
A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias

La estructura lógica y jerárquica de lenguajes de indización utilizados por las bibliotecas universitarias

The logical and hierarchical structure of indexing languages used by university libraries

Mariângela Spotti Lopes FUJITA (1), Luciana Beatriz Piovezan dos SANTOS (2)

(1) Universidade Estadual Paulista Júlio de Mesquita Filho, Av. Hygino Muzzi Filho, 737, Campus Universitário, CEP. 17525-900, Marília, SP, Brasil, fujita@marilia.unesp.br, (2) Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista, lbpiovezan@gmail.com

Resumen

El uso de lenguajes de indización realiza con precisión la mediación entre los vocabularios de los documentos y los de los usuarios durante la recuperación de información. Dichos lenguajes, sin embargo, deben ser actualizados a medida que avanza la generación e innovación del conocimiento. Dicha actualización depende de la estructura jerárquica para facilitar su inclusión y clasificación. Para la investigación propuesta sobre las estructuras jerárquicas de lenguajes de indización se adoptó el método del comité ascendente propuesto en la Directiva ANSI/NISO de 2005 para construir vocabularios controlados usando el modelo metodológico integral, comparando las jerarquías de las Library of Congress Subject Headings (LCSH), la Terminología de Assuntos da Biblioteca Nacional (TBN), el Linguagem BIBLIODATA y el Vocabulário de la Universidad de São Paulo en las áreas de física y matemáticas. Los resultados muestran que el lenguaje Bibliodata no posee algunas relaciones jerárquicas, lo que impide su comparación. El TBN tiene ocho niveles de jerarquía para las matemáticas; y los demás lenguajes tienen jerarquías de cuatro niveles para la física. Se concluye que es posible el desarrollo de jerarquías utilizando los métodos propuestos.

Palabras clave: Lenguajes de indización. Bibliotecas universitarias. Relaciones jerárquicas. Estructuras categoriales. Actualización. Brasil.

1. Introdução

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 da biblioteca. Ocorre que usuários do catálogo bibliográfico *on-line* ATHENA (2) não conhecem e nunca utilizaram linguagem de indexação por não estar disponibilizada na web junto à ferramenta de busca. Nunca existiu, então, a mediação da linguagem no catálogo ATHENA e,

Abstract

The use of indexing languages mediates accurately between document and user vocabularies during information retrieval. Nevertheless, it should be updated accordingly to the generation and innovation of new knowledge. Updating terms depends on the hierarchical structure that guides the inclusion and classification of terms. With the research proposal about hierarchical structures of indexing languages, the bottom-up committee method proposed in the ANSI/NISO-2005 Guidelines for the compilation of controlled vocabularies using the integrated methodological model was adopted, comparing the hierarchies of the Library of Congress Subject Headings (LCSH), the Subject Terminology of National Library (TBN), the BIBLIODATA Language and the USP Vocabulary in the fields of Physics and Mathematics. The results show that BIBLIODATA language does not have certain hierarchical relationships, making any comparison impossible. TBN has eight levels hierarchy for Mathematics and the other languages have four levels for Physics. The conclusion is that it is possible to develop hierarchies using these methods.

Keywords: Indexing languages. University libraries. Hierarchical relations. Categorical structures. Updating. Brasil.

em realidade, não precisaria ser utilizada pelos catalogadores, pois a linguagem natural é a mais utilizada pelos usuários como demonstraram os resultados obtidos no primeiro estudo de avaliação realizado por Boccato, Fujita e Gil Leiva (2011), que verificou o desempenho e o índice de precisão na recuperação de áreas científicas especializadas em 3 áreas de conhecimento no catálogo *on-line* ATHENA do sistema de bibliotecas da UNESP. A análise dos dados coletados utilizou, como indicadores de avaliação da linguagem de indexação, os índices de precisão e revocação pela recuperação da informação no catálogo Athena (Lancaster, 2004; Gil Leiva, 2008).

Essa constatação pode ser a mesma de outras bibliotecas que necessitam aprimorar a recuperação por assuntos de seus catálogos *on-line* e isso poderá ser resolvido pela escolha de uma linguagem de indexação. A decisão de usar uma só linguagem deve ser tomada em função da compatibilização de vocabulário a ser utilizado na indexação e na estratégia de busca para obter resultados de recuperação mais precisos, corretos e específicos.

O uso de uma linguagem de indexação adequada e atualizada para o contexto de contínua geração de novos conhecimentos e de inovação, no caso das bibliotecas da UNESP, foi determinante para a construção da Linguagem de indexação UNESP mediante processo de adequação e compatibilização de vocabulários de três linguagens de indexação de mesma raiz terminológica utilizadas por bibliotecas: Terminologia de assuntos da Biblioteca Nacional, *Library of Congress Subject Headings* (LCSH) e *Medical Subject Headings* (MeSH). O avanço desse processo de adequação e compatibilização exige, contudo, a incorporação de termos novos conforme previsto inicialmente pela própria necessidade do contexto de geração de novos conhecimentos científicos e tecnológicos das bibliotecas universitárias.

Ocorre que a criação de termos novos depende da existência de uma macroestrutura de categorias temáticas que oriente a inclusão e classificação dos termos para que se componha a estrutura lógica e hierárquica dos termos superordenados e subordinados.

Essa macroestrutura para organizar a ordem hierárquica dos assuntos ainda não existe na Linguagem de indexação UNESP e precisa ser criada para orientar a adequação, a compatibilização e a inclusão de termos novos pela organização de registros de autoridade. Aparentemente, a Linguagem da Biblioteca Nacional (BN) e a Linguagem da *Library of Congress Subject Headings* (LCSH) não possuem uma macroestrutura de categorias temáticas que organiza a ordem hierárquica dos assuntos, ainda que, todos os termos possuam uma organização hierárquica.

Considerando-se a observação desse problema de estruturação lógica e hierárquica da Linguagem de indexação UNESP, é importante o desenvolvimento de investigação sobre o processo de elaboração da estrutura lógica e hierárquica de linguagens de indexação na perspectiva de uma política de indexação. Para isso, propõe-se o estudo da macroestrutura de categorias da linguagem LCSH, Terminologia da Biblioteca Nacional, Vocabulário USP e Linguagem do BIBLIODATA para elaboração e desenvolvimento dessa

linguagem de indexação para bibliotecas universitárias.

2. Linguagens de indexação utilizadas por bibliotecas: aspectos históricos e de estrutura lógico-hierárquica, semântica e associativa

Para avançarmos no que tange ao uso da linguagem de indexação em bibliotecas examinaremos este instrumento a partir de sua concepção como Lista de Cabeçalhos de Assuntos e sua construção.

Veremos, então, que as listas de cabeçalhos de assunto são linguagens de indexação pelos aspectos estruturais e que, por isso, evoluíram e incorporaram aspectos de tesouros, tendo em vista as grandes mudanças de acesso e de recuperação da informação em sistemas de buscas por assunto.

Tal como o tesouro, as listas de cabeçalhos de assuntos são linguagens de estrutura alfabética, considerando-se que os termos que as integram são ordenados alfabeticamente. Tanto o tesouro como as listas de cabeçalhos de assuntos, possuem entre os termos relações hierárquicas, semânticas e associativas como se observa no Quadro I:

Descrição	Horticultura (subdividido geograficamente)
Remissiva Ver (US/UF)	Hortaliças - Cultivo Hortas
Remissiva Ver Também (TR)	TE: Aclimação (Plantas) TE: Cobertura dos solos TE: Cogumelos - Cultivo TE: Estufas para cultivo TE: Frutas - Cultivo TE: Hidroponia TE: Jardinagem paisagística TE: Olericultura TE: Plantas - Propagação TE: Plantio (Cultivo de plantas) TE: Poda TE: Produtos hortigranjeiros TG: Agricultura TR: Jardinagem
Fonte positiva dos dados	LCSH
Nota geral pública	Usado para obras que tratam do cultivo, para fins comerciais, de frutas, vegetais e plantas ornamentais
Outros vocabulários	
Library of Congress	Horticulture

Quadro I. Ficha da autoridade – Termo tópico “Horticultura” da TBN

O termo “Horticultura” tem relação semântica com os termos “Hortaliças-Cultivo” e “Hortas” como termo preferido ou de equivalência, relação hierárquica como termo superordenado com os termos específicos (TE) subordinados a partir do termo “Aclimação (Plantas)” até o termo “Produtos Hortigranjeiros” e como termo subordinado

com o termo genérico (TG) “Agricultura” e, por último, relação associativa (TR) com o termo “Jardinagem”.

As relações hierárquicas entre termos superordenados e subordinados demonstram a estrutura vertical de relações lógico-hierárquicas de uma linguagem de indexação que se constitui de vértices e níveis. O vértice das cadeias hierárquicas em sistemas de classificação, segundo explica Cintra et al. (1994, p. 32), é “constituído de disciplinas convencionais que se subdividem sucessivamente”. Nos tesouros e listas de cabeçalhos de assunto, os vértices equivalem às classes e são denominados de top terms ou termos genéricos superordenados que poderão ser o gênero ou o todo.

No exemplo do Quadro II, o *top term* da cadeia hierárquica é “Agricultura”, que possui 3 níveis constituídos de 3 vértices:

. AGRICULTURA
.. Horticultura
... Aclimação (Plantas)
... Cobertura dos solos
[...]
... Produtos hortigranjeiros

Quadro II. Vértices e níveis da estrutura lógico-hierárquica do termo “Horticultura” na TBN

Para a construção, atualização e manutenção de linguagens de indexação é fundamental o conhecimento e uso da estrutura lógico-hierárquica representada pelas cadeias hierárquicas de termos superordenados e subordinados compostas de vértices e níveis com a finalidade de representação do domínio de assunto de modo completo.

Para o entendimento das linguagens de indexação utilizadas por bibliotecas passamos a examinar as listas de cabeçalhos de assunto que são continuamente atualizadas e estão disponíveis para uso.

2.1. A linguagem da LCSH: história e configuração lógico-hierárquica, semântica e associativa

A *Library of Congress Subject Headings* (Lista de Cabeçalho de assunto da Biblioteca do Congresso) começou a ser construída em 1898 para atender às necessidades de representação dos conteúdos de livros e periódicos custodiados pela Biblioteca do Congresso dos Estados Unidos (*Library of Congress - LC*). Inicialmente intitulada *Subject Headings in the Dictionary Catalogues of the Library of Congress*, o título atual foi

adotado na oitava edição da lista, publicada em 1975 (Hoshy, 2001, p. 135-136).

Torres (2006) ressalta o papel de destaque da LC na disseminação da catalogação de assuntos e das listas de cabeçalhos de assuntos, que, desde sua criação e com a comercialização de seus registros bibliográficos e de autoridade, sedimentara o uso do cabeçalho de assunto na representação do conteúdo de documentos.

Deste modo, a LCSH tem sido amplamente utilizada em bibliotecas para promoção do acesso por assunto, tendo sido traduzida e adaptada para diversos idiomas e servido como base para construção de outras linguagens de indexação.

Considerada como a principal lista de cabeçalhos de assunto existente, a LCSH, embora seja originalmente uma linguagem pré-coordenada, estão hoje organizados em formato de tesouro, com seus cabeçalhos arranjados em uma estrutura de relacionamentos de equivalência, hierárquicos e associativos (Fujita, Boccato e Rubi, 2010).

No âmbito da LCSH, os cabeçalhos de assuntos são construídos a partir da criação de um registro de autoridade de assunto em formato *Machine Readable Cataloging - MARC21*, que será incorporado à lista de autoridades que compõe a LCSH (Hoshy, 2001).

O Registro de autoridades de assunto da LCSH é composto por notas de escopo e indicações de relacionamentos, tendo, como objetivos básicos, prover o cabeçalho de assunto autorizado ou direcionar ao cabeçalho autorizado, indicar se um cabeçalho pode ou não ser subdividido geograficamente, prover uma nota de escopo quando houver necessidade de explicar a cobertura a que o termo se aplica e indicar relacionamentos entre cabeçalhos.

Como dito, a LCSH está organizada em uma estrutura de relacionamentos de equivalência, hierárquicos e associativos, que permitem aos usuários ampliar, especificar ou redirecionar suas buscas. Tais relacionamentos e suas designações são padronizados para construção e gerenciamento de tesouros (Hoshy, 2001).

Desse modo, a estrutura da LCSH compreende uma ordenação sistemática que evidencia as relações que unem os cabeçalhos de assunto que a compõem. Nessa estrutura, as relações hierárquicas, expressas pelo código BT para termo geral e NT para termo específico, são feitas para indicar que há uma relação de superordenação/subordinação entre dois ou mais cabeçalhos (Yi e Chan, 2010).

Nesse tipo de relacionamento, há a possibilidade de um cabeçalho pertencer a mais de uma hierarquia lógica, casos em que serão feitas referências a cada hierarquia à qual o cabeçalho se liga, respeitando-se a ligação apenas ao termo imediatamente superior em cada hierarquia (Hoshy, 2001).

O relacionamento de equivalência, expresso pelos códigos USE e UF (usado para), existe entre termos não-preferidos remetendo para um cabeçalho autorizado, quando dois ou mais termos designam o mesmo conceito, de maneira a atender ao princípio do cabeçalho uniforme e controlar variantes semânticas e ortográficas (Hoshy, 2001).

O relacionamento associativo, expresso pelo código RT (termo relacionado) indica termos que estão ligados de alguma forma diferente das anteriormente expressas, de modo que se relacionem cabeçalhos relativos a assuntos entre os quais há uma associação mental.

2.2. Linguagens derivadas da LCSH: terminologia da Biblioteca Nacional e linguagem BIBLIODATA em comparação com a LCSH

A Lista de Cabeçalhos de Assunto da Rede BIBLIODATA começou a ser desenvolvida no âmbito do projeto BIBLIODATA/CALCO (Catalogação Legível por Computador), um esforço conjunto da Fundação Getúlio Vargas (FGV) e do Instituto Brasileiro de Bibliografia e Documentação (IBBD), atual IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia) com o objetivo de padronizar a utilização de cabeçalhos de assunto (Histórico, [2016]).

A Linguagem BIBLIODATA é disponibilizada eletronicamente a todas as instituições credenciadas pela Rede BIBLIODATA acompanhada de um manual de procedimentos que contém a sua estrutura formal e instruções de operacionalização da linguagem.

Verifica-se que a Linguagem BIBLIODATA é uma linguagem controlada e pré-coordenada, cujos termos são arranjados alfabeticamente, composta por meio de tradução e adaptação para a língua portuguesa da LCSH. De acordo com o Manual de cabeçalhos de assunto da Rede BIBLIODATA, a opção pela LCSH deu-se por sua característica multidisciplinar, além de sua confiabilidade (Fundação Getúlio Vargas, 1995).

A lista de cabeçalhos da Linguagem BIBLIODATA é composta por cabeçalhos tópicos e subdivisões de assunto, de forma, geográficas e cronológicas, remissivas (relações de equivalência) e relacionamentos hierárquicos e associativos. O registro de autoridade de assunto da Linguagem

BIBLIODATA compõe-se pelo cabeçalho, relacionamentos, nota de escopo e termo LC – em que se indica o termo correspondente na LCSH (Fundação Getúlio Vargas, 1995).

A Terminologia da Biblioteca Nacional (TBN) remonta a 1945, quando a instituição passou a adotar os princípios da *American Library Association* (ALA) para tratamento da informação, criando catálogos de autoridades de nomes pessoais e coletivos e, conjuntamente, passou a adotar a LCSH como padrão para criação de pontos de acesso por assunto para uso interno da instituição. Foi apenas no ano de 1982, com sua entrada na Rede BIBLIODATA/CALCO que a instituição passou a construir sua própria linguagem baseada na LCSH e cujo uso “foi sistematizado não só em âmbito interno como para utilização pelas bibliotecas participantes da Rede BIBLIODATA, mantendo-se os cabeçalhos originais e gerando, assim, um vocabulário bilíngue” (GRINGS, 2016, no prelo).

Em 1994, quando a FGV optou por uma mudança de *software* operacional para a Rede BIBLIODATA, a BN deixou de participar da Rede e ambas as linguagens passaram a ser desenvolvidas separadamente, mantendo como semelhança apenas o fato de possuírem a LCSH como matriz e base terminológica (GRINGS, 2016, no prelo).

2.3. Linguagem de indexação de bibliotecas universitárias: o Vocabulário USP

O Vocabulário USP é uma linguagem construída para a indexação de assuntos da coleção das bibliotecas universitárias da USP em seu Banco de Dados Bibliográficos DEDALUS. Para atender às necessidades relativas às atividades inerentes à universidade, o Vocabulário USP constitui-se de uma linguagem interdisciplinar, que abrange as diversas áreas do conhecimento (Universidade de São Paulo, 2015).

A criação do Vocabulário USP iniciou-se com a composição, em 1993, de uma equipe de bibliotecários da USP, responsável por atualizar e ampliar a Lista de Assuntos que era utilizada até o momento. A equipe de bibliotecários contou com o apoio de professores da Escola de Biblioteconomia da USP e também da UNESP (Universidade de São Paulo, [s.d.]).

Como resultado, em 2003 foi implementado o Vocabulário USP, disponível para consulta *on-line* na página web da universidade. O vocabulário constitui-se de descritores, não-descritores, qualificadores, remissivas e termos denominados como “elos falsos”, que são termos não-descritores que atuam como elemento aglutinador de

descritores mais específicos (Universidade de São Paulo, 2006).

O Vocabulário USP está disponível para consulta externa, podendo ser visualizado por sua apresentação em lista alfabética ou hierárquica, complementada por tabelas de qualificadores, além de poder ser consultado por sua macroestrutura, em que são explicitadas suas relações lógico-semânticas. Estima-se que atualmente o Vocabulário USP contenha aproximadamente 45 mil termos (Universidade de São Paulo, 2015).

Para uso interno, o vocabulário está disponibilizado na base de dados SIBIX 650 integrada ao sistema de gerenciamento para uso na indexação de itens do acervo. O Vocabulário também pode ser utilizado por outras instituições por meio de consulta à página web institucional (Universidade de São Paulo, 2006; Universidade de São Paulo, 2015).

As quatro linguagens de indexação apresentadas foram construídas para uso em bibliotecas, sendo que apenas o Vocabulário USP tem a especificidade de haver sido desenvolvido para bibliotecas universitárias; as demais são utilizadas em diversas modalidades de bibliotecas indistintamente. Tais linguagens apresentam estrutura alfabética de organização de termos, sendo que três são linguagens pré-coordenadas (LCSH, TBN e BIBLIODATA) e uma linguagem é pós-coordenada, o Vocabulário USP. Duas linguagens, LCSH e BIBLIODATA, são denominadas como listas de cabeçalhos de assunto, mas a primeira adota características de tesauro para organização de seus termos, da mesma forma que a TBN, que segue a estrutura e a terminologia da LCSH. As quatro linguagens apresentam complementos para os termos incluídos, como listas de qualificadores, subdivisões e remissivas.

3. O estudo das macroestruturas das linguagens de indexação: método e procedimentos

A construção de linguagens de indexação possui procedimentos específicos abordados por normatizações e literatura da área. As principais diretrizes, em diferentes momentos, abordaram a construção de vocabulários controlados. São elas: a do IBICT (1984), da UNESCO (1996) e da ANSI/NISO (2005) cujas etapas podem ser observadas e analisadas a partir de síntese elaborada por Cervantes (2009, p. 200) no quadro "Etapas de construção de um tesauro de acordo com as diretrizes", que contém 9 etapas: Trabalho preliminar, métodos de compilação, Registro de termos, Verificação dos termos, Determinação da estrutura e formatos de apresentação, Níveis de Especificidade, Inclusão e Exclusão de

termos, Uso de equipamento automático de processamento de dados (IBICT e UNESCO), Termos não designados (ANSI/NISO).

Conforme análise mais detalhada das etapas no quadro de Cervantes (2009, p. 200), verificou-se que é na etapa de "Métodos de compilação" que as três Diretrizes expõem a construção das classes mais amplas, denominadas de categorias a partir das quais se constroem as hierarquias de termos genéricos e específicos. As Diretrizes do IBICT e da UNESCO dividem os métodos em dedutivo, indutivo e combinação de métodos e a ANSI/NISO amplia para 4 os tipos de métodos, sendo abordagem de comitê, abordagem empírica, combinação de métodos e assistência da máquina.

Os métodos indutivo e dedutivo são considerados pela Diretriz ANSI/NISO (2005) como empíricos e estão incluídos em seu método de abordagem empírica. De forma geral, são descritos com conteúdo semelhante nas três diretrizes com enfoque para a orientação de inclusão de termos em categorias que devem ser identificadas pelos termos genéricos de primeiro nível hierárquico antes (indutivo) ou depois (dedutivo) da compilação dos termos. A combinação de métodos indutivo e dedutivo propõe método que incorpora o dedutivo e o indutivo de modo que sugere o estabelecimento das categorias dos termos de modo indutivo por um grupo editorial composto por indexadores e especialistas e depois de as hierarquias de termos serem examinadas a partir da perspectiva dedutiva.

A abordagem de comitê, sugerida pela diretriz ANSI/NISO (2005), propõe a atuação de um comitê composto por especialistas no domínio dos assuntos e especialistas em criação de vocabulários e identifica dois métodos para a atuação desse comitê: Top-down (descendente) e Bottom-up (ascendente). No descendente, os termos mais genéricos ou superordenados são identificados inicialmente e depois são construídas as hierarquias de termos subordinados ou mais específicos para atingir o nível de profundidade hierárquica necessário. No ascendente, é mais característica a situação de vocabulários controlados já iniciados que precisam incorporar termos derivados de novos conteúdos e, nesse caso, a estrutura de categorias e as relações hierárquicas vão sendo construídas ao contrário: iniciam-se pelos termos mais específicos e avançam para os termos superordenados. Desse modo, o método descendente é mais indicado para a construção de novos vocabulários que resultarão em linguagens de indexação e o ascendente, para linguagens de indexação já publicadas e que necessitam do acréscimo de novos termos para sua atualização e manutenção.

Portanto, para a situação desta proposta de acréscimo de termos novos combinados com a necessidade de criar uma estrutura de categorias e suas hierarquias para o conjunto de termos já existentes da Linguagem de indexação da UNESP, será adotado para a compilação o método de comitê ascendente, considerando-se que os termos específicos já existem e as hierarquias de termos necessitam da identificação das categorias.

Para a identificação das categorias serão analisadas as estruturas de categorias das linguagens de indexação da LCSH, da Terminologia de Assuntos da Biblioteca Nacional, do BIBLIODATA e

do Vocabulário USP utilizando o “Modelo Metodológico Integrado para construção de Tesouro” de Cervantes (2009, p. 164).

Observa-se que na coluna da direita estão os procedimentos terminográficos para cada etapa da construção do tesouro (coluna da esquerda) que aqui generalizamos para linguagem de indexação. Nosso enfoque será a aplicação dos procedimentos terminográficos propostos para o método de compilação de abordagem de comitê ascendente com a finalidade de construir é uma estrutura de categorias para a Linguagem UNESP e, para isso, definimos o desenvolvimento de suas etapas e fases:

<i>Modelo metodológico integrado para construção de tesouro</i>	
Sistematização de etapas da construção de tesouros (normalização, literatura e tesouros) - Procedimentos terminográficos	
1. Trabalho preliminar (Orientações gerais/ Uso de equipamento automático de processamento de dados)	Escolha do domínio e da língua do tesouro Delimitação do subdomínio Estabelecimento dos limites da pesquisa terminológica temática; Consulta a especialista do domínio/subdomínio
2. Método de compilação (Abordagem de compilação)	Coleta do corpus do trabalho terminológico Estabelecimento da árvore de domínio Expansão da representação do domínio escolhido
3. Registro de termos	Coleta e classificação de termos.
4. Verificação de termos (Admissão e exclusão de termos /Especificidade)	Verificação, classificação e confirmação de termos Elaboração de definições Uso do vocabulário de especialidade para o estabelecimento de relações entre os descritores e de relações entre descritores e não descritores Organização das relações entre descritores
5. Forma de apresentação de um tesouro	Trabalhos de apresentação do tesouro.

Quadro III. Sistematização das etapas de construção de tesouros

6. Coleta do *corpus* do trabalho terminológico

- Fase 1: identificação e descrição das linguagens de indexação selecionadas para análise do domínio da Física e da Matemática conforme apresentado nos itens 2.1, 2.2 e 2.3;
- Fase 2: identificação e descrição das estruturas de representação documentária das áreas de Física e da Matemática nas linguagens de indexação selecionadas;

7. Estabelecimento da árvore de domínio

- Fase 3: identificação e descrição das hierarquias de termos dos domínios de Física e Matemática das linguagens de indexação;

8. Expansão da representação de domínio

- Fase 4: elaboração da estrutura lógico-hierárquica do domínio da Física e da Matemática, tendo, como parâmetro, as estruturas das linguagens de indexação.

4. Análise e discussão sobre o estudo das macroestruturas das linguagens de indexação

Neste tópico será abordada a aplicação do “Modelo Metodológico Integrado para construção de Tesouro”, proposto por Cervantes (2009) relativamente à etapa Método de compilação, descrevendo o desenvolvimento das fases 2, 3 e 4.

4.1. Coleta do *corpus* do trabalho terminológico

Fase 2: Identificação e descrição das estruturas de representação documentária das áreas de Física e da Matemática nas linguagens de indexação selecionadas

Foram analisadas as linguagens de indexação identificadas previamente de modo a compilar e estruturar os termos relativos às áreas de Física e Matemática, agrupados em uma planilha, com o objetivo de identificar a sua estrutura terminológica, visto que apenas o Vocabulário USP disponibiliza de forma estruturada em sua macroestrutura. Por meio da identificação dos termos de primeiro nível hierárquico em ambas as áreas temáticas, discriminaram-se as categorias de assunto de cada linguagem analisada.

Na LCSH, a categoria da área de Física é indicada pelo termo *Physics*, ao qual estão agregados 67 termos imediatamente subordinados, e a categoria da área de Matemática é indicada pelo termo *Mathematics*, ao qual estão relacionados 77 termos imediatamente subordinados.

A TBN identifica a área da Matemática pelo termo Matemática, ao qual foram relacionados 45 termos subordinados, já o termo Física identifica a área de Física, ao qual são subordinados 47 termos.

No Vocabulário USP, a área de Física é identificada pela categoria CE520 – FÍSICA, posteriormente subdividida em 20 termos genéricos, e a área da Matemática é identificada pela categoria CE550 MATEMÁTICA, subdividida em 65 termos genéricos.

Diante do identificado nesta fase, em que se verificou que as três linguagens analisadas indicam as categorias de Física e de Matemática por estes termos, sugere-se que a Linguagem UNESP adote esta nomenclatura como denominação das categorias aqui tratadas.

4.2. Estabelecimento da árvore de domínio

Fase 3: identificação e descrição das hierarquias de termos dos domínios de Física e Matemática das linguagens de indexação

A partir do agrupamento dos termos relativos às áreas de Física e Matemática existentes nas linguagens de indexação analisadas, foi possível o estabelecimento de uma estrutura proveniente das categorias principais, identificadas anteriormente.

A estruturação dos termos em uma linguagem tem por objetivo a caracterização do domínio e se faz por meio da classificação dos termos nas

categorias e indicação dos relacionamentos entre estes termos (Fujita, 2011).

Observando-se a estrutura hierárquica das linguagens de indexação, verificou-se que na LCSH foi possível encontrar termos subordinados que avançaram até o quarto nível hierárquico para ambas as categorias analisadas.

LCSH	
TG Physics	TG Mathematics
TE Agricultural physics	TE Algebra
TE Soil physics	TE Axioms
TE Soil absorption and adsorption	TE Axiomatic set theory
TE Soil colloids	TE Axiom of choice
TE Archaeological physics	TE Banach-Tarski paradox

Quadro IV. Exemplo da hierarquia das categorias Física e Matemática na LCSH

Exemplo disso é o termo “*Soil colloids*”, que é um termo subordinado de quarto nível hierárquico, cujo termo imediatamente superordenado é “*Soil absorption and adsorption*”, que, por sua vez, é específico de “*Soil physics*”, todos pertencentes à hierarquia “*Agricultural physics*” na categoria Física. Um exemplo da categoria Matemática é “*Banach-Tarski paradox*” também subordinado de quarto nível hierárquico, que pertence à hierarquia de “*Axioms*”, completado pelos termos “*Axiomatic set theory*” e “*Axiom of choice*”.

Na Terminologia da Biblioteca Nacional, foram identificados termos subordinados que avançaram até o quarto nível hierárquico para a área de Física e até o oitavo nível hierárquico para a área de Matemática.

Na categoria Física, o termo “Articulações – Hipermobilidade” demonstra essa afirmação. Observa-se que é um termo subordinado de quarto nível hierárquico que se relaciona imediatamente ao termo superordenado “Articulações - Amplitude de movimento”, por sua vez termo específico de “Biomecânica” que completa a hierarquia do termo “Biofísica”. Para a categoria de Matemática, tem-se o exemplo da hierarquia do termo “Cálculo vetorial”, composta por oito níveis hierárquicos demonstrados nos termos subordinados “Cálculo tensorial”, “Espaços generalizados”, “Banach, Espaços de”, “Banach, Álgebra de”, “Análise harmônica”, “Harmônicos esféricos” e do termo “Grassmann, Teoria da extensão de”.

TBN	
TG Física	TG Matemática
TE Biofísica	TE Cálculo vetorial
TE Biomecânica	TE Cálculo tensorial
TE Articulações – Amplitude de movimento	TE Espaços generalizados
TE Articulações - Hipermobilidade	TE Banach, Espaços de
TE Calor	TE Banach, Álgebra de
TE Calor animal	TE Análise harmônica
TE Controle de temperatura	TE Harmônicos esféricos
TE Capilaridade	TE Grassmann, Teoria da extensão de

Quadro V. Exemplo da hierarquia das categorias Física e Matemática na TBN

A análise do Vocabulário USP identificou termos subordinados que avançaram até o quinto nível hierárquico na categoria Física e até o quarto nível na categoria referente à Matemática.

Vocabulário USP	
CE520 FÍSICA	CE550 - MATEMÁTICA
CE520.12 - MECÂNICA DOS LÍQUIDOS	CE550.1 – ÁLGEBRA
CE520.12.1 - HIDRÁULICA (MECÂNICA DOS LÍQUIDOS)	550.16 - ANÁLISE FUNCIONAL
CE520.12.2 - HIDROMECAÂNICA	CE550.16.51 - OPERADORES
CE520.12.2.1 - HIDROSTÁTICA	CE550.16.51.17 - OPERADORES LINEARES
CE520.12.2.1.1 - DENSIDADE DOS LÍQUIDOS	CE550.16.51.17.2 - C-SEMIGRUPOS

Quadro VI. Exemplo da hierarquia das categorias Física e Matemática no Vocabulário USP

Na categoria CE520 FÍSICA, tal estrutura pode ser vista no termo “CE520.12.2.1.1 - DENSIDADE DOS LÍQUIDOS”, termo mais específico subordinado à “CE520.12.2.1 – HIDROSTÁTICA”, que é um termo subordinado à “CE520.12.2 - HIDROMECAÂNICA”, cujo termo genérico é “CE520.12.1 - HIDRÁULICA (MECÂNICA DOS LÍQUIDOS)”. Já na categoria “CE550 - MATEMÁTICA”, é possível demonstrar a hierarquia do termo “CE550.16 - ANÁLISE FUNCIONAL”, ao qual o termo “CE550.16.51 - OPERADORES” é subordinado, chegando ao termo “CE550.16.51.17 - OPERADORES LINEARES”, ao qual o termo “CE550.16.51.17.2 - C-SEMIGRUPOS” se relaciona no quarto nível de profundidade.

Verificou-se que a linguagem da Rede BIBLIODATA não possui relacionamentos hierárquicos determinados, constituindo-se efetivamente como uma lista de cabeçalhos de assunto e, portanto, não há como compará-la com as demais

linguagens que possuem essa estrutura hierárquica estabelecida. Dessa forma, de modo a garantir a consistência e adequação da Macroestrutura da Linguagem Unesp, não é indicada a utilização da linguagem da Rede BIBLIODATA neste processo.

4.3. Expansão da representação de domínio

Fase 4: Elaboração da estrutura lógico-hierárquica do domínio da Física e Matemática tendo como parâmetro as estruturas das linguagens de indexação

Com as comparações expostas nos Quadros 6 (Física) e 7 (Matemática) em Apêndices na versão eletrônica, é possível observar que, entre as três estruturas, há poucas semelhanças entre os vértices por causa da estrutura do Vocabulário USP, porém entre as estruturas da LCSH e TBN existem várias semelhanças. Nas três estruturas de Física, existe semelhança, como, por exemplo, com o vértice “Física médica” e na Matemática com o vértice “Álgebra”. Com relação às diferenças entre vértices, elas existem somente na LCSH, como, por exemplo, “Física Agrícola” e “Física Arqueológica”, termos genéricos que demonstram a interdisciplinaridade científica de domínios presentes na linguagem. Na Matemática, “Geometria diferencial”, “Geometria diferencial clássica” e “Geometria não euclidiana” são exemplos de vértices que existem somente no Vocabulário USP.

A proposta para elaboração de estruturas lógico-hierárquicas para expansão da representação dos domínios de Física e Matemática 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 conforme Quadro 8 (Apêndice na versão eletrônica).

5. Considerações finais

Após a análise das estruturas de categorias e das hierarquias lógicas das linguagens de indexação, verificou-se que a escolha das linguagens LCSH, TBN, BIBLIODATA e Vocabulário USP foi adequada para o tipo de análise pretendida, considerando que se trata de linguagens desenvolvidas para contextos semelhantes de aplicação.

A linguagem BIBLIODATA, porém, demonstrou não possuir hierarquização de termos, inviabilizando sua utilização nas fases finais da pesquisa e, portanto, decretando que sua utilização para o propósito final desta análise, a construção de

uma proposta de macroestrutura para a Linguagem UNESP, é inviável. Desta verificação, conclui-se que a escolha das linguagens de indexação destinadas à análise de macroestruturas deve considerar linguagens constantemente atualizadas, como o são a LCSH, a TBN e o Vocabulário USP e que sejam direcionadas a contextos de uso semelhantes.

Quanto à estrutura das linguagens, é possível destacar que a LCSH sobressaiu-se por sua exatidão e diversidade terminológica, oferecendo opções para o trabalho do indexador, de modo que este possa atender às necessidades mais atuais de interdisciplinaridade científica.

A TBN, por ser uma versão da LCSH, possui com essa linguagem mais semelhanças, porém, não apresenta a mesma diversidade que a sua matriz, de modo que esta análise não poderia prescindir do uso de ambas as linguagens para a composição da representação documentária dos domínios.

Quanto ao Vocabulário USP, podemos destacar que sua estrutura diferenciada das demais linguagens indica que sua utilização será importante para a elaboração da macroestrutura da Linguagem UNESP, especialmente por haver sido construído no contexto de um sistema integrado de bibliotecas universitárias prezando, em sua constituição, características das áreas de domínio da universidade.

As três linguagens demonstram ser importantes para a construção de estruturas lógico-hierárquicas com profundidade adequada à busca contínua pela geração de novos conhecimentos e inovação no âmbito das bibliotecas universitárias.

Finalmente, a metodologia utilizada mostrou-se apropriada à proposta, destacando-se o uso da Diretriz ANSI/NISO (2005), que contempla e amplia as orientações propostas pelo IBICT (1984) e UNESCO (1996). O uso do método de comitê ascendente associado ao Modelo Metodológico Integrado para construção de Tesouro de Cervantes (2009) foi bem sucedido e conduziu a elaboração da proposta de estrutura lógico-hierárquica para expansão dos domínios de Física e Matemática. Dessa forma, a utilização de procedimentos terminográficos para a construção de macroestrutura lógica de linguagens de indexação permitiu a caracterização dos domínios e indica-se que seja utilizada esta opção metodológica para a construção da macroestrutura para a Linguagem UNESP.

Notas

- (1) Parte das atividades de pesquisa em desenvolvimento com auxílio regular FAPESP (Processo 2015/13410-8)

- (2) Catálogo *on-line* construído pela rede de bibliotecas universitárias da Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP, disponível em: <http://www.athe na.biblioteca.unesp.br>.

Referências

- American National Standards Institute; National Information Standards Organization (2005). ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, Ma: NISO Press, 2005. 184 p. http://www.niso.org/apps/group_public/project/details.php?project_id=46 (2016-04-31).
- Barbosa, Alice P. (1972). Projeto CALCO: adaptação do MARC II para implantação de uma central de processamento da catalogação cooperativa. 1972. Dissertação (Mestrado em Biblioteconomia e Documentação) – Instituto Brasileiro de Biblioteconomia e Documentação, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Boccatto, V. R. C., Fujita, M. S. L., Gil Leiva, I. (2011). Avaliação comparada do uso de linguagens de indexação em catálogos de bibliotecas universitárias para recuperação por assunto. // *Scire*. 17:1 (2011) 55-64.
- Cervantes, Brígida Maria Nogueira (2009). A construção de tesouros com a integração de procedimentos terminográficos. 2009. 209f. Tese (Doutorado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.
- Cintra, A. M. M.; et al. (1994) Para entender as linguagens documentárias. São Paulo: Polis, 1994. 72 p.
- Fujita, M. S. L. (2011). Modelos de categorização para a construção de tesouros: metodologia de ensino. // Boccatto, V. R. C.; Gracioso, L. de S. (org.) Estudos de linguagem em Ciência da Informação. Campinas, SP: Alínea, 2011. 35-67.
- Fujita, M. S. L.; Boccatto, V. R. C.; Rubi, M. P. (2010). O contexto da indexação para a catalogação de livros em abordagem sociocognitiva. // *Brazilian Journal of Information Science*. 4:2 (jul./dez. 2010) 22-40. <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/485> (2011-07-16)
- Fundação Getúlio Vargas. Rede BIBLIODATA/CALCO (1995) Manual de cabeçalhos de assunto: normas e procedimentos: versão 1.0. Rio de Janeiro.
- Gil Leiva, I. (2008) Manual de indización: teoría y práctica. Gijón: Ediciones Trea.
- Gil Urdiciain, B. (2004). Manual de lenguages documentales. Gijón: Ediciones Trea. 280 p.
- Grings, L. (2016). Controle de autoridades na Biblioteca Nacional do Brasil: breve histórico e práticas atuais. // *Revista Brasileira de Biblioteconomia e Documentação*. 12:1 (No prelo).
- Guimarães, J. A. C. (2008). A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO) // *Revista Ibero-Americana de Ciência da Informação*. 1:1 (jan./jun. 2008) 77-99.
- Histórico. ([2016]). Rede BIBLIODATA. Rio de Janeiro: IBICT. <http://bibliodata.ibict.br/geral/modelos/historicoantes.htm> (2016-03-05)
- Hoshy, L.M.E. (2001). Relationships in Library of Congress Subject Headings. // Bean, C.A. (ed.) Relationships in the organization of knowledge. Dordrecht: Kluwer Academic Publishers, 2001. 135-152.
- Lancaster, F. W. (2004). Indexação e resumos: teoria e prática. 2. ed. Brasília: Briquet de Lemos.

- Torres, L. M. C. (2006). Sistematização da sintaxe de cabeçalho de assunto. // Gomes, H. E. (coord.) BITI - Biblioteconomia, Informação & Tecnologia da Informação. Rio de Janeiro, 2006. <http://www.conexaorio.com/bitil/lecy/lecy.htm> (2013-10-22).
- UNESCO (1993) Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües. [2ª ed. rev. por Derek Austin e Peter Dale]; Trad. de Bianca Amaro de Melo; rev. de Lígia Maria Café de Miranda. Brasília: IBICT; SENAI.
- Universidade de São Paulo (2015) Sistema Integrado de Bibliotecas. Vocabulário Controlado USP. <http://www.sibi.usp.br/produtos/vocabulario-controlado-usp/> (2016-01-16).
- Universidade de São Paulo ([s.d.]). Vocabulário Controlado do SIBi/USP: histórico. <http://143.107.154.62/Vocab/imagens/Historico.htm> (2016-03-27).
- Universidade de São Paulo (2006) Sistema Integrado de Bibliotecas. Manual de indexação de assuntos com uso do Vocabulário Controlado USP: versão preliminar. São Paulo. 72 p. http://areatecnica.sibi.usp.br/wiki/index.php/Vocabul%C3%A1rio_Controlado. (2012-11-27)
- Yi, K.; Chan, L. M. (2010). Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment. // Journal of the american society for information science and technology. 61:4, 677-687. <http://onlinelibrary.wiley.com/doi/10.1002/asi.v61:4/is-suetoc> (2016-03-16).

Enviado: 2016-04-01. Segunda versão: 2016-06-29.
Aceptado: 2016-09-15.
