



TENDÊNCIAS ATUAIS E PERSPETIVAS FUTURAS EM ORGANIZAÇÃO DO CONHECIMENTO

ATAS DO III CONGRESSO ISKO ESPANHA-PORTUGAL
XIII CONGRESSO ISKO ESPANHA

Universidade de Coimbra, 23 e 24 de novembro de 2017

Com a coordenação de

Maria da Graça Simões, Maria Manuel Borges

TÍTULO

Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha

COORDENADORES

Maria da Graça Simões
Maria Manuel Borges

EDIÇÃO

Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20

ISBN

978-972-8627-75-1

ACESSO

<https://purl.org/sci/atas/isko2017>

COPYRIGHT

Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

OBRA PUBLICADA COM O APOIO DE



FLUC FACULDADE DE LETRAS
UNIVERSIDADE DE COIMBRA



CEIS 20
CENTRO DE ESTUDOS
INTERDISCIPLINARES
DO SÉCULO XX
UNIVERSIDADE DE COIMBRA

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

PROJETO UID/HIS/00460/2013



CONSTRUÇÃO DE TESAURO NO DOMÍNIO DOS METADADOS PARA DESCRIÇÃO DE RECURSOS BIBLIOGRÁFICOS

Jefferson Leite Oliveira Ferreira¹, Heliomar Cavati Sobrinho², Virgínia Bentes Pinto³

¹Mestrando Programa de Pós-Graduação em Ciência da Informação/Universidade Federal do Ceará
Jefferson.leite@ufc.br

²Programa de Pós-Graduação em Ciência da Informação/Universidade Federal do Ceará
heliomarcavati@yahoo.com.br

³Programa de Pós-Graduação em Ciência da Informação/Universidade Federal do Ceará
vbentes@ufc.br

RESUMO Descreve as etapas envolvidas na construção de um tesauro no domínio dos metadados aplicados à descrição de recursos bibliográficos. Discute as linguagens documentárias como instrumentos que agregam aportes teóricos-epistemológicos da lógica e da semântica descritiva, a partir do referencial teórico e metodológico de Cavati(2014) e Cervantes(2009). Adota o modelo metodológico integrado como diretriz para a organização sistemática do domínio notional e posterior desenvolvimento do percurso empírico da construção do vocabulário com a sistematização de suas etapas e discussão dos resultados.

PALAVRAS-CHAVE *Metadados descritivos, domínio bibliográfico, representação da informação, tesauro.*

ABSTRACT Describes the steps involved on building of metadata domain thesaurus applied to bibliographic resources description. Discusses documentary languages as tools that aggregates epistemological-theoretic backgrounds from logic and semantic descriptive, starting from Cavati (2014) and Cervantes (2009). Adopts integrated methodological model as guideline to systematic organization of notional domain and later development of empirical roadmap aiming the building of vocabulary with systematization of its steps and results discussions.

KEYWORDS *Descriptive metadata, bibliographic domain, information representation, thesaurus.*

COPYRIGHT Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

INTRODUÇÃO

Desenvolver linguagens documentárias implica empreender esforços que contemplam, além das tarefas de análise, seleção e organização de termos, o adequado tratamento da linguagem representativa do domínio. Não se trata de uma tarefa pragmática ou um exercício técnico linear e assistemático; envolve um arcabouço teórico-metodológico capaz de gerar uma representação enriquecida de significado semântico e contextual. Isto implica que no contexto de um vocabulário controlado, a representação promovida pela linguagem como instancia simbólica do real assume a dimensão das conexões possíveis não-lineares e dinâmicas. Para Caixeta e Souza (2013, p. 39) uma outra dimensão da linguagem era

representada pelos gregos pela palavra *logos*, que significava a palavra racional do conhecimento real (daí a lógica). No âmbito da estrutura lógica das linguagens documentárias enquanto instrumentos de representação do conhecimento, a representação pode ser tratada como “doutrina segundo a qual nosso conhecimento nunca é absoluto, mas sempre flutuante, como se estivesse em um *continuum* de incerteza e indeterminação”(COSTA; MOURA, 2013. p. 13). Isto posto, considera-se necessário reconhecer o caráter multidisciplinar do exercício de construção das LD’s, concordando com Cervantes (2009, p. 27) ao asseverar que o conceito de representação se relaciona com a Psicologia, Linguística, Semiótica, Ciência da Informação, especialmente, Documentação e outras.

Em particular e no que concerne às linguagens documentárias, a tarefa de tradução da linguagem natural para a linguagem do sistema ou controlada deve considerar a mínima perda ou ruído na pós-coordenação. Trata-se de signos que, embora estejam agregados estruturalmente, não devem prescindir de uma semântica capaz de propiciar relações significativas para o alcance da precisão terminológica.

No caso específico da representação da informação é possível estabelecer uma relação semiótica entre o conteúdo temático dos documentos e os termos e expressões adotados para representá-los nos sistemas de recuperação de informação. (COSTA; MOURA, 2013. p. 20)

Para a elaboração de um tesouro, os termos são estruturados, seguindo um triplo sistema de relações que servem de delimitador do conteúdo semântico de cada termo e os vínculos que os relacionam (equivalência, hierarquização e associação semântica) (CAVATI, 2014, p. 45).

Este artigo descreve as etapas da construção de um tesouro no domínio dos metadados para descrição de recursos bibliográficos, que segundo a IFLA, estes recursos, por ocasião dos avanços tecnológicos das últimas décadas, passaram a incorporar novas formas de expressão para além do documento textual, o que implicou no desenvolvimento de novos esquemas de metadados para sua representação.

METODOLOGIA

A partir do modelo integrado proposto por Cervantes (2009) e de sua implementação na pesquisa de Cavati (2014), ambos descritos em pesquisa de doutorado, será apresentado a seguir, a sistematização necessária ao desenvolvimento da linguagem proposta, que contempla os seguintes passos:

- Trabalho preliminar: consiste na escolha da área de especialidade, que envolve, também, a análise do usuário em termos de levantamento de suas demandas.
- Método de compilação: nesta etapa são consideradas as abordagens dedutiva e indutiva com ênfase no alcance de maior flexibilidade para a estruturação do tesouro. Inclui a coleta do *corpus* do trabalho terminológico, estabelecimento da árvore de domínio e a expansão da representação do domínio escolhido.
- Registro de termos: Envolve a coleta e classificação dos termos a partir do *corpus* selecionado na etapa precedente. Implica no julgamento do termo mais adequado para representar determinado conceito, segundo método analítico proposto por Rondeau (1984).
- Verificação de termos: esta etapa consiste na verificação da adequabilidade entre termos e conceitos, o que traz benefícios como aumento da precisão nas relações a serem estabelecidas e na classificação definitiva.

- Forma de apresentação do tesauro: apresenta a estrutura final do vocabulário em sua versão definitiva.

As etapas descritas acima não esgotam as diversas propostas para a construção de linguagens documentárias, mas atendem aos propósitos para sua elaboração na medida em que primam pela objetividade e contemplam os requisitos apoiados na literatura especializada para esta finalidade.

Ao justificar a adoção do modelo metodológico integrado para construção de tesouros, Cavati (2014, p. 97) classifica o modelo de Cervantes como exequível para construção de LD's. Acrescentamos a relevância dos requisitos necessários a esta tarefa, que incluem da caracterização do usuário às formas de apresentação do tesauro.

Das fases preliminares do trabalho, foram adotadas a escolha do domínio e da língua do tesauro bem como a delimitação do subdomínio, delineadas no quadro 1 abaixo. A delimitação do subdomínio corresponde a um recorte necessário para contemplar um universo de termos e conceitos e suas relações que, por si só, já tem o condão de representar com elevado rigor de detalhe o escopo conceitual ou ainda as redes nocionais. Cervantes (2009, p. 147).

Para esta pesquisa foi escolhido o domínio bibliográfico como categoria mais geral para análise terminológica, em seguida, circunscrevemos o subdomínio com o repertório de descritores extraídos dos seguintes vocabulários: Biblioteca Nacional, Rede Pergamum e *Library of Congress Subject Heading*.

Os descritores analisados e extraídos correspondem aos metadados empregados na descrição de recursos bibliográficos, que, por sua vez, integram o domínio ora em estudo.

Quadro 1: Definição das etapas preliminares. Adaptado de Cervantes (2009, p. 163)

ETAPAS PRELIMINARES	
Escolha do domínio e da língua do tesauro	<ul style="list-style-type: none"> • Domínio bibliográfico em tesauro de língua portuguesa e inglesa
Delimitação do subdomínio	<ul style="list-style-type: none"> • Metadados empregados na catalogação de recursos bibliográficos e extraídos dos cabeçalhos da Biblioteca Nacional, Rede Pergamum e <i>Library of Congress</i>.

A compilação inicial dos descritores consta no apêndice A, que inclui a terminologia mais geral do subdomínio consultado bem como as fontes utilizadas para tanto. Atenção especial foi dispensada para eventuais homógrafos e sinônimas com vistas à garantia da consistência na representação conceitual e na posterior construção das hierarquias e associações condizentes ao domínio e usualmente adotadas pela comunidade.

Foram compilados um total de 50 termos representativos do domínio bibliográfico e divididos em categorias para análise e delineamento da árvore de domínio, quadro 2 abaixo, de acordo com a hierarquização nas fontes consultadas. O repertório terminológico compilado foi inicialmente arrolado em planilha Excel e separado por fontes consultadas.

Agência	Qntd. de descritores identificados
Rede Pergamum	17
Biblioteca Nacional	15
Library of Congress	18

Na ocorrência de duplicidades de um mesmo descritor, foi dada preferência àqueles constantes no cabeçalho da *Library of Congress*, pela sua relevância como fonte de informação, aplicando-se o mesmo raciocínio entre a Rede Pergamum e a BN. Neste caso, a escolha recaiu para a segunda opção em virtude da cobertura temática do referido repertório terminológico.

Estas etapas envolvem o julgamento dos termos mais adequados, ou candidatos, a inclusão no tesauro e implica a análise dos conceitos atrelados do geral ao específico (top down) ou, ainda, do específico ao geral (bottom up) segundo modelo proposto por Cervantes (2009, p. 202) .

O levantamento dos conceitos adotou como fonte de informação os próprios cabeçalhos de autoridade, uma vez que tais conceitos remetem a entradas autorizadas empregadas na catalogação o que atesta sua validade e, por outro lado, compila, organiza e sistematiza a relação terminológica no domínio bibliográfico com mais exatidão e razoável exaustividade.

Na etapa denominada “abordagem empírica” Cervantes arrola dois métodos que auxiliam a tarefa do indexador, ou comitês, no julgamento da organização terminológica. Trata-se do método dedutivo que desdobra os termos em relações hierárquicas do mais geral ao mais específico e do indutivo que segue o caminho inverso com ênfase particular na inclusão de novos termos.

Ao desenvolver o tesauro de domínio bibliográfico foram adotados, recursivamente, os dois métodos acima descritos com o objetivo de assegurar a coerência interna do arranjo hierárquico, sem prescindir da garantia literária mediante consultas a artigos de periódicos, anais de eventos, coletâneas, resenhas críticas e bases de citações. A título de exemplo, o descritor “*Information Retrieval*” consta no cabeçalho da *Library of Congress* como um termo genérico a partir do qual se vincula outros como *Cross-language information retrieval* e *Information Services*. Tal arranjo é ratificado na literatura em Biblioteconomia na medida em que as fontes supracitadas validam a hierarquia conceitual e conseqüentemente o desdobramento terminológico.

A forma de apresentação do tesauro adotou o modelo hierárquico convencional com a indicação do termo geral seguido dos termos relacionados constantes no apêndice B.

CONCLUSÃO

Definir estratégias de representação da informação por meio de vocabulários terminológicos constitui prática de forte engajamento inter e multidisciplinar. O modelo metodológico integrado para sua elaboração, na proposta de Cervantes e Cavati, descreve uma seqüência operativa que abstrai detalhes específicos da incorporação do vocabulário controlado a dispositivos tecnológicos, na medida em que se concentra nos aspectos indispensáveis à sua elaboração, tais como a participação de comissões especializadas, delimitação de domínio e subdomínio, registro e verificação de termos dentre outros.

A construção do tesauro no domínio dos metadados para descrição bibliográfica traz diversos e potenciais benefícios ao campo da Biblioteconomia e da Ciência da Informação, dentre os quais o suporte a levantamentos historiográficos do domínio bibliográfico e a inserção, ao longo do tempo, das tecnologias de informação na abordagem dos metadados.

Sabe-se que a finalização de um tesauro não implica a conclusão definitiva de uma tarefa. Conceitos são regularmente revistos, categorias e subcategorias de termos são modificados na medida em que

novas construções teóricas e novos paradigmas forçam o abandono de um termo em detrimento a outro e assim por diante. Na condição de um vocabulário, o tesouro para ser consistente na sua representação e útil como fonte para indexação e recuperação da informação, deve estar inserido em políticas contínuas de revisão e edições que se fazem necessárias

REFERÊNCIAS BIBLIOGRÁFICAS

Caixeta, M., & Souza, R. R. (2013). Representação do conhecimento: história, sentimento e percepção. *Inf. Inf.*, 18(3), 45-67.

Cavati, H. (2009). *A representação documentária do domínio da economia: análise de estruturas de representação em linguagens documentárias e documentos específicos de economia*. São Paulo / SP: UNESP.

Cervantes, Brígida Maria Nogueira (2009). *A construção de tesouros com a integração de procedimentos terminográficos*. Marília / SP , Brazil: UNESP.

Costa, Mateus Uerlei Pereira da, & Moura, M. A. (2013). A representação da informação em contextos de comunicação científica: a elaboração de resumos e palavras-chave pelo pesquisador- autor. *Inf. Inf.*, 18(3), 45-67.

International Federation of Library Associations and Institutions. (2016). *Functional Requirements for Bibliographic Records*. IFLA, 142.

International Organization for Standardization. (1986) ISO 2788-1986: Documentation - Guidelines to establishment and development of monolingual thesauri.

Geneva : ISO.

Rondeau, G. (1984) *Introduction à laterminologie*. (2a ed). Québec, Canadá: Gaëtan Morin