

Natação (não “nadar”)

4.2.1.5 ARTIGOS INICIAIS

4.2.1.5.1 OMISSÃO

O uso de artigos iniciais nos termos do tesouro deve geralmente ser evitado. Se necessário, um qualificador entre parênteses deve ser usado.

Exemplo:

Estado (entidade política) em vez de “O estado”

4.2.1.5.2 RETENÇÃO

Se o artigo inicial é parte integrante de um nome e precisa ser pesquisável, ele deve ser incluído no termo do tesouro em ordem direta. Caso contrário, o artigo deve ser omitido ou o termo invertido. Se a omissão

causar ambiguidade, um qualificador deve ser adicionado. Se o artigo é ou não considerado parte integrante do nome, depende da linguagem e do contexto.

Exemplos:

“A priori”

“O fantástico na literatura”

Onde um termo pode ser procurado com ou sem o artigo, uma referência deve ser feita a partir da forma não preferida.

4.2.1.6 CAPITALIZAÇÃO, PONTUAÇÃO E CARACTERES ESPECIAIS

4.2.1.6.1 CAPITALIZAÇÃO

Um estilo consistente deve ser usado para a apresentação dos termos do tesauro. No Tesauro Unesp, deve-se usar letras minúsculas com iniciais maiúsculas.

Exceções podem ser feitas para abreviações, acrônimos, nomes próprios ou termos que são convencionalmente escritos em um estilo especial. Nesses casos, o estilo mais aceito entre os usuários previstos do tesauro deve ser usado.

Exemplos:

ActiveX (Programa de computador)

DNA pH Fotocópias

4.2.1.6.2 CARACTERES NÃO ALFABÉTICOS

O uso de sinais de pontuação, sinais diacríticos e outros caracteres especiais pode causar problemas na digitação de termos. Apóstrofes, parênteses, hífens e caracteres numéricos podem complicar as expressões de busca.

Exemplo 1 (de evitar caracteres especiais):

Raios beta	não	Raios β
Pesquisa e desenvolvimento	não	Pesquisa & desenvolvimento

No entanto, como seu uso não pode ser completamente evitado, eles devem ser mantidos onde quer que a terminologia seja ambígua, incorreta ou inaceitável para a comunidade de usuários do tesouro, particularmente em abreviações, nomes químicos, nomes próprios e marcas registradas, ou se esses termos pertencerem a um vocabulário padronizado no domínio.

Exemplo 2 (onde caracteres especiais não podem ser evitados):

Micro-ondas

Calculadora eletrônica HP-48G

C++ (Linguagem de programação de computador)

Controle H [Infinito]

Ácidos graxos Ômega-6

Hífens e diacríticos também devem ser mantidos em termos não preferidos se estiverem de acordo com a ortografia oficial.

Sempre que possível (mas não no caso de nomes de produtos químicos e outros onde fazem parte integrante do termo), o uso de parênteses deve ser limitado a qualificadores.

4.2.1.7 FORMAS SINGULARES OU PLURAIS

4.2.1.7.1 FATORES CULTURAIS E LINGUÍSTICOS

Diferentes convenções existem em diferentes linguagens naturais com relação ao uso de singulares ou plurais. Os indexadores em algumas comunidades de idiomas, por exemplo, o francês e o alemão, tendem a usar a forma singular para que o usuário possa abordar e usar o tesouro da mesma forma que abordaria e usaria um dicionário. Em inglês e espanhol,

no entanto, é comum basear a escolha sobre se um termo específico é um substantivo contável ou um substantivo não contável. A última convenção ajuda a distinguir entre um processo como “pintura”, que só pode ser expresso no singular e o produto do mesmo processo, neste caso “pinturas”. A forma dos termos em cada idioma deve basear-se nas convenções aplicadas nesta linguagem. Como consequência de tal prática, um tesauro multilíngue é suscetível de ter entradas em que um termo no singular em francês ou alemão tem um equivalente em inglês no plural.

Exemplo:

fr:

maison de:

Haus en:

houses es:

casas por:

Habitações

4.2.1.7.2 TRATAMENTO DE NOMES CONTÁVEIS

Os nomes contáveis são nomes de entidades contáveis que estão sujeitos à pergunta “Quantos?”, mas não “quanto?” No Tesauro Unesp nomes contáveis devem ser expressos no plural.

Exemplo 1 (de substantivos contáveis):

Documentos

Pinguins

Partidos políticos

Janelas

Uma exceção são os nomes de partes do corpo, que geralmente são expressas no singular.

Exemplo 2 (de exceções):

Aparelho digestivo

Cabeça

Outra exceção é o nome dos organismos vivos. Muitas espécies, como *Escherichia coli*, não têm um nome comum e são conhecidas apenas pelo seu nome científico, que por convenção é expresso em latim, no singular. No interesse da consistência, quando os nomes latinos e os nomes aceitos como inglês coexistem em uma compilação, o singular pode ser aplicado por toda parte.

4.2.1.7.3 TRATAMENTO DE SUBSTANTIVOS NÃO CONTÁVEIS

Os nomes não contáveis são nomes de conceitos como materiais ou substâncias que estão sujeitos à pergunta “quanto?”, mas não “quantos?” No Tesouro Unesp nomes não contáveis devem ser expressos no singular.

Exemplo 1 (de substantivos não contáveis):

Algodão

Alumínio

Vapor

No entanto, se a comunidade de usuários servidos pelo tesauro considerar uma determinada substância ou material como uma classe com mais de um membro, a classe deve ser expressa no plural.

Exemplo 2 (de exceções):

Venenos

Os nomes de fenômenos abstratos, propriedades, sistemas de crença, atividades e disciplinas são frequentemente nomes não contáveis que devem ser expressos em suas formas singulares.

Exemplo 3 (nomes adicionais não contáveis):

Fenômenos abstratos:	Personalidade; Inverno
Propriedades:	Fragilidade; Solubilidade; Radioatividade
Sistemas de crença:	Catolicismo; Xintoísmo; Comunismo
Atividades ou processos:	Corte; Imigração; Envelhecimento
Disciplinas:	Astronomia; Sociologia

No entanto, quando um conceito abstrato é considerado como uma classe com mais de um membro, o termo que representa a classe deve ser expresso no plural.

Exemplo 4 (exceções adicionais):

Reações químicas

Testes de inteligência

Ciências físicas

4.2.1.7.4 COEXISTÊNCIA DO SINGULAR E PLURAL

Em qualquer idioma, onde as formas singular e plural de um termo se referem a conceitos diferentes, ambos devem ser inseridos no tesauro. A distinção entre eles deve ser reforçada pela adição de uma nota de escopo e, se possível, um termo ou frase de qualificação.

Exemplo 1:

Costa

Costas (Anatomia)

O qualificador adicionado se torna parte integrante do termo; não constitui uma nota de escopo.

Onde singular e plural denotam o mesmo conceito, mas a sua grafia difere de tal forma que os termos seriam separados por termos não relacionados quando arquivados alfabeticamente, uma referência deve ser feita a partir da forma não preferida.

Exemplo 2:

mouse USE mice

Quando singular e plural denotam o mesmo conceito, e a sua grafia difere muito pouco, ex. “avental” e “aventais”, é usual fornecer uma entrada apenas para o termo preferido, de acordo com as regras para substantivos contáveis e não contáveis.

4.2.2 CLARIFICAÇÃO E DESAMBIGUAÇÃO DOS TERMOS DO TESAURO

No discurso ordinário, um único termo pode ter mais de um significado, e o significado mais comum nem sempre é o necessário para representar o conceito desejado no tesauro. Se o contexto fornecido pela hierarquia de conceitos mais amplos e restritos ligados ao conceito em questão é insuficiente para elucidar o escopo pretendido, medidas adicionais devem ser aplicadas. Embora todos os termos do tesauro devam ser expressos da forma mais inequivocamente possível, é especialmente importante formular o termo preferido de um determinado conceito de tal forma que ele transmita o escopo pretendido a qualquer usuário. Por exemplo, o termo multissignificado “depressão” poderia ser reformulado como “depressão econômica” ou “depressão mental”, conforme apropriado. Como alternativa, um qualificador (ver 4.2.2.1) pode ser usado. Nos casos em que essas medidas não sejam adequadas ou suficientes, ou onde informações adicionais possam ajudar a esclarecer o significado e tornar o uso mais consistente, uma nota explícita de escopo deve ser usada (ver 4.2.2.2.1).

4.2.2.1 HOMÓGRAFOS E QUALIFICADORES

Homógrafos (às vezes referidos pelo termo mais amplo “homônimos”) são palavras com a mesma ortografia, mas significados diferentes.

Quando homógrafos são necessários como termos do tesauro, o significado de cada termo deve ser esclarecido e a maneira tradicional de fazer isso é adicionar um qualificador entre parênteses. O qualificador deve ser o mais breve possível, idealmente composto por uma palavra. Muitas vezes, um termo mais amplo, o qualificador deve indicar o contexto ou área de assunto a que o conceito pertence. Faz parte do termo e não serve como uma nota de escopo (ver 4.2.2.2.1).

Exemplo 1:

Harmonia (Estética)

Harmonia (Música)

Um qualificador deve ser adicionado a cada termo homográfico, mesmo quando um de seus sentidos é mais comum que o outro no domínio do tesauro.

Se apenas um dos termos homográficos for usado em um tesauro especializado e seu significado for óbvio para os usuários do tesauro, o qualificador poderá ser excluído. No entanto, deve-se considerar a possibilidade de que o escopo do tesauro possa ser ampliado no futuro, ou que a interoperabilidade com outros vocabulários cobrindo diferentes âmbitos seja necessária. No Tesauro Unesp recomenda-se esclarecer ambos os termos homográficos.

Como os qualificadores tornam os termos difíceis de aplicar, e como alguns sistemas eletrônicos têm dificuldades em aplicá-los, seu uso (especialmente em termos preferidos) deve ser evitado se outro meio de resolver a ambigüidade puder ser encontrado. Por esse motivo, um termo com várias palavras deve ser preferido a um termo de palavra única com um qualificador, contanto que a forma composta ocorra em linguagem natural.

Exemplo 2:

Por exemplo, “pluralismo cultural” deve ser preferido a “pluralismo (Ciências sociais)”. A última expressão, no entanto, pode ser inserida como um termo não preferencial.

Qualificadores entre parênteses não devem ser usados para fornecer um termo na forma invertida.

Exemplo 3:

Por exemplo, a forma invertida “canetas (tinteiro)” está incorreta e a forma direta “canetas-tinteiro” deve ser usada no lugar. “Tinteiro” é usado no último exemplo para indicar um tipo de caneta em vez de desambiguar a palavra “caneta”. (Veja também 4.2.3.6, a ordem das palavras em termos de várias palavras).

No caso de siglas e abreviaturas, que são muitas vezes homógrafos, o qualificador deve ser a forma completa do termo (ver também 4.2.2.3.7).

Os qualificadores de siglas e abreviaturas só devem ser usados quando seu significado for considerado ambíguo para os usuários do tesauro, ou seja, quando tiverem outro significado conhecido no domínio coberto pelo tesauro ou na linguagem comum.

4.2.2.2 NOTAS

4.2.2.2.1 NOTAS DE ESCOPO

O escopo pretendido de um conceito no tesauro nem sempre é o significado mais comum associado ao seu termo preferido (ou qualquer um de seus termos não preferenciais) no discurso ordinário. No tesauro, o contexto é geralmente fornecido pela hierarquia de conceitos mais amplos e restritos ligados ao conceito em questão e isso ajuda a elucidar o escopo pretendido. Muitas vezes é possível formular o termo preferido do conceito de maneira inequívoca que transmite o escopo pretendido para qualquer usuário. No entanto, nos casos em que essas medidas não sejam adequadas ou suficientes, ou onde informações adicionais ajudem a esclarecer o significado e tornar o uso mais consistente, uma nota de escopo explícita deve ser usada. Uma nota de escopo deve ser usada para esclarecer os limites de um conceito, especialmente quando o significado do termo preferido no discurso ordinário pode ser interpretado de forma muito ampla ou restrita, ou para distinguir entre termos preferidos que têm significados sobrepostos em linguagem natural. Ele também pode ser usado para fornecer outros conselhos sobre o uso do termo para o indexador ou para o pesquisador. Uma nota de escopo não precisa ser uma definição completa, mas deve esclarecer o uso pretendido de um termo dentro do tesauro. Outros tipos de nota podem ser transmitidos separadamente; veja 4.4.2.2.2 para definições e 4.2.2.2.3 para notas históricas. Ao contrário dos qualificadores uma nota de escopo não é considerada como parte do termo ao qual ela está anexada.

Exemplo 1:

Produtos primários

Nota Usado para obras que tratam de matérias-primas e bens semi-acabados.

Exemplo 2:

Profissionais de nível superior

Nota Usado para obras que tratam dos graduados em universidades como um grupo sócioeconômico.

Quando é feita referência a outros conceitos em uma nota de escopo, uma nota de escopo recíproco deve ser geralmente fornecida para cada conceito mencionado.

Exemplo 3:

Projeção (Psicologia)

Nota Usado para obras que tratam dos conceitos psicanalíticos da projeção como mecanismo de defesa. Obras sobre projeção como teste ou diagnóstico técnico para revelar atributos de personalidade ou fatores psicopatológicos usar Técnicas projetivas.

Técnicas projetivas

Nota Usado para obras que tratam da projeção como um teste ou diagnóstico técnico, para revelar atributos da personalidade ou fatores psicopatológicos. Obras sobre o conceito psicanalítico de projeção como um mecanismo de defesa, entram em Projeção (Psicologia).

Mesmo onde o escopo de apenas um dos conceitos requer esclarecimento, é útil fazer uma nota editorial no registro para qualquer conceito adicional, de onde ele foi citado. O propósito da referência recíproca é assegurar que quando uma mudança é feita em um dos conceitos, ou é deletada, o efeito no outro conceito é considerado.

4.2.2.2.2 DEFINIÇÕES

Uma definição completa geralmente não é necessária para esclarecer a maneira pela qual um termo preferido deve ser usado. No

entanto, se uma definição for necessária por outro motivo, um campo de anotações deverá ser estabelecido para as definições. A fonte de cada definição deve ser registrada ao lado da própria definição. No Tesouro Unesp as notas de definição também aparecem como “Nota”.

Exemplo:

Amálgama dentário

Nota Liga usada na odontologia restauradora que contém mercúrio, prata, estanho, cobre, e possivelmente zinco.

4.2.2.2.3 NOTAS HISTÓRICAS

Quando um tesouro é atualizado, as mudanças nos termos podem afetar a capacidade de recuperação de um conceito. Em tais casos, uma nota de histórico pode ser usada.

Uma nota de histórico pode ser aplicada a um único termo preferido ou não preferido ou a um conceito. Ela deve ser usada quando um novo termo preferido é adicionado ao tesouro ou uma alteração é feita em um termo existente que afeta o escopo do conceito em diferentes períodos de aplicação. Embora seja possível incluir essas informações na nota de escopo, uma nota de histórico separada é preferível. Uma nota de histórico pode registrar a data de introdução ou pode dar conselhos mais complexos sobre como pesquisar o conceito em momentos anteriores ou posteriores. No Tesouro Unesp as notas de histórico aparecem como “Nota”.

Exemplo 1:

Ensino de primeiro grau

Nota Usado para obras que tratam do ensino no Brasil a partir de 1971, isto é, da Lei 4.024 que fixa as diretrizes e bases da educação formal que vai da primeira à oitava série. Obras sobre o mesmo assunto editadas até 1970, correspondendo à níveis de escolaridade de primeira à quarta série entram em Ensino primário. Obras sobre o mesmo assunto editadas até 1970, correspondendo à níveis de escolaridade de quinta à oitava série entram em Ensino ginásial. Obras a partir da Lei 9394 de 20.12.1996, correspondendo à níveis de escolaridade de primeira à oitava série e obras a partir da Lei n. 11.274/2006 que fixa em nove anos o ensino fundamental,

entram em Ensino fundamental. Para países estrangeiros usar sempre Ensino primário.

Exemplo 2:

Ensino fundamental

Nota Usado para obras posteriores a Lei 9.394 de 20.12.96 que trata do ensino formal da primeira à oitava série. Obras sobre o mesmo assunto editadas até 1970, pertinentes a níveis de escolaridade de primeira à quarta série entram em “Ensino primário”. Obras sobre o mesmo assunto editadas até 1970, pertinentes a níveis de escolaridade de quinta à oitava série entram em “Ensino ginásial”. A partir de 1971, isto é, da Lei 4.024 que fixa as diretrizes e bases da educação nacional, pertinentes à níveis de escolaridade de primeira à oitava série entram em “Ensino de primeiro grau”. Para países estrangeiros usar sempre “Ensino primário”.

4.2.2.3 SELEÇÃO DA FORMA PREFERIDA

Às vezes existe mais de um termo na mesma língua para representar o mesmo conceito. Quando múltiplos termos são incluídos correspondentes ao mesmo conceito, um deles deve ser designado como o termo preferido e os outros termos não preferidos.

Onde há uma escolha entre formas de expressão sinônimas, as preferências da comunidade a ser atendida pelo tesauro devem ser adotadas (sujeito a evitar ambiguidades ou linguagem que possa ser ofensiva para alguns usuários).

4.2.2.3.1 ORTOGRAFIA

A ortografia mais aceita das palavras deve ser adotada para os termos preferidos. Se grafias variantes existirem e forem comumente reconhecidas, cada uma deve ser inserida no tesauro como um termo não preferido e uma referência deve ser feita do termo não preferido para o termo preferido.

Exemplo 1:

Queijo muçarela

USE Queijo mussarela

Sempre que possível, a ortografia deve seguir a prática de um dicionário ou glossário bem estabelecido na linguagem do tesouro.

Palavras com erros ortográficos não são aceitáveis em termos preferidos. No entanto, alguns erros de ortografia são tão comuns que podem fornecer pontos de entrada úteis, especialmente em tesouros eletrônicos, e podem ser usados como ou em termos não preferidos.

Exemplo 2:

Micro-ondas

UP Microondas

Pontos de entrada com erros de ortografia devem ser fornecidos somente quando:

- a) o erro de ortografia é comum o suficiente para valer a pena; e
- b) não há perigo de que o erro de ortografia esteja relacionado a um termo diferente do termo preferido indicado.

4.2.2.3.2 TERMOS DE EMPRÉSTIMO E TRADUÇÕES DE TERMOS DE EMPRÉSTIMOS (ESTRANGEIRISMO)

Termos de empréstimo ou termos emprestados são termos estrangeiros que foram absorvidos pela língua portuguesa e podem ser usados como termos preferidos se estiverem bem estabelecidos.

Exemplo 1 (de termos de empréstimo):

Marketing

Designers

Flashbacks

Stress

Ocasionalmente, um termo de empréstimo e uma tradução putativa coexistem. Se o termo de empréstimo for mais amplamente aceito, ele deve ser tratado como o termo preferido, enquanto a tradução pode ser preferida se estiver bem estabelecida. Referências recíprocas devem ser feitas entre os dois termos.

Exemplo 2 (de termos de empréstimo com traduções):

Marketing

UP Mercadologia

Mercadologia

USE Marketing

Flashbacks

UP Analepse

Analepse

USE Flashbacks

4.2.2.3.3 TRANSLITERAÇÃO

Ao transliterar termos de idiomas com alfabetos diferentes, um esquema reconhecido deve ser seguido. O esquema utilizado deve ser especificado na introdução do tesauro.

4.2.2.3.4 NEOLOGISMOS, TERMOS DE GÍRIAS E JARGÕES

Devem ser adotados os termos estabelecidos para os conceitos ao invés de termos de gíria ou jargão, uma vez que estes são frequentemente transitórios.

Neologismos úteis, no entanto, são frequentemente cunhados em resposta a tecnologias emergentes ou desenvolvimentos na sociedade ou na política. Eles podem ser adotados como termos do tesauro assim que forem julgados úteis para a recuperação de informações.

Exemplo 1 (de neologismos úteis):

Transfobia

Smartphones

RSS feeds

Rave (Festa)

Entradas de gírias ou jargões podem ser necessárias em circunstâncias como as seguintes:

a) Um conceito recém-emergente é expresso por um termo originário de uma subcultura ou grupo social específico e não existe uma alternativa amplamente aceita. O termo de gíria ou jargão deve então ser aceito como um termo preferido.

Exemplo 2 (de termos de gíria ou jargão aceitáveis):

Cultura cosplay

Selfies

b) Onde um termo de gíria ou jargão é amplamente utilizado no lugar do termo estabelecido, muitos usuários podem considerar um ponto de entrada útil. Um termo não-preferido deve ser fornecido da gíria ou jargão ao termo preferido.

Exemplo 3 (de termos de gíria ou jargão com status não preferido):

Preservativos

UP Camisinha (Preservativo masculino)

Camisinha (Preservativo masculino)

USE Preservativos

c) Onde o público esperado tem uma forte preferência por gírias ou jargões, pode ser útil usá-los como termos preferidos, inserindo os termos alternativos como não preferidos. Isso pode acontecer, por exemplo, com uma platéia de jovens (preferindo gírias).

4.2.2.3.5 NOMES COMUNS E NOMES COMERCIAIS

Um produto é freqüentemente conhecido por um nome comercial reconhecido. Quando um nome comum adequado também existe, normalmente deve ser adotado como o termo preferido e o nome comercial deve ser admitido como um termo não preferido somente se for provável que sirva como ponto de acesso de um usuário.

Exemplo:

Aspirina

UP Ácido acetilsalisílico

Ácido acetilsalisílico

USE Aspirina

Exceções podem ocorrer, por exemplo, nos casos em que um produto genérico é vendido sob vários nomes comerciais concorrentes e os usuários desejam distinguir entre as diferentes marcas. Neste caso, o termo preferido para o produto genérico deve ser o nome comum, e os nomes comerciais podem ser usados como termos mais restritos.

4.2.2.3.6 NOMES POPULARES E NOMES CIENTÍFICOS

Se um nome popular e um científico se referirem ao mesmo conceito, a preferência deve ser dada à forma mais provável de ser procurada pelos usuários do tesauro. Por exemplo, “pinguins” pode ser escolhido como o termo preferido em um tesauro geral, mas o equivalente científico, “Sphenisciformes”, pode ser preferido em um tesauro de zoologia. Referências recíprocas devem ser feitas nesses casos.

Exemplo:

Orquídeas

UP Orquidáceas

4.2.2.3.7 ABREVIACÕES E SIGLAS

Abreviações e siglas são frequentemente ambíguas, porque podem se referir a mais de um conceito. Portanto, a forma completa do nome deve funcionar normalmente como o termo preferido, com uma referência recíproca da forma abreviada.

Exemplo 1 (de desambiguação acrônimo / abreviação):

Produto Interno Bruto

UP PIB

PIB

USE Produto Interno Bruto

Acidente Vascular Cerebral

UP AVC

AVC

USE Acidente Vascular Cerebral

Excepcionalmente, abreviações ou acrônimos podem ser usados como termos preferidos quando são amplamente conhecidos, não ambíguos e facilmente compreendidos dentro do campo coberto pelo tesouro, especialmente se eles se tornaram tão bem estabelecidos que a forma completa do nome é raramente usada ou é geralmente ignorada. Referências recíprocas ainda devem ser feitas entre o termo completo e sua abreviação.

Exemplo 2 (de siglas / abreviaturas que não precisam de desambiguação):

Vírus da Imunodeficiência Humana USE HIV

HIV

UP Vírus da Imunodeficiência Humana

4.2.2.3.8 NOMES PESSOAIS PRÓPRIOS, NOMES DE LUGARES E NOMES DE INSTITUIÇÕES

Nomes próprios, nomes de lugares e nomes de instituições não fazem parte do escopo do Tesauro Unesp. Para a inclusão destes na base de autoridades da Unesp consultar o Padrão de Qualidade de Registros de Autoridade da Unesp e o AACR2.

4.2.3 CONCEITOS COMPLEXOS

Os conceitos variam do muito simples ao mais complexo. Os conceitos simples são noções como “prata” ou “pessoas”. Conceitos mais complexos seriam “bules de prata georgianos” ou “defensores dos direitos humanos”. A última expressão, por exemplo, combina pelo menos três conceitos distintos em um conceito mais complexo.

Conceitos complexos são muitas vezes transmitidos por termos compostos, isto é, termos que podem ser divididos morfológicamente em dois ou mais componentes. Alguns deles são termos com várias palavras, e alguns consistem em apenas uma palavra.

A disponibilidade de diversas opções apresenta ao editor de tesouros uma decisão difícil e subjetiva: adotar o conceito complexo ou se deve confiar em conceitos mais simples para os usuários aplicarem combinação. Onde o tesauro permite várias opções, uma consideração, o importante é guiar todos os usuários para usarem a mesma combinação para um determinado conceito. A consistência torna-se ainda mais difícil de alcançar se o conceito for dividido em mais de dois componentes, como no exemplo de “pessoas + informação + gerenciamento” ou Gerenciamento de informação pessoal.

Outra consideração é que a introdução de conceitos complexos, seja representada por palavra única ou termos com várias palavras, tende a aumentar a especificidade. Maior especificidade ajuda os usuários a alcançar maior consistência. Por exemplo, se a “biodegradabilidade” é admitida como um termo mais específico de “degradabilidade”, os pesquisadores têm uma maneira fácil de recuperar documentos relevantes sem itens indesejados sobre a “degradabilidade” por meios químicos, ação da luz ultravioleta, etc. No entanto, há uma desvantagem quando o tesauro

se torna maior e às vezes separa itens semelhantes, dificultando para os usuários identificar o termo correto.

Exemplo:

Bioacústica

Biocatálise

Bioclimatologia

4.2.3.1 A NATUREZA DOS TERMOS COMPOSTOS

Os termos compostos geralmente consistem em mais de uma palavra.

Exemplo 1:

Gerenciamento de configurações de software

Segurança alimentar

Aves de rapina

As partes da maioria desses termos podem ser divididas da seguinte maneira:

a) O foco ou cabeça, isto é, o componente substantivo que identifica a classe geral de conceitos à qual o termo como um todo se refere;

Exemplo 2 (de foco no termo):

o componente substantivo “bibliotecas” no termo “bibliotecas escolares”
o substantivo “hospitais” na frase preposicional “hospitais infantis”

b) A diferença ou modificador, isto é, um ou mais componentes adicionais que servem para estreitar o escopo do foco e, desse modo, especificar uma de suas subclasses.

Exemplo 3 (de diferença de termo):

1. o adjetivo “impresso” no termo “índices impressos”

2 . a combinação preposição-mais-substantivo “para crianças” no termo “hospitais para crianças”

Exemplo 4 (expressões que representam conceitos determinados):

Crime contra o meio ambiente

Fé e razão

Trabalhadores da agroindústria açucareira

Indenização por acidentes

Uma análise semelhante pode às vezes ser aplicada a termos compostos de uma só palavra.

Exemplo 5 (de termos compostos de palavra única):

Termo	Foco	Diferença Micro
Microfinanças	Finanças	(prefixo associado a pequeno)
Biocombustíveis	Combustíveis	Bio (prefixo associado à vida)

4.2.3.2 DECIDIR SE DEVE OU NÃO ADMITIR UM CONCEITO COMPLEXO

4.2.3.2.1 AS OPÇÕES

Tomando um exemplo direto, um conceito como “segurança viária” poderia ser representado usando uma combinação dos conceitos mais simples “estradas” e “segurança”. À medida que a complexidade aumenta, mais escolhas devem ser consideradas. Assim, um software de gerenciamento de tesouros pode ser representado por Gerenciamento de recursos de informação: gerenciamento + recursos + informação.

As opções principais a seguir devem ser consideradas para lidar com um conceito complexo proposto.

a) Admitir o conceito, representando-o com um único termo preferido.

Exemplo 1:

Biodegradação

UP Degradação biológica

b) Admita o conceito, representando-o como uma combinação de dois ou mais termos. Neste caso, um termo não-preferido para todo o conceito deve ser inserido, apontando para a combinação apropriada de termos preferidos mais simples. Esta opção é muitas vezes referida como “dividir” o conceito.

Exemplo 2:

Biologia espacial

UP Cosmobiologia

c) Rejeitar o conceito complexo, caso seja muito raramente solicitado, irrelevante ou inadequado, mas mantenha um conceito mais amplo e admita um termo não preferencial que aponte para o termo preferido.

Exemplo 3:

Narcóticos sintéticos USE Opióides

Pacientes Codependentes USE Codependentes

4.2.3.2.2 FATORES A CONSIDERAR

A decisão sobre a admissão é muitas vezes difícil e subjetiva. Fatores como os seguintes devem ser levados em conta:

a) Frequência de uso: o termo é amplamente utilizado e compreendido pelo público.

b) Grau de discriminação necessário. Se houver um grande volume de documentos relevantes a serem indexados, pode ser importante discriminar entre o conceito complexo e conceitos similares.

Exemplo 1:

Responsabilidade (Direito);

Responsabilidade civil.

c) Multiplicidade de opções de pesquisa, levando à recuperação incompleta. A disponibilidade de um grande número de termos preferidos relacionados pode complicar a escolha de opções. Por exemplo, se o tesauro tiver todos os termos “comércio”, “comércio de cereais”, “comércio de sementes”, “comércio de calçados”, “comércio de livros” e “comércio de hortaliças”, o usuário poderá precisar experimentar todos esses termos. Se o número total de documentos sobre comércio for pequeno, a pesquisa poderá ser mais fácil e eficaz se nenhum dos conceitos complexos for aceito como um termo preferencial e uma expressão de pesquisa pós-coordenada como “cereais +comércio” for a única opção de recuperação.

Exemplo 2:

Comércio de cereais

USE Cereais - Comércio

d) Termos simples foram utilizados em que o foco “agentes” foi qualificado por apenas uma diferença, por exemplo, “penitenciários”. Maior complexidade ocorre com termos como “Agentes de segurança penitenciária”, para a qual o foco é qualificado simultaneamente por duas diferenças. Termos como esses multiplicam ainda mais as opções de busca e impedem a recuperação direta e, portanto, devem ser evitados.

Exemplo 3:

Agentes de segurança penitenciária

USE Agentes penitenciários

e) Um termo pode ser útil em um tesauro para servir especialistas em recuperação de informações, mas pode ser confuso se o tesauro for destinado a um público mais geral.

f) Conceitos de componentes podem ser combinados de maneiras diferentes, tendo diferentes significados. Por exemplo, uma combinação de “bibliotecas” e “ciência” poderia ser usada para representar “biblioteconomia” ou “bibliotecas científicas”.

g) Complicações multilíngues. No caso de tesouros multilíngues, restrições ou necessidades específicas encontradas em qualquer uma das línguas podem influenciar a inclusão ou exclusão de um conceito complexo em todos os idiomas.

h) Ambiente de pesquisa antecipado. Se o software de pesquisa oferecer recursos sofisticados, como agrupamento de resultados de pesquisa, personalização para refletir os interesses de determinados usuários ou comunidades de usuários, etc., o nível de especificidade necessário no tesouro poderá ser afetado. Esse fator deve ser equilibrado com a probabilidade de que o software e outros aspectos do ambiente possam mudar no futuro.

4.2.3.2.3 CIRCUNSTÂNCIAS QUE FAVORECEM A DIVISÃO DE UM CONCEITO COMPLEXO

As circunstâncias a seguir favorecem a divisão:

a) A divisão pode ser considerada quando o conceito é bastante específico e está fora do escopo principal do tesouro. A inclusão de um grande número de termos periféricos aumenta o volume e a complexidade do vocabulário sem dar muito benefício de recuperação.

b) Se muito poucos documentos puderem ser indexados com o termo proposto, a inclusão como um termo preferencial pode não valer a pena, e a divisão é uma opção a ser considerada.

c) O conceito complexo deve ser dividido se o foco for qualificado por mais de uma diferença. Por exemplo, “câmeras cinematográficas eletrônicas” devem ser divididas em “câmeras cinematográficas” e “câmeras eletrônicas”.

d) O conceito complexo geralmente deve ser dividido se o foco representar uma propriedade, parte ou componente da diferença.

Exemplos:

Aviões - Motores

Crianças - Assistência hospitalar

No entanto, exceções podem ocorrer se o conceito for bem definido e distinto, por exemplo, “abajures” ou “temperatura corporal”. É provável que as referências à temperatura corporal sejam significativamente diferentes das referências à temperatura de outros objetos inanimados ou substâncias que possam ser discutidas na mesma coleção de documentos.

4.2.3.2.4 CIRCUNSTÂNCIAS CONTRÁRIAS À DIVISÃO DE UM CONCEITO COMPLEXO

Um conceito complexo não deve ser dividido se as seguintes condições se aplicarem:

a) Um termo para o conceito tornou-se tão familiar no uso comum, ou no campo coberto pelo tesauro, que sua expressão como elementos separados dificultaria a compreensão.

Exemplo 1:

Processamento de dados

Produto Interno Bruto

b) Dividir o conceito em suas partes levaria a uma perda de significado ou ambiguidade.

Exemplo 2:

Alimentos vegetarianos

Plantas comestíveis

c) O conceito é representado por um termo estabelecido que é um nome próprio ou incorpora um nome próprio.

Exemplo 3:

Psicologia junguiana

Dewey, Classificação decimal de

d) A diferença em um termo estabelecido perdeu seu significado original.

Exemplo 4:

Bode expiatório

e) O conceito é representado por um termo contendo uma diferença que sugere uma semelhança, como uma metáfora, uma coisa ou evento não relacionado.

Exemplo 5:

Árvore da vida

f) As partes do conceito quando separadas não transmitem o conceito geral. Assim, seria enganoso indexar os documentos correspondentes com termos que representam os conceitos da parte.

Exemplo 6:

Seguro de responsabilidade do empregador

Parafusos ósseos (ortopedia)

Peste negra

Sequestro de carbono

g) O conceito é representado por um termo em que o foco tem outro significado na ausência da diferença.

Exemplo 7:

Queijo de soja

Flores artificiais

Ovos de chocolate

(Seria enganoso usar o termo “ovos” para indexar itens que lidam com ovos de chocolate, uma vez que os de chocolate são um tipo de confeitaria em vez de um tipo de ovo.)

4.2.3.3 COMO DIVIDIR UM CONCEITO COMPLEXO

Em casos como, por exemplo, o de segurança de computadores, o conceito pode ser adequadamente expresso usando um termo multi palavra, ou seja, “computadores” + “Medidas de segurança”. A justificativa para fazer isso em um tesauro específico pode ser que qualquer documento que lide com segurança de computadores provavelmente contenha informações relevantes para os dois termos constitutivos e que alguém que esteja procurando informações sobre computadores provavelmente julgue um artigo sobre segurança de computadores como relevante.

Quando um conceito é dividido, o conceito constituinte pode ser representado por termos preferidos que diferem das palavras do termo com várias palavras. Muitas vezes as palavras constituintes precisam ser modificadas para transmitir o conceito correto.

Exemplo:

Segurança em computadores

USE Computadores - Medidas de segurança

4.2.3.4 RETENÇÃO DE TERMOS CONSTITUINTES

No caso de um termo composto ser aceito como um termo preferencial deve-se considerar a inclusão de termos preferidos representando cada um de seus conceitos constituintes, se eles ainda não estiverem presentes no tesauro. Por exemplo, se “Crianças - Assistência hospitalar” for aceito, o tesauro deverá normalmente conter também os termos preferidos “Crianças” e “Assistência hospitalar”.

4.2.3.4.1 PARTES E COMPONENTES

A confusão pode surgir quando a diferença do termo composto é uma máquina ou outra montagem complexa da qual o foco se refere a um componente. Por exemplo, se “motores de aviões” forem divididos em “aviões” e “motores”, uma busca pelo termo “aviões” por si só retornará registros de todos os componentes de aviões.

Isso pode ser aceitável para algumas coleções, mas em outras resultará em uma sobrecarga do termo “Aviões” e não fornecerá meios para isolar os documentos que tratam de aviões inteiros. Uma solução é criar um termo preferido, como “componentes de aviões”. Este termo pode ser usado em combinação com “motores” ou “instrumentação” ou “subcarruagens” ou qualquer outro tipo de componente.

4.2.3.5 CONSISTÊNCIA NO TRATAMENTO DE CONCEITOS COMPLEXOS

Em geral, a consistência da prática é um objetivo útil, pois incentiva a aplicação consistente pelos usuários. No entanto, a consistência absoluta na admissão de conceitos complexos é difícil de alcançar e nem sempre é necessária. Como a especificidade necessária em áreas centrais do tesouro é geralmente maior do que para áreas periféricas, algumas inconsistências são inevitáveis. Esse tipo de inconsistência não tem um efeito negativo no desempenho de recuperação, desde que haja entradas claras e úteis suficientes no tesouro e, portanto, possa ser desconsiderado.

4.2.3.6 ORDEM DAS PALAVRAS EM TERMOS DE VÁRIAS PALAVRAS

Os termos preferidos que são frases nominais adjetivas ou preposicionais devem ser introduzidos no tesouro em ordem de linguagem natural, não como termos invertidos. No entanto, a forma invertida de uma frase preposicional pode ser inserida como um termo não preferido.

Exemplo:

Rapina, aves de

USE Aves de rapina

Para frases adjetivas (por exemplo, “pão integral” ou “camas de casal”), uma entrada invertida pode não ser necessária - porque o substantivo focal geralmente está presente como um termo mais amplo. A entrada para o termo mais amplo (por exemplo, “pão” ou “camas”) fornece um ponto de entrada a partir do qual todos os seus termos mais restritos podem ser encontrados.

4.2.4 RELAÇÃO DE EQUIVALÊNCIA

4.2.4.1 A RELAÇÃO DE EQUIVALÊNCIA, EM UM CONTEXTO MONOLÍNGUE

A relação de equivalência é a relação entre um termo preferido e seu correspondente não preferido na mesma língua natural. Essa relação se aplica entre termos e não entre conceitos.

A relação é recíproca e a reciprocidade entre termos preferenciais e não preferenciais deve ser expressa pelas seguintes convenções:

USE, escrito como um prefixo para o termo preferido;

UP (usado para), escrito como um prefixo para o termo não preferido.

Exemplos:

Plantas aquáticas

UP Plantas da água

Plantas da água

USE Plantas aquáticas

A equivalência é estabelecida em quatro situações gerais:

- a) os termos são sinônimos (ver 4.2.4.1.1);
- b) os termos são quase-sinônimos (ver 4.2.4.1.2);
- c) o termo é considerado desnecessariamente específico e é representado por outro termo com escopo mais amplo (ver 4.2.4.1.3);
- d) o termo é considerado desnecessariamente específico e é representado por uma combinação de dois ou mais termos (conhecido como “equivalência composta”).

4.2.4.1.1 SINÔNIMOS

Vários tipos de sinônimos são encontrados na prática. A lista a seguir não é exaustiva, mas indica algumas das classes mais comuns de sinônimos.

a) Termos de origem lingüística diferente:

Exemplo 1:

Alimentos; Comida

Concorrência; Competição

b) Nomes populares e nomes científicos:

Exemplo 2:

Orquídea; Orquidáceas

Mosca-branca; Bemisia tabaci

c) Nomes comuns e nomes comerciais:

Exemplo 3:

Aspirina (Exemplo de nome de marca consagrado na linguagem corrente);
Ácido acetilsalicílico

Ibuprofeno (exemplo de nome comum preferido ao nome de marca);
Salprofen

d) Nomes variantes para conceitos emergentes:

Exemplo 4:

Transfobia; Fobia a transgêneros

RSS feeds; Newsfeeds (RSS); Webfeeds (RSS)

e) Termos atuais ou favorecidos em relação a termos desatualizados ou obsoletos.

Exemplo 5:

Países em desenvolvimento; Países subdesenvolvidos

Antropologia física; Somatologia

f) Grafias variantes, incluindo variante da raiz, ordem de palavras invertidas e plurais irregulares. Entre as grafias variantes, pode ser útil incluir alguns erros ortográficos.

Exemplo 6:

Queijo mussarela; Queijo muçarela

Micro-ondas; Microondas

g) Termos originados de diferentes culturas compartilhando um idioma comum.

Exemplo 7:

Jeans (Vestuário)

h) Abreviaturas ou siglas e nomes completos.

Exemplo 8:

HIV; Vírus da Imunodeficiência Humana

Produto Interno Bruto; PIB

i) Nomes comuns e gírias ou termos jargão.

Exemplo 9:

Café solúvel; Café instantâneo

Nestes casos e similares, os termos preferidos devem ser selecionados para atender às necessidades da maioria dos usuários. Por uma questão de previsibilidade, estes critérios devem ser aplicados de forma consistente em todo o tesauro. Se, por exemplo, é decidido que os nomes populares, em vez de nomes científicos devem servir como termos preferidos, esta decisão deve ser aplicada de forma consistente, exceto quando um nome popular apropriado não existe. Da mesma forma, convenções ortográficas consistentes devem ser adotadas. As convenções editoriais e os critérios devem ser anotados em uma introdução ao tesauro.

4.2.4.1.2 QUASE-SINÔNIMOS

Quase-sinônimos frequentemente representam pontos em uma sequência, e podem até ser antônimos (opostos).

Em outros casos, os conceitos estão tão intimamente relacionados que a discussão de um provavelmente interessará a um usuário que busca o outro.

Exemplo 1 (de opostos tratados como quase-sinônimos):

Igualdade, Desigualdade Migração, Imigração

Exemplo 2 (de conceitos relacionados tratados como quase-sinônimos):

Voto, Sufrágio

Casamento entre homossexuais, União homoafetiva

A extensão em que os termos são tratados como quase-sinônimos depende, em grande parte, do campo de assunto coberto pelo tesouro. Em um tesouro especializado em sexualidade, os termos “Casamento entre homossexuais” e “União homoafetiva” podem ser estabelecidos como termos preferenciais, enquanto um tesouro mais geral pode tratar os dois termos como quase-sinônimos. A decisão deve basear-se em antecipar o grau de discriminação exigido no momento da pesquisa.

4.2.4.1.3 TERMOS ESPECÍFICOS INCLUÍDOS EM UM CONCEITO MAIS AMPLO

Às vezes, é útil tratar o nome de uma classe e também os nomes de seus membros, como um conjunto de equivalências, com o termo mais amplo funcionando como o termo preferencial.

Exemplo:

Neve

UP Neve granular

UP Neve artificial

UP Neve de primavera

UP Neve em pó

Essa técnica às vezes é usada para reduzir o número de termos preferidos em um tesauro. Ele limita a especificidade da indexação e a busca que pode ser alcançada, e é benéfico quando a coleção a ser indexada tem muito pouca informação sobre a área de assunto em questão. A presença dos termos não preferenciais fornece pontos de entrada extras para auxiliar o acesso do usuário. A substituição de conceitos mais restritos em um conceito mais amplo não deve, no entanto, ser usada quando há uma riqueza de informações sobre o assunto em questão e maior especificidade é necessária para identificar conceitos com precisão.

4.2.4.2 EQUIVALÊNCIA ENTRE IDIOMAS

Num tesauro multilíngue, todas as línguas devem ter o mesmo status e, se possível, cada conceito deve ser representado em todas as línguas do tesauro. Uma coleção que foi indexada usando qualquer uma das linguagens pode então ser pesquisada de forma igualmente eficaz usando qualquer um dos outros idiomas.

Nos casos em que um termo adequado não está disponível em um idioma, muitas vezes é possível modificar o escopo do conceito para que possa ser representado por um termo adequado em cada uma das línguas do tesauro.

4.2.4.2.1 GRAUS DE EQUIVALÊNCIA

Durante a preparação de um tesauro multilíngue frequentemente são encontrados os seguintes graus de equivalência:

a) Equivalência exata:

Nesta situação ideal, um conceito é representado em todas as línguas de um tesauro, e é possível identificar termos preferidos semanticamente e culturalmente equivalentes.

Exemplo 1:

pt: Física / en: physics

b) Equivalência inexata ou não equivalência:

Os termos encontrados em diferentes idiomas às vezes carregam uma pequena diferença no geral, talvez devido a diferenças de cultura, conotação ou apreciação (semelhante ao caso de quase sinônimos dentro de um idioma). Se, no entanto, os termos forem aceitos como próximos o suficiente para representar o mesmo conceito quando usado no tesouro, então um termo preferido para cada idioma deve ser selecionado, a equivalência entre eles estabelecida e o escopo do conceito entendido como incluindo todas as variações culturais.

Exemplo 2:

pt: recuperação da informação / en: information retrieval

c) Equivalência parcial, também conhecida como equivalência mais ampla/mais estreita:

Às vezes uma língua não tem o termo que representa com precisão o conceito para o qual um rótulo é procurado, mas tem um termo que normalmente é utilizado para representar um conceito mais amplo ou mais restrito.

Se o escopo do conceito representado por um termo cai completamente dentro do escopo do conceito representado pelo outro, os termos são ditos parcialmente equivalentes. Se a diferença no escopo for pequena o suficiente, pode ser aceitável admitir os termos para o tesouro, tratando os dois como equivalentes que representam o mesmo conceito (semelhante ao caso de termos específicos incluídos em um conceito mais amplo, dentro de um idioma).

Exemplo 3: pt: Liberdade / en: freedom, liberty

Um caso especial de equivalência parcial pode surgir quando uma das línguas oferece duas ou mais equivalências que, em combinação, representam a

totalidade do conceito representado por um termo no idioma de origem. Esta situação é por vezes descrita como “equivalência composta”.

Exemplo 4:

pt: Segurança / en: safety, security

4.2.4.2.2 PROBLEMAS E SOLUÇÕES TÍPICOS

Na situação ideal, para cada conceito, um termo preferido pode ser encontrado em cada um dos idiomas, e estes termos são equivalentes exatos no discurso normal.

Mas problemas adicionais são frequentemente causados por quase-sinônimos ou homógrafos não identificados em um ou mais idiomas, ou por completa ausência de um termo para representar um conceito em uma ou mais línguas. Raramente há uma solução correta única. É importante ponderar os méritos das alternativas em cada caso e considerar o que é melhor para a maioria das situações de recuperação.

4.2.4.2.2.1 PROBLEMAS CAUSADOS POR QUASE-SINÔNIMOS OU HOMÓGRAFOS

Se o problema é causado por homografia, assim como para a decisão de se adotar ou não um termo como preferido na língua portuguesa, deve-se seguir as regras previstas em 4.2.2.1, adicionando qualificadores quando apropriado. A mesma solução também pode ser aplicada mesmo quando os termos têm significado muito mais próximo, na medida em que poderiam ser tratados como quase-sinônimos.

4.2.4.2.2.2 AUSÊNCIA DE UM TERMO EQUIVALENTE EM UM OU MAIS IDIOMAS

Às vezes, nenhum termo em um idioma pode ser encontrado, mesmo que parcialmente, mas conseguimos representar um conceito que é necessário no tesauro. Os termos do Tesauro Unesp têm como equivalentes na língua inglesa, termos advindos da Library of Congress Subject Heading (LCSH) e do Medical Subject Heading (MeSH). Quando não

são encontrados termos equivalentes nessas linguagens deve-se criar um termo equivalente (inventado) na língua inglesa para o termo novo em português que será incorporado no tesouro. A pesquisa pode ser realizada em fontes de dados confiáveis, como vocabulários controlados e dicionários da área. Também se recomenda consulta entre indexadores, especialistas em idiomas e especialistas no assunto.

Termos criados/inventados podem ser criados das seguintes maneiras (não em ordem de preferência).

- a) Por tradução literal do termo do idioma de origem ou seus componentes semânticos;
- b) Por construção de um termo ou frase que expresse o significado geral do termo em linguagem de origem;
- c) Pela invenção de um neologismo, que deve ser o mais conciso possível para promover a aceitação (essas invenções às vezes se aproximam de traduções literais).

Para todos os casos a fonte de dados utilizada para a tradução do termo deve ser indicada.

4.2.5 RELAÇÃO ENTRE CONCEITOS

Além da relação de equivalência, existem vários outros tipos de relacionamentos que podem ser estabelecidos em um tesouro.

4.2.5.1 O RELACIONAMENTO HIERÁRQUICO

O relacionamento hierárquico deve ser estabelecido entre um par de conceitos quando o escopo de um deles cair completamente dentro do escopo do outro. Deve basear-se em graus ou níveis de superordenação e subordinação, onde o conceito superordenado representa uma classe ou um todo, e conceitos subordinados referem-se a seus membros ou partes.

As seguintes tags devem ser usadas, reciprocamente:

TG (ou seja, termo amplo, genérico), escrito como um prefixo para o termo superordenado;

TE (ou seja, termo mais restrito, específico), escrito como um prefixo para o termo subordinado.

Exemplo 1:

Vertebrados

TE Mamíferos

Mamíferos

TG Vertebrados

A relação hierárquica pode ser de três tipos, correspondendo a três situações logicamente diferentes a seguir:

- a) o relacionamento genérico;
- b) a relação hierárquica de todo/parte;
- c) o relacionamento de instância.

Todo conceito subordinado deve pertencer à mesma categoria inerente como seu conceito superordenado, tanto o termo mais amplo e mais restrito deve representar uma coisa, ou uma ação, ou uma propriedade, etc.

Exemplo 2:

- a) “metais” (uma classe de materiais) e “fundição” (uma ação) representam diferentes tipos de conceito e, portanto, não pode ser relacionado hierarquicamente;
- b) “metais” e “ouro” representam materiais e podem, portanto, estar relacionados hierarquicamente.

A principal função das relações hierárquicas é ajudar tanto os indexadores quanto os pesquisadores a escolher um nível de especificidade.

Uma busca pode ser ampliada ou reduzida movendo-se para cima ou para baixo, respectivamente, a hierarquia.

4.2.5.1.1 AS RELAÇÕES GENÉRICAS

O relacionamento genérico é o link entre uma classe ou categoria e seus membros ou espécies.

Exemplo:

Alguns membros da classe “pássaros” são conhecidos como “papagaios”, e todos os “papagaios”, por definição e independentemente do contexto, são considerados “aves”. Este exemplo geralmente garante que um termo como “papagaios” não seja subordinado a uma classe como “animais de estimação”, como nem todos os papagaios são animais de estimação, ou seja, alguns membros da classe “animais de estimação” são “papagaios”, e apenas alguns papagaios são considerados animais de estimação. Esses termos não devem, portanto, receber uma ligação TG / TE.

Observação: Ao indexar um trabalho em “papagaios como animais de estimação”, os termos preferidos que representam esses dois conceitos devem ser atribuídos.

4.2.5.1.2 O RELACIONAMENTO HIERÁRQUICO DE TODO/PARTE

A relação hierárquica de todo/parte cobre uma gama limitada de situações em que uma parte de um entidade ou sistema pertence unicamente a um todo possuidor particular.

Isso se aplica a quatro classes principais de termos:

a) Sistemas e órgãos do corpo;

Vasos sanguíneos

Artérias

Veias

b) Disciplinas ou campos do discurso;

Língua portuguesa

Linguística

Fonética

c) Estruturas sociais hierárquicas.

Ciência militar

Exércitos

Soldados

A maioria dos outros casos da relação de todo/parte não é elegível para uma ligação hierárquica porque poderia pertencer a mais de um todo. Por exemplo, uma relação TG / TE não deve ser estabelecida entre “bicicletas” e “rodas” porque uma roda pode fazer parte de um automóvel, um carrinho de mão ou um de muitos outros artefatos. Uma pesquisa avançada por bicicletas recuperaria muito material indesejado se fosse estendido a todos tipos de roda.

4.2.5.1.3 O RELACIONAMENTO DE INSTÂNCIA

O relacionamento de instância vincula um conceito geral, como uma classe de coisas ou eventos e uma instância individual dessa classe, que é frequentemente representada por um nome próprio.

Exemplos:

Regiões montanhosas { classe

Alpes {instâncias

Himalaia

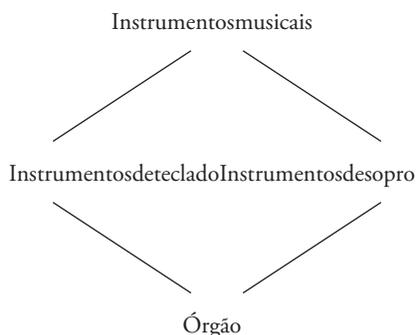
Neste exemplo, “Alpes” e “Himalaia” são atribuídos a posições subordinadas em uma hierarquia. No entanto, eles não são nem tipos nem partes de “regiões montanhosas”, mas representam instâncias individuais.

4.2.5.1.4 RELAÇÕES POLI-HIERÁRQUICAS

Alguns conceitos podem pertencer, em bases lógicas, a mais de um grupo ou classe ao mesmo tempo. Em tais casos, uma ligação hierárquica direta deve ser estabelecida para todos os conceitos mais amplos apropriados e a estrutura do tesouro é dita poli-hierárquica. Isso contrasta com uma estrutura mono-hierárquica que, para um conceito particular, permitiria que apenas um dos links “TG” válidos fosse estabelecido.

Exemplo 1:

Figura 1- Traduzido de INTERNATIONAL STANDARD (2011, p. 62)

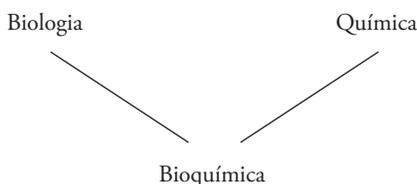


Fonte: INTERNATIONAL STANDARD (2011, p. 62).

No exemplo 1, o termo “órgãos” é atribuído a posições subordinadas com base em sua relação genérica com dois termos mais amplos. Em outros casos, links polihierárquicos podem ser baseados em relacionamentos inteiros.

Exemplo 2:

Figura 2: Traduzido de INTERNATIONAL STANDARD (2011, p. 62)

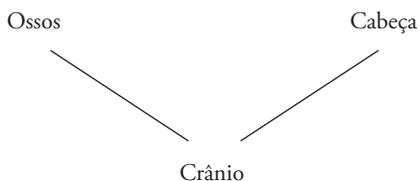


Fonte: INTERNATIONAL STANDARD (2011, p. 62).

Em alguns casos, os links poli-hierárquicos podem ser baseados em relacionamentos logicamente diferentes.

Exemplo 3:

Figura 3: Traduzido de INTERNATIONAL STANDARD (2011, p. 62)



Fonte: INTERNATIONAL STANDARD (2011, p. 62).

No Exemplo 3, o link entre “ossos” e “crânio” é baseado no relacionamento genérico (o crânio é uma espécie de osso), enquanto a ligação entre “cabeça” e “crânio” é baseada na relação hierárquica de todo/parte (o crânio é uma parte da cabeça).

Onde um conceito tem mais de um conceito mais amplo, todos os relacionamentos aplicáveis devem ser exibidos. Como opção, a diferença no tipo de relacionamento pode ser explicitamente mostrada usando as tags apropriadas.

Exemplo 4: Crânio

TG Ossos

Esqueleto

4.2.5.2 O RELACIONAMENTO ASSOCIATIVO

O relacionamento associativo abrange associações entre pares de conceitos que não estão relacionados hierarquicamente, mas são semanticamente ou conceitualmente associados a tal ponto que a ligação entre eles precisa ser explícita no tesauro, sob a alegação de que pode sugerir termos adicionais ou alternativos para uso em indexação ou recuperação. O relacionamento é indicado pela tag “TR” (termo relacionado) e deve ser aplicado reciprocamente.

Aves

TR Ornitologia

Ornitologia

TR Aves

Uma diretriz geral para relacionamentos associativos é que sempre que o termo preferido para um conceito é usado, o outro deve ser sempre aplicado dentro dos quadros comuns de referência compartilhados pelos usuários do tesouro. Além disso, um dos termos é muitas vezes um componente necessário em qualquer explicação ou definição do outro; o termo “aves”, por exemplo, constitui uma parte necessária da explicação da “ornitologia”.

É particularmente importante estabelecer um relacionamento associativo entre os conceitos que se sobrepõem no escopo.

4.2.5.2.1 TERMOS E CONCEITOS COM SIGNIFICADOS SOBREPOSTOS

No discurso ordinário, um par de termos pode ser usado de forma intercambiável em alguns contextos, mas não em outros. Por exemplo, pode-se dizer que os termos “navios” e “barcos” representam conceitos diferentes porque embora tenham muito em comum, nenhum deles cobre adequadamente o escopo do outro. Se dois conceitos são inseridos no tesouro e representados por termos preferenciais distintos, é essencial fornecer uma relação associativa entre os dois. Isso lembra os pesquisadores que eles podem precisar usar os dois termos para ter certeza de encontrar todo o material relevante.

Exemplo: Barcos

TR Navios

Navios

TR Barcos

Um par de termos preferidos com significados sobrepostos muitas vezes podem ser irmãos, ou seja, eles compartilham um termo mais amplo comum. No entanto, não é necessário inter-relacionar todos os termos irmãos dessa maneira. Por exemplo, não há necessidade de associar termos como “cavalos” e “burros”, alegando que eles compartilham termo mais amplo, por ex. “equinos”, uma vez que os significados dos termos não se sobrepõem neste caso.

4.2.5.2.2 OUTROS CASOS DE LIGAÇÃO ASSOCIATIVA

Quando um termo é fortemente implicado por outro, há motivos suficientes para associar os dois termos. Os seguintes grupos são oferecidos apenas como exemplos representativos de situações relacionais típicas e prática:

a) Uma disciplina ou campo de estudo e os objetos ou fenômenos estudados;

Exemplo 1:

Neurologia

TR Sistema Nervoso

Sistema Nervoso

TR Neurologia

b) Uma operação ou processo e seu agente ou instrumento;

Exemplo 2:

Calor

TR Controle de temperatura

Controle de temperatura

TR Calor

c) Uma ação e o produto da ação;

Exemplo 3:

Carpintaria

TR Trabalhos em madeira

Trabalhos em madeira

TR Carpintaria

d) Uma ação e seu destinatário ou destino;

Exemplo 4:

Cultivos agrícolas TR Agronomia

Agronomia

TR Cultivos agrícolas

e) Objetos ou materiais e suas propriedades definidoras;

Exemplo 5:

Venenos

TR Compostos bioativos

Compostos bioativos

TR Venenos

f) Um artefato e suas partes, se não se qualificarem para o relacionamento hierárquico de todo/parte;

Exemplo 6: Armas

TR Armaduras

Armaduras

TR Armas

g) Conceitos ligados por dependência causal;

Exemplo 7:

Temperatura corporal TR Febre

Febre

TR Temperatura corporal

h) Um objeto ou processo e seu agente contrário;

Exemplo 8:

Inflamação

TR Agentes anti-inflamatórios

Agentes anti-inflamatórios

TR Inflamação

i) Um conceito e sua unidade de medida;

Exemplo 9:

Tempo - Medição TR Relógios

Relógios

TR Tempo - Medição

j) Um termo composto e o substantivo que é seu foco, se os dois não
tiverem um

relacionamento hierárquico verdadeiro;

Exemplo 10:

Governança pública

TR Administração pública

Administração pública

TR Governança pública

k) Um organismo ou substância criado ou derivado de outro.

Exemplo 11:

Morfina

TR Ópio

Ópio

TR Morfina

REFERÊNCIAS

FUNDAÇÃO GETÚLIO VARGAS. *Manual de cabeçalhos de assunto: normas e procedimentos: versão 1.0*. Rio de Janeiro: FGV, 1995.

INTERNATIONAL STANDARD. ISO 25964-1: *Information and documentation: thesauri and interoperability with other vocabularies: part 1: thesauri for information retrieval*. Switzerland: ISO, 2011. p. 21-67.

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

5.1.2.2 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

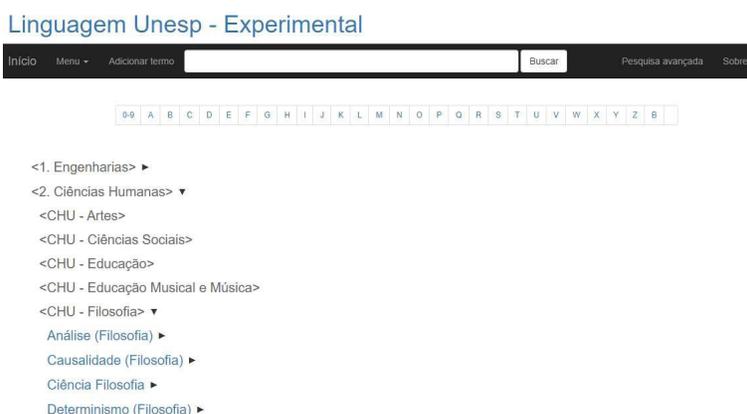
1. Engenharias (13)	2. Ciências Humanas (11)	3. Ciências Sociais aplicadas (14)	4. Ciências Biológicas (2)	5. Ciências Exatas e da Terra (7)
ENG - Engenharia Aeronáutica e Aeroespacial	CHU - Artes	CSA - Administração	CBA - Biociências	CET - Ciências da Computação
ENG - Engenharia Civil	CHU - Ciências 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

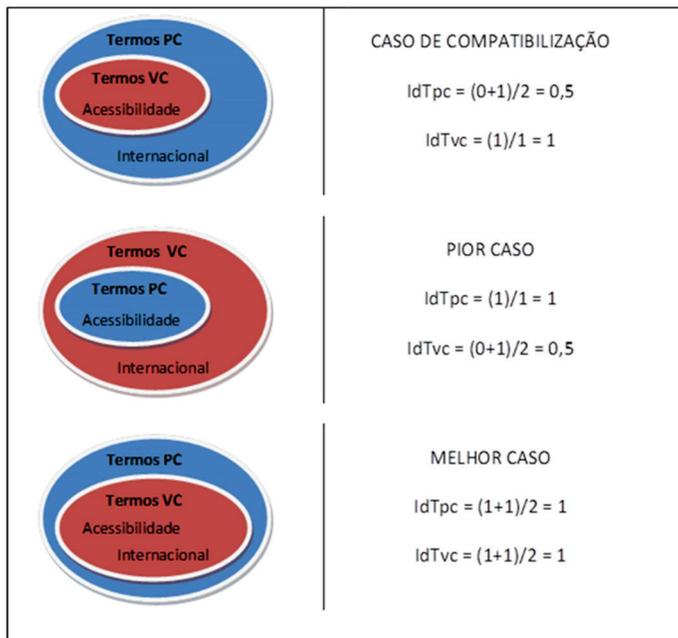
CSU - Enfermagem	Enfermeiros
CHU - Filosofia	Filósofos
CET - Geologia	Geólogos
CET - Química	Químicos
CSA - Comunicações	Radiação

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 termo 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 ►

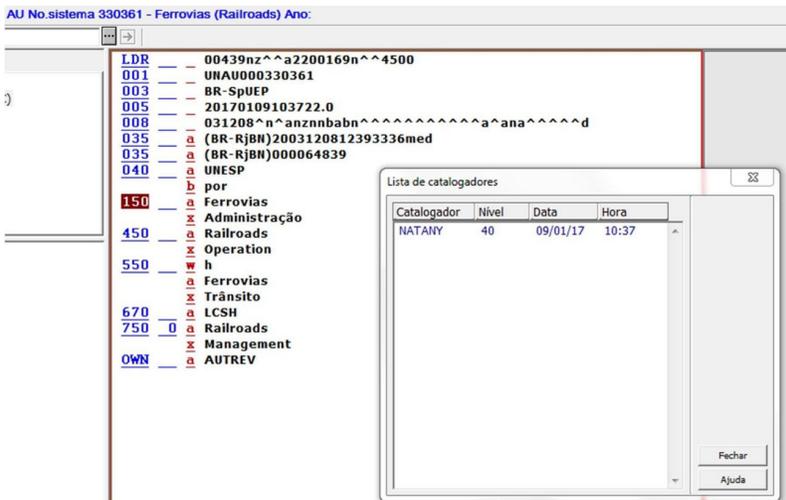
Fonte: Elaborado pelos autores.

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.

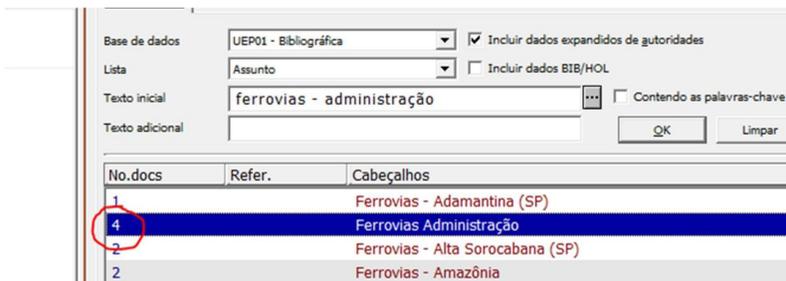
Como exemplo, o termo tópico Ferrovias – Administração foi incorporado conforme figura 4 pela catalogadora e na Figura 5 se constata que 4 registros bibliográficos estão vinculados a esse descritor.

Figura 4 – Registro de autoridade de incorporação do descritor



Fonte: Software Aleph – Módulo de Catalogação

Figura 5 – Vinculação de registros bibliográficos ao descritor

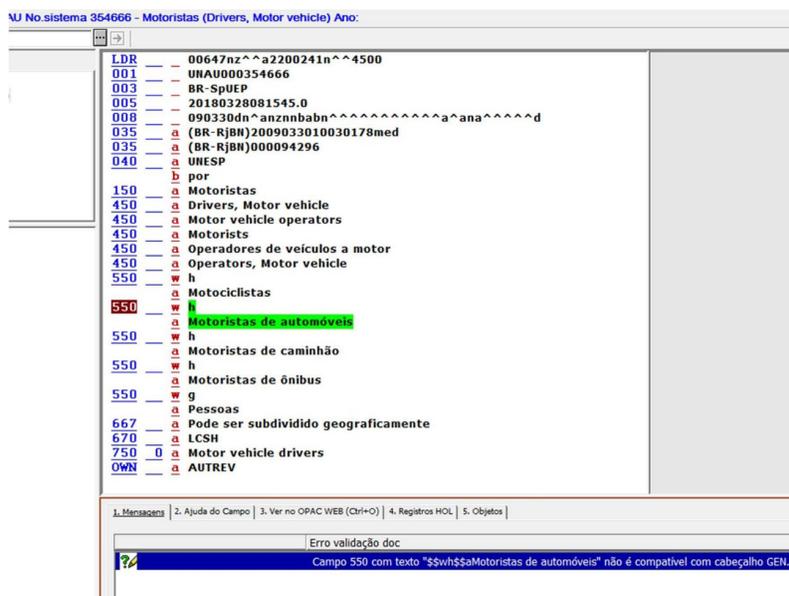


Fonte: Software Aleph – Módulo de Catalogação

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



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

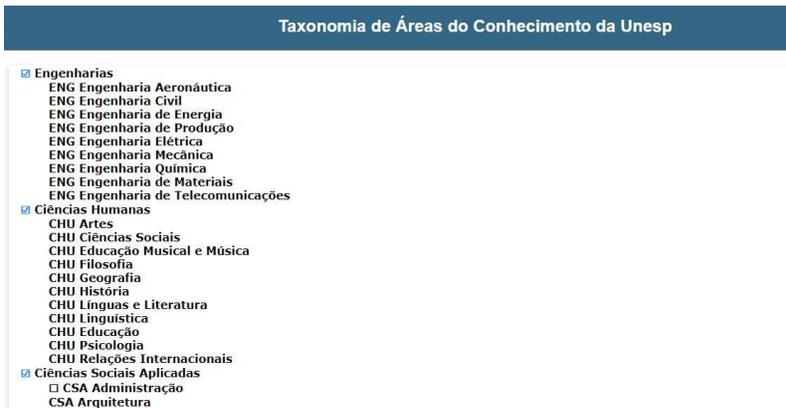
orientações são as mesmas para o campo 550 com texto “\$\$\$atermo” (termos relacionados por associação e não hierarquia).

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	w h a Mecatrônica	termo remissiva ver também (relação hierárquica ou associativa)
550	w h a Movimentos mecânicos	termo remissiva ver também (relação hierárquica ou associativa)
550	w h g Engenharia	termo remissiva ver também (relação hierárquica ou associativa)
550	a Máquinas	
667	aPode ser subdividido geograficamente	
670	a LCSH	Fonte positiva de dados
680	IUsado 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	aMechanical Engineering	Ligação de cabeçalhos autorizados (Termo da LC)
MET	ENG – Engenharia Mecânica	Metatermo
OWN	aAUTREV	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.

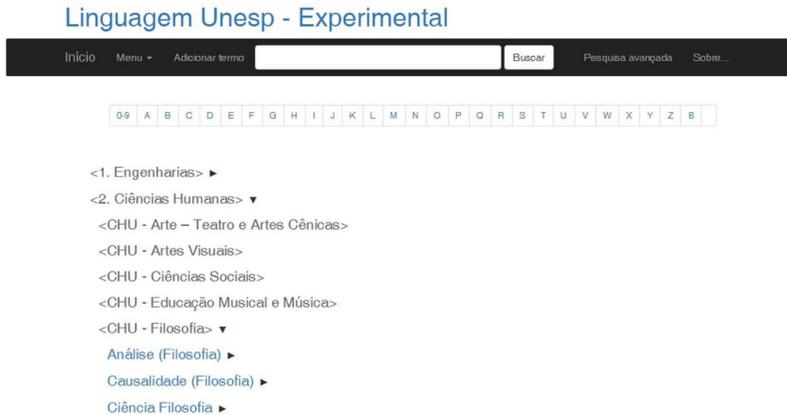
REFERÊNCIAS

- 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 topé](#) [Cancelar](#)

118 termo/s encontrados na busca *filosofia*.

Tipée 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.

6. PLANEJAMENTO DA ESTRUTURA E OPERACIONALIZAÇÃO DO TESAURO NO TEMATRES

Walter Moreira

José Carlos Francisco dos Santos

Luciana Beatriz Piovezan Rio Branco

Cibele Araújo C. M. dos Santos

Deise Maria Antonio Sabbag

Érica Fernanda Vitorini

- 6.1 Introdução
- 6.2 Estrutura do registro de autoridades
- 6.3 Apresentação e leiaute
- 6.4 A escolha do software de gerenciamento do tesauro
- 6.5 Formato de intercâmbio: registros de autoridade em formato MARC
- 6.6 Interoperabilidade com outras linguagens de indexação

Referências

Apêndice D - Manual de conversão registros de autoridades em formato MARC21
- XML para texto etiquetado

6.1 INTRODUÇÃO

Dentre os aspectos que caracterizam o tesauro, duas características podem ser destacadas como elementos que demandam o seu planejamento cuidadoso: o controle do vocabulário (pela identificação de termos preferidos, de relações de equivalência entre os termos, de padronização etc.) e a estruturação do vocabulário (pela identificação e representação dos conceitos e das relações conceituais no conjunto de termos que o compõe).

Considerando-se o caráter dinâmico inerente às linguagens, de modo geral, e os aspectos pragmáticos da linguagem documentária, de modo particular, a construção e a manutenção de um tesauro é, inevitavelmente, um processo contínuo, que irá requerer esforço e atenção permanentes.

O processo de construção e de manutenção do Tesauro Unesp vem sendo realizado desde 2013 pelo Grupo de Linguagem da Rede de Bibliotecas da Unesp. O Tesauro Unesp foi criado para uso na interface de busca integrada de todas as bases de dados da Universidade. O processo teve início com a utilização da base de registros de autoridades de assuntos da Lista de Cabeçalhos de Assuntos da Rede BIBLIODATA (LCARB) na base de dados Aleph do catálogo Athena. Esse movimento propiciou a compatibilização dos registros de autoridades de assuntos com três outras linguagens: Library Congress Subject Headings (LCSH), Terminologia da Biblioteca Nacional (TBN) e Medical Subject Headings (MeSH) para áreas da Saúde (SANTOS; FUJITA; MOREIRA, 2018).

O Grupo de Linguagem da Rede de Bibliotecas da Unesp, formado por catalogadores e pesquisadores ligados a essa instituição, foi criado e capacitado para lidar com as questões que envolviam a inclusão de termos novos no que ainda se chamava de “Linguagem Unesp”, de modo a promover a substituição gradativa da LCARB adotada anteriormente. Em suas diversas fases, o processo de construção do Tesauro UNESP foi subsidiado teórica e metodologicamente pelo desenvolvimento de variadas investigações, dentre as quais podem ser citadas: a compatibilização de registros de autoridades (FUJITA; PIOVEZAN; SANTOS, 2017; FUJITA; RUBI, 2015), a elaboração de estruturas hierárquicas (FUJITA; SANTOS, 2016; FUJITA et al., 2018) e a análise e seleção de software de gestão de tesouros (FUJITA et al., 2017).

Tendo em vista a complexidade dos procedimentos que envolvem a produção e manutenção de um tesauro, ainda mais quando se trata de um instrumento complexo e com um grande número de entradas, como é o caso do Tesauro Unesp, faz-se necessário organizar de modo sistemático os procedimentos adotados. Nesse sentido, o objetivo deste manual é orientar os procedimentos metodológicos de integração semiautomática dos registros de autoridades da Unesp com o TemaTres, software de gestão de vocabulários utilizado no processo.

6.2 ESTRUTURA DO REGISTRO DE AUTORIDADES

Os pontos de acesso por assuntos são construídos a partir da criação de um registro de autoridade de assunto por meio de um arquivo de autoridade em formato Machine Readable Cataloging - MARC21 (FUJITA; PIOVEZAN; SANTOS, 2017).

Conforme estabelece o MARC21 (2005), o registro de autoridade consiste no estabelecimento de formas padronizadas de nomes (pessoais, corporativos, eventos, jurisdição, título uniforme, combinações nome/título) e assuntos (termos tópicos, nomes geográficos, nomes com subdivisão de assunto, termos e subdivisão de assunto), sendo ambos utilizados como pontos de acesso. Considerando-se isso, foram padronizadas as entradas principais e secundárias no registro bibliográfico com o termo “nome” para entrada principal e secundária e o termo “assunto” no catálogo online construído por meio de software de gestão e manutenção de catálogos (FUJITA; PIOVEZAN; SANTOS, 2017).

O registro de autoridade de assunto é composto por indicações de relações e, sempre que necessário à clareza da delimitação do termo, são complementados com notas de escopo. As principais funções do registro de autoridade de assunto são: a) fornecer o termo de assunto autorizado ou referir ao termo autorizado; b) indicar se um termo pode ou não ser subdividido geograficamente; c) fornecer uma nota de escopo, quando houver necessidade de explicar a cobertura conceitual do termo e d) indicar relações entre termos (FUJITA; SANTOS, 2016). Para fins de ilustração, apresenta-se exemplo de registro de autoridade de assunto de termo tópico do software Aleph em formato OPAC de visualização web e no registro MARC21, do módulo de catalogação (Figuras 1 e 2).

O registro de autoridade define a forma autorizada do cabeçalho no sistema de informação do assunto principal. Também chamada de “termo tópico” ou “assunto tópico”, esse registro compõe o campo MARC 150 do registro de autoridades de um documento. Registram-se também as formas não autorizadas, ou seja, as que não são estabelecidas como assunto principal, mas que podem ser considerados como termos equivalentes (remissivas) e que são dispostas no campo MARC 450 do registro de autoridades. A função da rede de remissivas é promover a conexão precisa entre a busca do usuário e a recuperação efetiva da informação, de modo que o usuário não tenha que pensar em uma sucessão de possíveis termos relacionados ao documento buscado. Com esse recurso, o usuário não precisa buscar necessariamente pelo termo principal, pois o sistema opera tanto com o termo principal quanto com as suas remissivas. Assim, se um usuário pesquisa um assunto que é uma remissiva, o sistema vai rastrear o termo tópico no qual a remissiva está registrada, sendo interligado diretamente com o registro bibliográfico, pois o campo 150 do registro de autoridade é o equivalente ao campo MARC 650 (Assunto – termo tópico) do registro bibliográfico, que contém o termo de indexação (FUJITA; PIOVEZAN; SANTOS, 2017).

Outro tipo de relação entre termos que é indicada no registro de autoridade de assunto é a que ocorre entre o termo genérico e o termo específico, por meio da qual é indicada a posição do termo na hierarquia de assuntos. Um termo genérico é indicado no campo MARC 550 do registro de autoridades, dispondo-se o código “g” no subcampo “w”. Como essa relação é recíproca e assimétrica, o sistema deve ser configurado para incluir a relação inversa automaticamente. Esse tipo de relação entre termos é útil ao catalogador como subsídio à decisão sobre qual termo de assunto indicar como ponto de acesso ao documento, bem como ao usuário, que poderá navegar pela hierarquia de termos para decidir em que nível de generalidade o assunto é mais relevante ao seu interesse (FUJITA; PIOVEZAN; SANTOS, 2017).

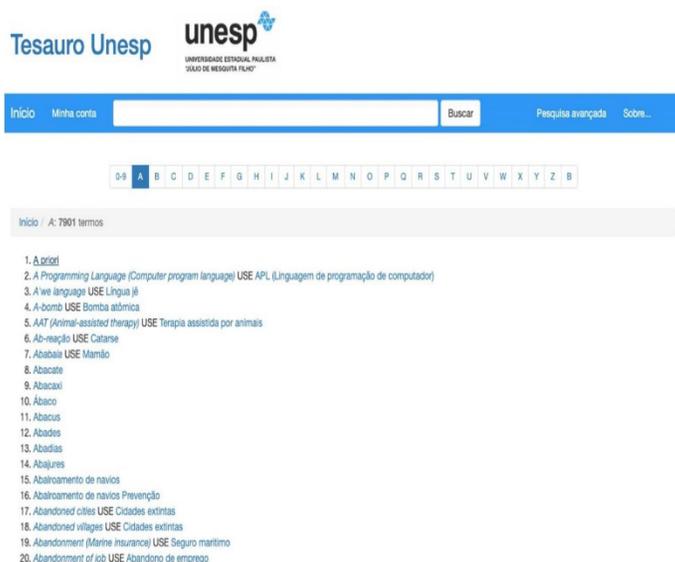
Indica-se também no registro de autoridade as conexões entre termos relacionados, ou termos associados. Nesse tipo de relação vinculam-se entre si os termos que possuem alguma conexão diferente das relações anteriores (de equivalência e de hierarquia). Essa relação também é indicada

no campo MARC 550 subcampo “a” do registro de autoridades (FUJITA; PIOVEZAN; SANTOS, 2017).

6.3 APRESENTAÇÃO E LEIAUTE

Há, tradicionalmente, dois modos básicos de exibição de um tesouro: um arranjo alfabético e um arranjo sistemático. O arranjo alfabético (Figura 3) permite localizar e acessar os conceitos a partir das palavras inicialmente expressas pelo usuário, com função semelhante à de um índice. O arranjo sistemático (Figura 4) permite visualizar os outros conceitos com os quais o conceito selecionado mantém relação na estrutura do tesouro.

Figura 3 – Apresentação alfabética do Tesouro Unesp



Fonte: Página web do Tesouro Unesp

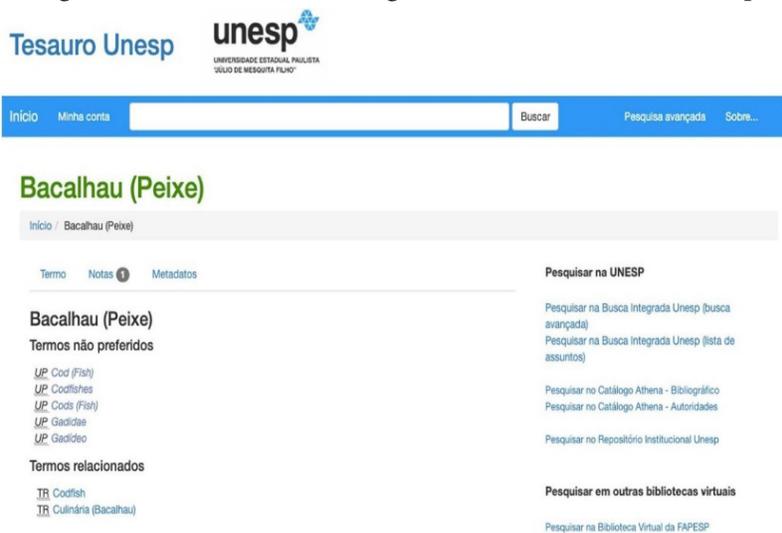
Figura 4– Apresentação sistemática do Tesauro Unesp



Fonte: Página web do Tesauro Unesp

A seleção de um determinado termo, seja a partir do arranjo alfabético seja a partir do arranjo sistemático, exibe a visualização do registro único, mostrando o termo preferido, o(s) termo(s) não preferidos e quaisquer outras relações ou notas acrescentadas ao termo em tela (Figura 5). Nesse modo de visualização apresenta-se também, do lado direito da tela, a integração do tesauro com os diversos sistemas de catálogos online da Unesp e de outras bases.

Figura 5 – Visualização de registro único no Tesouro Unesp



Fonte: Página web do Tesouro Unesp

Somam-se a todos os procedimentos de interoperabilidade dos registros de autoridades entre os softwares Aleph e o TemaTres os resultados quantitativos a partir da execução importação ao TemaTres, conforme o detalhamento apresentado na subseção 5.4. No Quadro 1 apresenta-se o total dos termos distribuídos por níveis, sendo os níveis caracterizados pelas relações hierárquicas que cada termo possui.

Quadro 1 – Quantidade de termos por níveis hierárquicos no Tesouro Unesp

Níveis	Quantidade de termos
1	16764
2	4877
3	2908
4	1356
5	693
6	397

7	393
8	240
9	131
10	49
11	10
12	2

Fonte: Elaborado pelos autores (2019).

No cômputo total, foram importados 89.581 termos, com 46.988 relações entre eles. Destes termos 40.456 são termos preferidos e 49.255 são termos não preferidos. Foram acrescentadas 16.657 notas de escopo aos registros de autoridade importados da base de dados do Aleph.

6.4 A ESCOLHA DO SOFTWARE DE GERENCIAMENTO DO TESAURO

O software TemaTres, em função das características apresentadas na sequência, foi escolhido para a gestão e manutenção do Tesauro Unesp com a importação dos registros de autoridade compatibilizados. O Tesauro Unesp está atualmente disponível junto à interface de pesquisa integrada das bases de dados da Unesp¹.

A disponibilização do Tesauro Unesp para consulta é parte fundamental da Política de Indexação das bibliotecas Universitárias da Unesp² (FUJITA, 2016). Considera-se que a intermediação do tesauro nas ações de representação de informação pelo indexador e de busca de informação pelo usuário melhora o processo como um todo, com ganhos significativos na recuperação da informação, tornando-a mais precisa e mais eficaz.

O TemaTres é um software livre³ desenvolvido por Diego Ferreyra, na Argentina, em 2004 (RODRÍGUEZ GAIRÍN; RUSSO GALLO; SULÉ DUESA, 2008). Trata-se de um aplicativo web que pode

¹ http://parthenon.biblioteca.Unesp.br:1701/primo_library/libweb/action/search.do

² <https://www.biblioteca.Unesp.br/portal/arquivos/manual-politica-indexacao-2017.pdf>

³ Software livre é uma denominação que indica que depois de adquirido o software pode ser usado, copiado, estudado, modificado e redistribuído livremente.

ser empregado no gerenciamento de vocabulários, tesouros, taxonomias e representações formais de conhecimento. O TemaTres permite desenvolver estruturas de navegação web e pode funcionar como complemento de um sistema de gerenciamento de bibliotecas, para gerenciar a linguagem documentária adotada pela instituição.

Atualmente em sua versão 3.1, o TemaTres está disponível para download gratuito na página web do projeto⁴, na qual também é possível encontrar um tesouro de demonstração construído com o software. A interface de uso do TemaTres é multilíngue, incluindo os idiomas espanhol, alemão, francês, inglês, italiano e português.

De acordo com o manual de utilização do software, a instalação do TemaTres requer (TEMATRES, 2019): um servidor web que suporte PHP; uma instalação de PHP em funcionamento com versão 4.3.0 ou superior e um servidor de bases de dados. O processo de instalação é relativamente simples, com três passos básicos: a) descarregamento (download) e descompactação do arquivo de instalação; b) configuração da conexão com a base de dados MySQL no arquivo /include/db.tematres.php e c) cumprimento das instruções de instalação do arquivo.

O TemaTres possibilita a criação de linguagens orientadas para o controle de vocabulário e possibilita: a) descrever e controlar, em número ilimitado, diversas relações de equivalência (USE-UP), relações hierárquicas (TG-TE) e relações associativas (TR) entre os termos; permite ainda a criação de novos tipos de relações entre os termos; b) aplicar ferramentas de consistência que possibilitam o estabelecimento de relações entre termos equivalentes, não equivalentes e parcialmente equivalentes; c) atribuir múltiplas notas aos termos, tais como notas de aplicação, de escopo, notas explicativas, históricas, bibliográficas etc.; d) a identificação do status dos termos em relação à sua função no conjunto de termos (termo candidato, termo aceito, termo rejeitado); e) navegar entre termos, de modo sistemático ou alfabético, com sugestão de termos de busca; f) o desenvolvimento e a integração de vocabulários polihierárquicos; g) a exportação completa do vocabulário em formatos que visam à interoperabilidade entre linguagens controladas, por meio de padrões tais como: RDF SKOS-Core; Dublin Core (ISO 158362003) e BS 8723 (Structured Vocabularies for Information Retrieval), entre outros;

⁴ <https://www.vocabularyserver.com/>

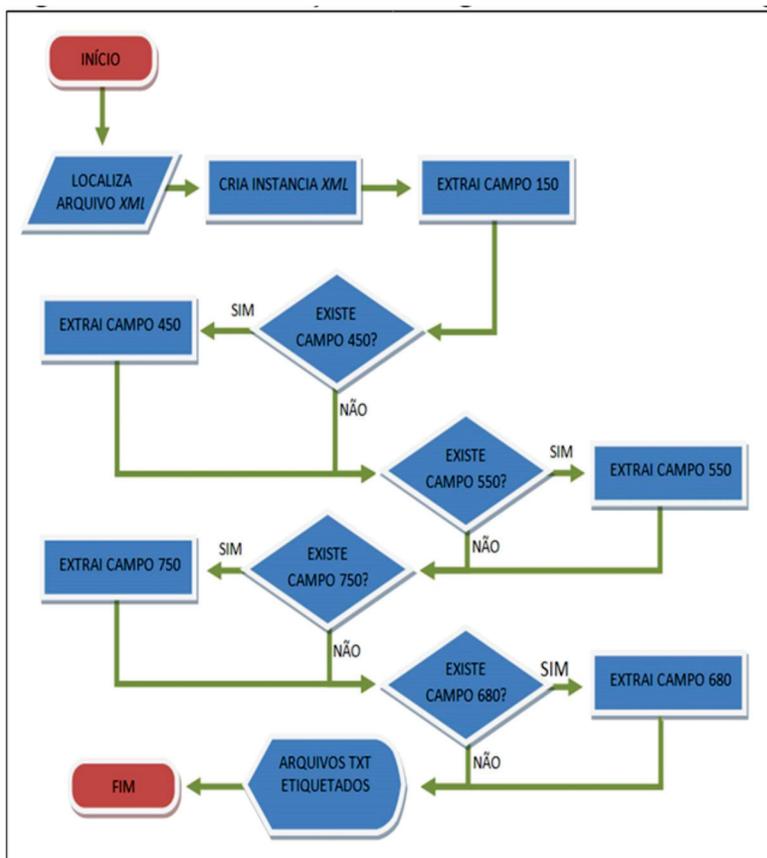
h) a interligação de dados entre vocabulários por meio do gerenciamento de suas relações; i) a auditoria de termos, isto é, a busca por termos livres e repetidos; j) a administração de usuários e editores, considerando-se os níveis de acesso ao sistema; k) a produção de relatórios estatísticos, tais como: média de palavras por termo, termos por número de termos genéricos, termos por número de termos específicos e termos por nível de profundidade.

6.5 FORMATO DE INTERCÂMBIO: REGISTROS DE AUTORIDADE EM FORMATO MARC

A amostragem dos registros de autoridade foi realizada por meio de consulta ao banco de dados do software Aleph, regulados pela aplicação de filtros de registros compatibilizados em que constavam registros bibliográficos vinculados. A identificação dos padrões de tags exportadas no formato MARC21 em suporte XML foi realizada a partir de amostragens em números menores de registros.

O planejamento e desenvolvimento do algoritmo de interoperabilidade para estruturação dos registros resultou da identificação dos padrões, sendo necessária a análise dos registros na íntegra neste formato. Foi estruturado um diagrama de blocos para organizar a extração dos termos e suas relações. Na Figura 6 apresenta-se a sistematização do algoritmo. Este tratamento automático foi desenvolvido em linguagem PHP, para receber e carregar o arquivo XML. A partir dos dados carregados, passou-se a identificação dos termos tópicos e, sequencialmente, à extração das relações (SANTOS; FUJITA; MOREIRA, 2018).

Figura 6 - Sistematização em diagrama de blocos do algoritmo



Fonte: Santos; Fujita; Moreira (2018)

A estruturação dos registros de autoridades em linguagem de importação do TemaTres é realizada a partir da execução do algoritmo proposto na Figura 6. Optou-se pela linguagem etiquetada do TemaTres no suporte de arquivo .txt. Após este tratamento automático, foram gerados os arquivos em .txt com limitações de mil registros em cada arquivo (SANTOS; FUJITA; MOREIRA, 2018).

A linguagem de exportação do sistema de gerenciamento do acervo, neste caso o Aleph, é o formato MARC21 e em suporte de arquivo

.xml dos registros de autoridade. A Figura 7 ilustra um exemplo de um registro nesse formato (SANTOS; FUJITA; MOREIRA, 2018).

Figura 7 - XML em formato MARC dos registros de autoridade

```
1 <?xml version = "1.0" encoding = "UTF-8"?>
2 <collection xmlns="http://www.loc.gov/MARC21/slim"
3 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4 xsi:schemaLocation="http://www.loc.gov/MARC21/slim
5 http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
6 <record xmlns="http://www.loc.gov/MARC21/slim"
7 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
8 xsi:schemaLocation="http://www.loc.gov/MARC21/slim
9 http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
10 <leader> nz 2200217n 4500</leader>
11 <controlfield tag="001">UNAU000204154</controlfield>
12 <controlfield tag="003">BBD</controlfield>
13 <controlfield tag="005">20170912154512.0</controlfield>
14 <controlfield tag="008">960517 n anznbnabn a ana d</controlfield>
15 <datafield tag="040" ind1=" " ind2=" ">
16 <subfield code="a">UNESP</subfield>
17 <subfield code="b">por</subfield>
18 </datafield>
19 <datafield tag="150" ind1=" " ind2=" ">
20 <subfield code="a">Projeto estrutural</subfield>
21 </datafield>
22 <datafield tag="360" ind1=" " ind2=" ">
23 <subfield code="i">Como subcabecalho usar Projetos e construção sob tipos de estruturas</subfield>
24 </datafield>
25 <datafield tag="550" ind1=" " ind2=" ">
26 <subfield code="w">g</subfield>
27 <subfield code="a">Projetos de engenharia</subfield>
28 </datafield>
29 <datafield tag="550" ind1=" " ind2=" ">
30 <subfield code="a">Projeto arquitetônico</subfield>
```

Fonte: Santos; Fujita; Moreira (2018).

O registro é marcado com a tag <record>, nas linhas com a tag <controlfield> descrevem-se as características do registro. As tags <datafield> apresentam a estrutura dos termos. No Quadro 2 apresentam-se as codificações dos registros de autoridades em MARC21 (SANTOS; FUJITA; MOREIRA, 2018).

Quadro 2 - Descrições dos termos em codificação MARC2

Código	Descrição	Exemplo tag XML	Etiqueta TemaTres
150	Termo tópico – termo utilizado	<datafield tag="150" ind1=" " ind2=" ">	-
450	Equivalência	<datafield tag="450" ind1=" " ind2=" ">	UF
550	Relacionados	<datafield tag="550" ind1=" " ind2=" ">	-
w	h – Específico	<subfield code="w">h</subfield>	NT
w	g – Genérico	<subfield code="w">g</subfield>	BT
750	Equivalente na LC	<datafield tag="750" ind1=" " ind2="0">	RT
680	Nota de escopo	<datafield tag="680" ind1=" " ind2=" ">	NA

Fonte: Santos; Fujita; Moreira (2018).

No Quadro 2 foram apresentados os códigos e as tags equivalentes no formato MARC21 exportado no suporte de arquivo .xml. Pode-se perceber que os atributos das tags são as indicações de referência do termo. No primeiro exemplo <datafield tag="150" ind1=" " ind2=" ">, a expressão datafield demonstra que os dados do campo serão descritos; logo na sequência o atributo tag indica a codificação a que se refere esse termo, neste caso seu valor é 150, indicando um termo tópico. Subordinadamente é apresentado a tag <subfield code="a"> Termo exemplo </subfield>, ilustrado na Figura 7. O subfield com o atributo code valorado em a indica o conjunto de caracteres que representam o termo, neste caso o termo é um Termo exemplo (SANTOS; FUJITA; MOREIRA, 2018).

Nos testes realizados para determinação do padrão de dados dos termos em .xml, percebeu-se a existência de termos compostos e representados em dois ou mais subfield, na Figura 8 é possível visualizar, na linha 209 e 210 do XML e no texto etiquetado está assinalado com # os dois subcampos (SANTOS; FUJITA; MOREIRA, 2018).

Figura 8 - Representação XML e texto etiquetado – termos compostos

<pre>195 <record xmlns="http://www.loc.gov/MARC21/slim" 196 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" 197 xsi:schemaLocation="http://www.loc.gov/MARC21/slim 198 http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"> 199 <leader> nz 2200217n 4500</leader> 200 <controlfield tag="001">UNAU000208442</controlfield> 201 <controlfield tag="003">BBD</controlfield> 202 <controlfield tag="005">20170912100903.0</controlfield> 203 <controlfield tag="008">950327dn anznbabn a ana d</control 204 <datafield tag="040" ind1=" " ind2=" "> 205 <subfield code="a">UNESP/BGI</subfield> 206 <subfield code="b">por</subfield> 207 </datafield> 208 <datafield tag="150" ind1=" " ind2=" "> 209 <subfield code="a">Aço</subfield> 210 <subfield code="x">Estruturas</subfield> 211 </datafield> 212 <datafield tag="450" ind1=" " ind2=" "> 213 <subfield code="a">Estruturas de aço</subfield> 214 </datafield> 215 <datafield tag="450" ind1=" " ind2=" "></pre>	<p>Aço#Estruturas (UNAU000208442) UF: Estruturas de aço UF: Structural steel BT: Engenharia civil BT: Materiais de construção BT: Vigas RT: Construção metálica RT: Pontes metálicas RT: Steel,Structural</p>
--	---

Fonte: Santos; Fujita; Moreira (2018).

Em outros casos, as variações ocorrem nos atributos das tags indicando as referências conforme apontado no Quadro 2. Portanto, o resultado dessa atividade centrou-se na coleta dessas informações, as quais por sua vez, não são tão perceptíveis visualmente. Adicionalmente, alguns testes foram realizados para que se pudesse observar o comportamento dos registros. Na Figura 9, apresenta-se a interface de importação do TemaTres, referindo-se a um dentre os padrões de importação. O texto etiquetado foi selecionado para este estudo (SANTOS; FUJITA; MOREIRA, 2018).

Figura 9 - Padrão de importação no TemaTres

Tesouro Unesp unesp
UNIVERSIDADE ESTADUAL PAULISTA
"SOLDO DE MESQUITA FILHO"

Início Minha conta Buscar Pesquisa avançada

Importar tesouro de um arquivo txt tabulado

Selecionar formato: Texto etiquetado

Arquivo: Escolher arquivo Nenhum arquivo selecionado

Salvar

Texto tabulado:

```
South America
  Argentina
    Buenos Aires = Bs As
  Brazil
  Uruguay
```

Texto etiquetado:

```
South America
ET: America
NT: Argentina
UF: South-america
RT: Latin America
```

Fonte: Santos; Fujita; Moreira (2018).

Os subsídios dessa atividade de importação darão suporte para a programação do algoritmo de transição de linguagem, interoperabilidade, a partir da extração e conversão em texto etiquetado (SANTOS; FUJITA; MOREIRA, 2018).

A partir da execução da etapa anterior passa-se para a organização e implementação do código do algoritmo de interoperabilidade na linguagem PHP. Alguns requisitos são elementos necessários para a compilação do código fonte e análise dos testes. Para tanto, utilizou-se ferramenta de edição de programas em PHP (Komodo Edit 10) e servidor Apache (EasyPHP – Devserver 17). O algoritmo é executado em servidor local simulando a web. Na Figura 10, é apresentada parte do código do algoritmo, que por sua vez sistematiza a proposta ilustrada na Figura 6 (SANTOS; FUJITA; MOREIRA, 2018).

Figura 10 - Codificação em PHP do algoritmo

```
27 foreach ($records as $record) {
28     ++$cont;
29
30     // echo 'controledfield: '. $record->controlfield[0];
31     $id=$record->controlfield[0];
32     $datafields= $record->datafield;
33     foreach ($datafields as $datafield){
34     switch ($datafield->attributes()-> tag){
35         case '150':
36             if (isset ($datafield->subfield[0])) {
37                 $termot= $datafield->subfield[0];
38             }
39             if (isset ($datafield->subfield[1])) {
40                 $termot.= ' '.$datafield->subfield[1];
41             }
42             //$termot.= ' ('.$id.')';
43             $texto.="\\r\\n". $termot;
44             break;
45         case '450':
46             $termoUF= 'UF: ';
47             if (isset ($datafield->subfield[0])) {
48                 $termoUF.= $datafield->subfield[0];
49             }
50             if (isset ($datafield->subfield[1])) {
51                 $termoUF.= ' '.$datafield->subfield[1];
52             }
53             $texto.= "\\r\\n". $termoUF;
54             break;
55     }
```

Fonte: Santos; Fujita; Moreira (2018)

Nas linhas 27 e 33 da Figura 10, ocorrem as sistematizações dos registros do arquivo XML, de maneira a proporcionar a navegabilidade entre os registros para extração dos termos e suas relações. A verificação da qualificação do termo é realizada nas linhas 35 e 45 da mesma figura.

O trabalho de estruturação dos registros de linguagens de importação do TemaTres é o resultado da execução do desenvolvimento do algoritmo de interoperabilidade. Como resultado apresenta-se, na Figura 11, a codificação etiquetada, a partir dos resultados do algoritmo (SANTOS; FUJITA; MOREIRA, 2018).

Figura 11 - Registro no formato etiquetado

```
2 Placas (Engenharia)
3 UF: Disks (Mechanics)
4 UF: Panels
5 UF: Structural plates
6 BT: Placas e cascas elásticas
7 BT: Análise estrutural (Engenharia)
8 RT: Cascas (Engenharia)
9 RT: Plates (Engineering)
10 NA: (UNAU000201540) Usado para trabalhos sobre placas como estruturas de engenharia. Para trabalhos sobre cascas
11
12 Trajes História
13 UF: Indumentária medieval
14 UF: Trajes medievais
15 UF: Costume, Medieval
16 RT: Costume History
17 NA: (UNAU000201549)
18
19 Contos espíritas
20 BT: Ficção espírita
21 RT: Spiritual short stories
22 NA: (UNAU000201603)
23
24 Tabeua caraíba
25 UF: Caraibeira
26 UF: Caroba-do-campo
27 UF: Cinco-em-rama
28 UF: Craibeira
29 UF: Ipá-amarelo-do-cerrado
30 UF: Para-tudo
```

Fonte: Santos; Fujita; Moreira (2018).

A identificação dos registros aparece etiquetada como NA (nota de aplicação) e alguns exemplos de sua utilização podem ser visualizados nas linhas 10, 17 e 22 da Figura 11. Em cada arquivo .txt, são incluídos 1000 registros e conseqüentemente nomeados com a terminação de enumeração baseada no controle do último registro nele incluído, por exemplo: arquivo_texto_1000.txt, arquivo_texto_2000.txt, arquivo_texto_3000.txt e assim sucessivamente até o término dos registros. Essa medida foi utilizada para não congestionar a submissão do arquivo e para que a execução pudesse ser mais rapidamente efetuada no servidor local (SANTOS; FUJITA; MOREIRA, 2018).

O resultado da etapa final deste trabalho é a submissão, a partir dos resultados gerados pelo algoritmo de interoperabilidade, dos arquivos .txt etiquetados conforme os padrões estabelecidos para importação de termos apresentados na Figura 11. Isso foi realizado com o recurso da função de importação dos arquivos conforme a interface do Tema Três para tal procedimento, localizada no menu de administração do vocabulário controlado, ilustrado na Figura 9 (SANTOS; FUJITA; MOREIRA, 2018). A execução dessas atividades para obtenção da importação é descrita no Apêndice C, no Manual de conversão registros de autoridades em formato MARC21 - XML para texto etiquetado.

6.6 INTEROPERABILIDADE COM OUTRAS LINGUAGENS DE INDEXAÇÃO

Zeng e Chan (2004) utilizam a definição elaborada pela American Library Association (2000) para conceituar a interoperabilidade como a capacidade de diferentes sistemas ou componentes de trocarem informações e de utilizarem essas informações sem esforço especial por parte dos sistemas. Assim entendida, a interoperabilidade é a capacidade de sistemas distintos se comunicarem com informações transparentes por meio da admissão de padrões comuns e protocolos que viabilizem o compartilhamento e o uso da informação (MOURA, 2009).

Na literatura podem-se encontrar vários termos conceitualmente próximos ao conceito de interoperabilidade, tais como: compatibilização, unificação, integralização, interoperabilidade, convertibilidade. Ainda que não seja objetivo deste capítulo esclarecer tais diferenças, é importante registrar que, na condição de termos pertencentes ao mesmo campo conceitual, tais expressões são utilizadas algumas vezes como sinônimas, o que pode causar prejuízos à compreensão do processo.

Para que a interoperabilidade ocorra, faz-se necessário o estabelecimento das camadas de representação estrutural, sintática, semântica e lógica: Unicode/URI; XML/Namespace/XML Schema; RDF/RDF Schema; Ontológica; Lógica; Prova e Confiança. A camada ontológica realiza a integralização terminológica articulando a terminologia visando minimizar os conflitos. Ou seja, é necessário que seja estabelecido um método de compatibilização de linguagens, no caso, um método de compatibilização de linguagens de indexação.

O método de compatibilização (ou medida de compatibilização) realiza a similaridade entre duas ou mais linguagens introduzindo o conceito de graus de compatibilidade e fazendo com que se estabeleça a distribuição entre compatibilidade no plano semântico e no plano linguístico.

Conforme Campos (2017), a compatibilidade é a qualidade de articulação entre vocabulários afins (direta ou indiretamente) definindo equivalências conceituais entre os termos por meio de relações de semelhanças.

Assim, para que a compatibilização aconteça é necessário: criar uma matriz de compatibilização (mapeamento das potencialidades semânticas e linguísticas das linguagens); identificar a taxa de coincidência

verbal (análise da medida de similaridade entre o símbolo linguístico e seu conteúdo conceitual); e estabelecer o grau de compatibilidade conceitual (coincidência conceitual, correspondência conceitual e correlação conceitual).

Neste sentido e orientado por essa abordagem, o processo de construção do Tesouro Unesp utilizou elementos de cinco linguagens existentes, a saber, Lista de Cabeçalhos de Assunto da Rede BIBLIODATA (LCARB); Terminologia da Biblioteca Nacional (TBN); Library of Congress Subject Headings (LCSH) e Medical Subject Headings (MeSH), Vocabulário Controlado da USP (VocaUSP).

A partir do recurso a essas quatro linguagens foi realizada a compatibilização dos registros conforme segue: a) registros de autoridade: a base de registros da LCARB foi internalizada no catálogo Athena possibilitando a construção da Linguagem Unesp a partir da compatibilização dos registros de autoridades existentes e atualização as informações pela TBN; b) para registros não traduzidos pela LCARB foram utilizadas as linguagens da LCSH e o MeSH.

A escolha da TBN foi realizada por meio de estudos comparativos que demonstraram a viabilidade desta linguagem em substituir a LCARB. A TBN possui maior atualidade sendo uma tradução da LCSH que tem como características uma estrutura hierárquica visível com relações associativas de tesouro (FUJITA et al., 2018).

Dessa forma, os bibliotecários catalogadores passaram a importar registros de autoridade com termos novos constantes das linguagens LCSH, TBN e MeSH. Algumas necessidades importantes foram levadas em consideração para a construção da linguagem, notadamente a atualização e a garantia de compatibilização para os registros importados da LCARB (FUJITA et al., 2018).

A linguagem desenvolvida e utilizada pela USP (VocaUSP) foi utilizada por sua semelhança de propósitos com a Linguagem Unesp e por ter sido desenvolvida com embasamento terminológico nas áreas específicas de conhecimento com as quais a USP trabalha, mesmo que esses fundamentos terminológicos sejam diferentes da LCSH (FUJITA et al. 2018).

REFERÊNCIAS

- AMERICAN LIBRARY ASSOCIATION (ALA). *Committee on Cataloging: Description and Access (CC:DA) Task Force on Metadata*. Final report. 2000. Disponível em: <http://www.libraries.psu.edu/tas/jca/ccda/reports.html>. Acesso em: 05 nov. 2019.
- BRASILEIRO, I. B.; SANTOS, J. C. F.; MOREIRA, W. TemaTres e o Koha. In: SHINTAKU, M. (org.). *Guia do usuário do TemaTres*. Brasília: Ibict, 2019. p. 121146.
- CAMPOS, M. L. A. Interoperabilidade semântica e compatibilização de linguagens em ambientes heterogêneos: a questão do acesso aberto e as possibilidades de tratamentoterminológico. In: ENCONTROS PRÉ-CONFOA: GESTÃO DE CONTEÚDO E CRIS, 3., 2017. Rio de Janeiro. *Encontros [...]*. Rio de Janeiro: Fiocruz/ICICT, 2017. p. 1-47.
- FUJITA, M. S. L. *et al.* Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias. *Scire*, Zaragoza, v. 23, n. 2, p. 7181, 2017.
- FUJITA, M. S. L. *et al.* Construction and evaluation of hierarchical structures of indexing languages for online catalogs of libraries: an experience of the São Paulo State University (UNESP). *Knowledge Organization*, Baden-Baden, v. 45, n. 3, p. 220-231, 2018.
- FUJITA, M. S. L. *Política de indexação para bibliotecas: elaboração, avaliação e implantação*. Marília; São Paulo: Oficina Universitária; Cultura Acadêmica, 2016.
- FUJITA, M. S. L.; PIOVEZAN, L. B.; SANTOS, N. S. A função do registro de autoridade de assunto na construção e uso de linguagens de indexação para catálogos online. In: SIMÕES, M. G.; BORGES, M. M. (coord.). *Tendências atuais e perspectivas futuras em organização do conhecimento: atas do III Congresso ISKO Espanha-Portugal; III Congresso ISKO Espanha*. Coimbra: Universidade de Coimbra, 2017. p. 577-586.
- FUJITA, M. S. L.; RUBI, M. Compatibilidade entre linguagens documentais para construção, atualização e adequação de vocabulário de bibliotecas universitárias. In: RODRÍGUEZ MUÑOZ, J. V. *et al.* (org.). *Organización del conocimiento: sistemas de información abiertos: actas del XII Congreso ISKO España y II Congreso ISKO España y Portugal*. Murcia: Universidad de Murcia, 2015 p. 345-356.
- FUJITA, M. S. L.; SANTOS, L. B. P. A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias. *Scire*, Zaragoza, v. 22, n. 2, p. 37-46, 2016.
- MARC21. *MARC 21: formato condensado para dados de autoridade*. Tradução e adaptação de Margarida M. Ferreira. Marília: FUNDEPE, 2005.

MOURA, M. A. Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias. *Informação & Sociedade*, João Pessoa, v. 19, n. 1, p. 59-73, jan. /abr. 2009.

RODRÍGUEZ GAIRÍN, J. M.; RUSSO GALLO, P.; SULÉ DUESA, A. A virtual exhibition of open source software for libraries, 2008. In: *BOBCATSSS SYMPOSIUM*, 16., 2008, Zadar (Croácia). Disponível em: <http://hdl.handle.net/10760/11151>. Acesso em: 13 dez. 2015.

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

TEMATRES. *TemaTres*: servidor de vocabularios controlados. 2019. Disponível em: <http://r020.com.ar/tematres/manual/>. Acesso em: 30 out. 2019.

ZENG, M. L.; CHAN, L. M. Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society or Information Science and Technology*, New York, v. 55, n. 5, p. 377–395, 2004.

APÊNDICE D

MANUAL DE CONVERSÃO REGISTROS DE AUTORIDADES EM FORMATO MARC21 - XML PARA TEXTO ETIQUETADO

Este manual tem por finalidade orientar o usuário no processo de preparação dos dados e arquivos para importação no TemaTres. A apresentação deste manual foi publicada no capítulo 5 do livro Guia do Usuário do TemaTres (BRASILEIRO; SANTOS; MOREIRA,2019). Para efeito de facilidade para o usuário está representado o passo-a-passo. O sistema de conversão tem como principal interface apresentada na Figura 1.

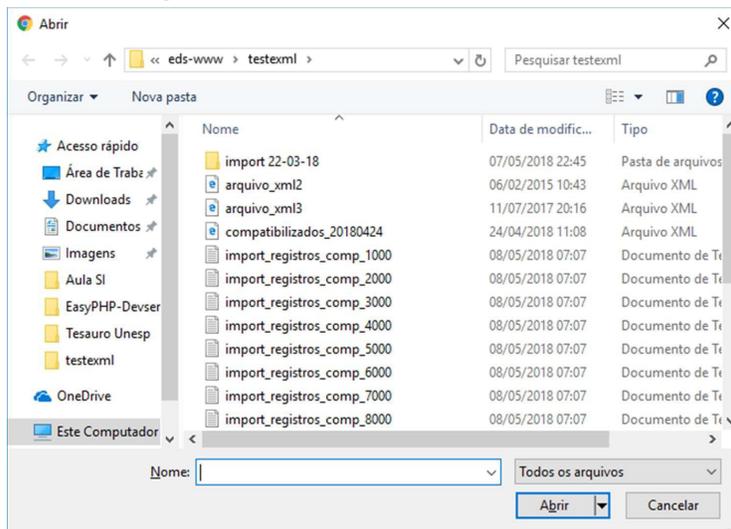
Figura 1 – Interface do sistema de conversão

Selecione o arquivo xml: Nenhum arquivo selecionado

Fonte: Brasileiro; Santos; Moreira (2019, p. 142)

Clique em “Escolher arquivo”, conforme Figura 1. Após esse procedimento será aberta a janela para selecionar o arquivo para ser carregado (Figura 2).

Figura 2 – Janela de seleção de arquivos



Fonte: Brasileiro; Santos; Moreira (2019, p.143)

Ao selecionar o arquivo clique em abrir, logo o nome do arquivo será mostrado ao lado do botão “Escolher arquivo”, apresentado na Figura 3. Para executar os processos de conversão clique em “Enviar”.

Figura 3 – Interface do sistema com a seleção do arquivo

Selecione o arquivo xml: compatibiliz...20180424.xml

Fonte: Brasileiro; Santos; Moreira (2019, p. 143)

Após o término da execução serão listados os arquivos em formato .txt, etiquetados no padrão de importação no software TemaTres, conforme Figura 4.

Figura 4 – Lista de arquivos convertidos para downloads

Arquivo enviado com sucesso!
Gerando os arquivos para o TemaTres...
Arquivo com registros até 1000: [BAIXAR](#)
Arquivo com registros até 2000: [BAIXAR](#)
Arquivo com registros até 3000: [BAIXAR](#)
Arquivo com registros até 4000: [BAIXAR](#)
Arquivo com registros até 5000: [BAIXAR](#)
Arquivo com registros até 6000: [BAIXAR](#)
Arquivo com registros até 7000: [BAIXAR](#)
Arquivo com registros até 8000: [BAIXAR](#)
Arquivo com registros até 9000: [BAIXAR](#)
Arquivo com registros até 10000: [BAIXAR](#)
Arquivo com registros até 11000: [BAIXAR](#)
Arquivo com registros até 12000: [BAIXAR](#)
Arquivo com registros até 13000: [BAIXAR](#)
Arquivo com registros até 13534: [BAIXAR](#)

Fonte: Brasileiro; Santos; Moreira(2019,p.144)

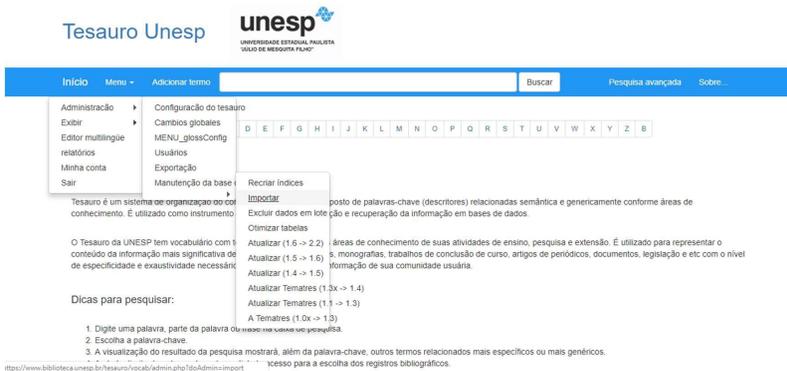
Clique em “BAIXAR” para salvar no computador. Após acessar o tesouro no endereço <https://www.biblioteca.Unesp.br/tesouro>, faça o login, conforme disponibilizado pelo departamento de TI (Figura 5).

Figura 5 – Interface do TemaTres para efetuar login



Fonte: Elabora do pelos autores, 2019 .

Uma vez logado no sistema do Tesauro Unesp, acesse o item Menu>Administração>Manutenção da Base>Importar, conforme Figura 6.



Fonte: Elabora do pelos autores, 2019.

Conforme mostra a janela de importação, apresentada na Figura 7, selecione o formato a ser importado, neste caso “Texto etiquetado”, e depois clique em “Arquivo” para abrir a janela de seleção de arquivos conforme Figura 8.

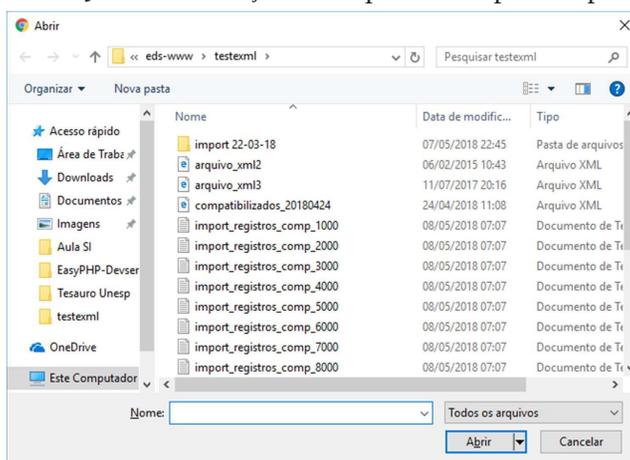
Figura 7 – Interface TemaTres para importação de registros em TXT



Fonte: Elaborada pelos autores, 2019.

Selecione o arquivo e clique em “Abrir” para executar a importação. Ao término do processo será exibida a mensagem de finalização. Fique atento para não clicar em qualquer outro botão ou clicar mais de uma vez para “Salvar”. O processo demora alguns instantes, pode ser percebido que está executando devido ao navegador representar o carregamento de uma página da web.

Figura 8 – Janela de seleção de arquivos .txt para importação



Fonte: Elaborado pelos autores, 2019.

SOBRE OS AUTORES

CIBELE ARAÚJO CAMARGO MARQUES DOS SANTOS



Doutora em Ciências da Informação e Bacharel em Biblioteconomia e Documentação pela ECA/USP. Professora do Departamento de Informação e Cultura da ECA e no Programa de PósGraduação em Ciência da Informação (PPGCI/ECA/USP) na linha de Organização da Informação e do Conhecimento. Desenvolve pesquisa sobre organização do conhecimento, linguagens documentárias, organização de arquivos, de informações e de imagens. Participa da equipe de gestão do Vocabulário Controlado da USP como assessora acadêmica. cibleac@usp.br

DEISE MARIA ANTONIO SABBAG



Bacharel em Biblioteconomia pela Universidade Estadual Paulista – UNESP de Marília. Doutora em Ciência da Informação pela Universidade Estadual Paulista (Unesp). Professora Doutora da Universidade de São Paulo, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto. É docente permanente na Pós-Graduação, linha de pesquisa Produção e Organização da Informação, do Programa de Ciência da Informação da Unesp de Marília. deisesabbag@usp.br

ÉERICA FERNANDA VITORINI



Doutora em Ciência da Informação no PPGCI na Universidade Estadual Paulista Julio de Mesquita Filho - Campus de Marília em 2020. Mestra em Ciência da Informação no PPGCI na Universidade Estadual Paulista Julio de Mesquita Filho - Campus de Marília em 2015, com o projeto Uso da linguagem documentária na busca da informação em bibliotecas universitárias: a perspectiva dos deficientes visuais. Bacharel de Biblioteconomia e Ciência da Informação na Universidade Federal de São Carlos- Campus de São Carlos em 2010. Seus interesses em pesquisa são linguagens documentais, organização do conhecimento e acessibilidade em Bibliotecas Universitárias. erica.vitorini@unesp.br

JOSÉ CARLOS FRANCISCO DOS SANTOS



Doutor em Ciência da Informação na Universidade Estadual Paulista “Júlio de Mesquita Filho” – Unesp (2020), Mestre em Ciência da Informação pela Universidade Estadual de Londrina - UEL (2015). Docente do Programa de Mestrado Profissional em Direito das Faculdades Londrina. Gerente de Projetos do Esap - Instituto de Estudos Avançados e Pós-Graduação. Como Pesquisador atua nos Grupos de Pesquisa “Análise Documentária” na Unesp, Direito, Inovação, Tecnologias e Desenvolvimento das Faculdades Londrina. Atua na área de Ciência da Informação, Organização do Conhecimento, com ênfase em Controle de Vocabulário em periódicos científicos eletrônicos. Integrante do Grupo de Linguagem da Rede de bibliotecas da Unesp.
jose.cf.santos@unesp.br

LUCIA SILVA PARRA



Graduada em História e Biblioteconomia pela Universidade de São Paulo (USP). Especialista em Gerência de Sistemas e Serviços de Informação pela Fundação Escola de Sociologia e Política de São Paulo, (FESPSP). Mestra em Estudos Culturais e doutoranda em Educação na Universidade de São Paulo (USP). Bibliotecária da Coordenadoria Geral de Bibliotecas, Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), campus de São Paulo. Integrante do Grupo de Linguagem da Rede de bibliotecas da Unesp.
lucia.parra@unesp.br

LUCIANA BEATRIZ PIOVEZAN RIO BRANCO



Doutora em Ciência da Informação na Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP com pesquisa sobre indexação de construção e gerenciamento de vocabulários controlados. Possui Mestrado em Ciência da Informação e bacharelado em Biblioteconomia obtido na mesma instituição. Seus outros interesses de pesquisa são indexação de assuntos, política de indexação e organização do conhecimento em bibliotecas universitárias. Integrante do Grupo de Linguagem da Rede de bibliotecas da Unesp. lbpiovezan@gmail.com

MARIA LUIZA DE ALMEIDA CAMPOS



Formada em Biblioteconomia e doutorado em Ciência da Informação pelo IBICT/UFRJ. Professora e pesquisadora do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal Fluminense. Líder do Grupo de Pesquisa Estudos Ônticos e Ontológicos em Contextos Informacionais. Desenvolve pesquisas em Organização e Representação do Conhecimento, Ontologia, Interoperabilidade semântica, Terminologia e Classificação. Integrante do Grupo de Linguagem da Rede de bibliotecas da Unesp. marialuizalmeida@gmail.com

MARIA MARLENE ZANIBONI



Bacharel em Biblioteconomia, especialista em Tecnologias em Informação e Gestão de Bibliotecas Públicas. Bibliotecária da Biblioteca da Universidade Estadual Paulista (Unesp) – Câmpus de Bauru. É membro do Grupo de Linguagem da Rede de Bibliotecas da Unesp da Rede de Bibliotecas da Unesp, onde colabora na com atividades relacionadas do Tesauro Unesp. marlene.zaniboni@unesp.br

MARIÂNGELA SPOTTI LOPES FUJITA



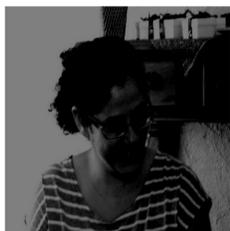
Bacharel em Biblioteconomia pela Fundação Municipal de São Carlos. Mestre e Doutora em Ciências da Comunicação pela Universidade de São Paulo, Livre Docente em Análise Documentária e Linguagens Documentárias Alfabéticas, Professora Titular da UNESP. Docente Permanente Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista – UNESP de Marília. Líder do Grupo de Pesquisa “Representação Temática da Informação”. Professora aposentada da UNESP de Marília. Coordenadora do Grupo de Linguagem da Rede de Bibliotecas da Unesp. mariangela.fujita@unesp.br

MILENA MARIA RODRIGUES PEDROZO



Bacharel em Biblioteconomia com MBA em Gestão de Unidades de Informação. Bibliotecária na Seção Técnica de Aquisição e Tratamento da Informação da Faculdade de Ciências e Letras - Campus de Araraquara da Universidade Estadual Paulista (UNESP). Atua com os cursos de Ciências Humanas, classificação, catalogação e indexação no formato MARC21 (bibliográfico e autoridade). É membro do Grupo de Linguagem da Rede de Bibliotecas da Unesp, onde colabora com os registros de autoridade e termos para o Tesouro Unesp. milena.rodrigues@unesp.br

MINERVINA TEIXEIRA LOPES



Bacharel em Biblioteconomia, especialista em Gestão de Bibliotecas Públicas e Gestão de Direção e Desenvolvimento de Pessoas, bibliotecária da Universidade Estadual Paulista (UNESP), Campus de Bauru, onde atua como supervisora da Seção Técnica de Aquisição e Tratamento da Informação. É membro do Grupo de Linguagem da Rede de Bibliotecas da Unesp, onde colabora na revisão, correção e criação de registros de autoridade para o Tesouro Unesp. minervina.lopes@unesp.br

ROSANE RODRIGUES DE BARROS RIBAS



Bacharel em Biblioteconomia, bibliotecária responsável pelo Grupo Técnico de Informação e Documentação da Reitoria da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), onde atua na área jurídica, principalmente em pesquisa de doutrina, jurisprudência e legislação e em catalogação no formato MARC21 (bibliográfico e autoridade). É membro do Grupo de Linguagem da Rede de Bibliotecas da Unesp da Rede de Bibliotecas da Unesp, onde colabora na revisão e correção de registros de autoridade e na criação de novos termos para o Tesouro Unesp.
rosane.barros@unesp.br

ROSIMARA LOFIEGO



Bacharel em Biblioteconomia pela Universidade Estadual Paulista/UNESP de Marília e Licenciada em Matemática UNESP Bauru. Trabalha desde 1994 na Biblioteca da UNESP Campus de Bauru. É membro do Grupo de Linguagem da Rede de Bibliotecas da Unesp, onde colabora na revisão e correção de registros de autoridade para o Tesouro Unesp.
rosimara.lofiego@unesp.br

TELMA JAQUELINE DIAS SILVEIRA



Bacharel em Biblioteconomia pela Unesp, atualmente é diretora da Biblioteca da Unesp Câmpus de Marília, onde atuou durante dez anos como supervisora da Seção Técnica de Aquisição e Tratamento da Informação. É membro do Grupo de Linguagem da Rede de Bibliotecas da Unesp da Rede de Bibliotecas da Unesp.
telma.silveira@unesp.br

WALTER MOREIRA



Bacharel em Biblioteconomia pelas Faculdades Integradas Teresa D'Ávila (FATEA). Mestre em Biblioteconomia e Ciência da Informação pela Pontifícia Universidade Católica de Campinas (PUCCamp). Doutor em Ciência da Informação pela Universidade de São Paulo (USP). Livre Docente em Sistemas de Organização do Conhecimento pela Universidade Estadual Paulista (UNESP). Professor do departamento de Ciência da Informação da UNESP, atuando nos cursos de graduação em Arquivística e em Biblioteconomia e na pós-graduação em Ciência da Informação. Desenvolve pesquisas sobre sistemas de organização do conhecimento, linguagens documentárias, terminologia, ontologias e teoria da classificação. walter.moreira@unesp.br

SOBRE O LIVRO

CATALOGAÇÃO

Telma Jaqueline Dias Silveira
CRB 8/7867

NORMALIZAÇÃO

Maria Elisa Valentim Pickler Nicolino
CRB - 8/8292
Mariana Macedo

CAPA E DIAGRAMAÇÃO

Gláucio Rogério de Moraes

PRODUÇÃO GRÁFICA

Giancarlo Malheiro Silva
Gláucio Rogério de Moraes

ASSESSORIA TÉCNICA

Renato Geraldi

OFICINA UNIVERSITÁRIA

Laboratório Editorial
labeditorial.marilia@unesp.br

FORMATO

16 x 23cm

TIPOLOGIA

Adobe Garamond Pro

Papel

Polén soft 70g/m2 (miolo)
Cartão Supremo 250g/m2 (capa)

TIRAGEM

100

IMPRESSÃO E ACABAMENTO

Gráfica
unesp
Campus de Marília



Um manual de construção e manutenção de tesouros para bibliotecas não é usual porque não é comum que as bibliotecas construam tesouros e os mantenham atualizados. Porém, acreditamos que essa prática é necessária tendo em vista a necessidade constante de atualização de vocabulário especializado e multidisciplinar exigido pela divulgação científica. Este livro foi idealizado nos moldes do desenvolvimento da criação do Tesouro Unesp e foi organizado de modo a espelhar o planejamento e a organização de um tesouro no contexto de bibliotecas universitárias. Assim como outras bibliotecas, as bibliotecas universitárias estão sujeitas à evolução tecnológica digital que afeta de modo especial as práticas bibliotecárias e que as obriga a se reinventar constantemente. Por isso, muitas soluções apresentadas neste livro respeitam as condições ideais dessa mudança renovadora, mas respeitam, igualmente, as suas condições de exequibilidade.

ISBN 978-65-5954-068-6



9 786559 540686