

VOCABULÁRIO CONTROLADO PARA A REPRESENTAÇÃO DOCUMENTÁRIA EM ARQUIVOS CORRENTES DA UNESP

Controlled vocabulary for documentary representation in current files of UNESP

MOREIRA, W. (walter.moreira@marilia.unesp.br),

FUJITA, M.S.L. (fujita@marilia.unesp.br)

DAVANZO, L. (luciana.davanzo@gmail.com)

PIOVEZAN, L.B. (lbpiovezan@gmail.com)

Universidade Estadual Paulista / Faculdade de Filosofia e Ciências /
Programa de Pós-graduação em Ciência da Informação
Av. HyginoMuzzi Filho, 737, Bairro: Mirante
17.525-000 - Marília, SP

Grupo de trabalho: GT3

RESUMO

A área de organização e representação do conhecimento e suas dimensões interdisciplinares dispõem de teorias e de metodologias que se configuram nos mais diversos caminhos para a construção de sistemas de organização do conhecimento, tais como tesouros, taxonomias, sistemas de classificação, ontologias e vocabulários controlados. Dada a necessidade de representação documentária da produção e organização documental em arquivos correntes universitários da UNESP e sua contínua atualização para atender a especificidade e exaustividade de suas atividades meio e fim, foi construído vocabulário controlado com o aproveitamento de listas alfabéticas de palavras existentes nas Seções Técnicas de Comunicação. Após a identificação dos “termos” empregados, realizou-se pesquisa terminológica visando a padronização de termos e termos não preferidos, criando-se rede de remissivas. Verificou-se que será necessário, em continuidade, avaliar o uso do Vocabulário Controlado no atendimento às necessidades da produção e gestão documental dos Arquivos Correntes da UNESP no Sistema SIGAD.

Palavras-chave: Vocabulário controlado. Organização e representação do conhecimento. Seções Técnicas de Comunicação.

ABSTRACT

The organization and knowledge representation area and its interdisciplinary dimensions have theories and methodologies that are configured in various ways to the construction of knowledge organization systems such as thesauri, taxonomies, classification systems, ontologies and controlled vocabularies. Given the need for representation of documentary production and documentary organization in university current archives of UNESP and its continuously need of update to meet the specificity and completeness of their middle and end activities, it was built a controlled vocabulary with the use of alphabetical lists of existing words in the Technical Sections of Communication. After identifying the "terms" employed it was produced a terminology research aiming to standardize terms and non-preferred terms, creating cross references network. It will be necessary, in continuity, to evaluate the use of the Controlled Vocabulary in meeting the needs of production and document management of UNESP current files in SIGAD System.

Keywords: Controlled vocabulary. Organization and representation of knowledge. Communication sections.

1INTRODUÇÃO

Os estudos sobre organização e representação do conhecimento (ORC), tomando-se essa disciplina em sua relação com a ciência da informação, surgem da confluência de algumas outras ciências e disciplinas, tais como as ciências cognitivas, a linguística, a terminologia, a ciência da computação e a lógica.

Trata-se, portanto, de uma disciplina científico-aplicada, “cujo objetivo é melhorar a circulação da informação dentro e por meio de sistemas mediadores – os centros de seleção, armazenamento, recuperação e disseminação de informação – para produzir novo conhecimento e/ou facilitar o acesso ao [conhecimento] existente” (GARCÍA MARCO, 1993, p. 100, tradução livre).

A institucionalização social da ORC é legitimada também por meio dos seus instrumentos de organização, representação e recuperação da informação, tais como os sistemas de classificação, as listas de cabeçalhos de assuntos, os tesouros, as taxonomias, as ontologias. Em comum, tais instrumentos, organizam relações conceituais e primam pelo controle vocabular. Considerando-se, portanto, esse prisma, tais instrumentos podem ser subsumidos à expressão genérica “vocabulários controlados”.

O nível e a eficácia do controle do vocabulário varia conforme a configuração estrutural do instrumento que será utilizado para esse fim, conforme as condições de sua aplicação e, naturalmente, está relacionado aos objetivos de sua aplicação. Há maior precisão terminológica num tesouro, por exemplo, do que numa lista de cabeçalhos de assunto, do mesmo modo que uma ontologia precisa apresentar relações conceituais e terminologia mais acuradas do que uma taxonomia.

Na arquivologia, as discussões sobre aspectos teóricos e metodológicos da construção de vocabulários controlados ainda não são realizadas com a frequência e o aprofundamento com que são realizadas ciência da informação de modo amplo e na biblioteconomia e na documentação, de modo especial. Mesmo a terminologia arquivística referente a esses instrumentos de ORC ainda é incipiente.

Em vista do volume de informações e de documentos em circulação e da impossibilidade de gerenciá-los sem o recurso de instrumentos adequados, a construção e o uso de vocabulários controlados tornou-se necessária em arquivos em outras unidades de informação, de modo geral. No caso da Universidade Estadual Paulista

(UNESP), em particular, considerando-se que essa é uma instituição de ensino superior com algumas características peculiares (uma única instituição de ensino superior com 34 unidades distribuídas em 24 cidades do Estado de São Paulo), o vocabulário controlado é imperativo ao funcionamento eficaz do fluxo de informação arquivística. Sua construção foi iniciada em função da implantação do Sistema de Gestão eletrônica de Documentos - SIGAD que previu o campo de palavras-chaves cujo controle de vocabulário não existia.

Para a construção do Vocabulário de Termos Arquivísticos da UNESP (VTArq-UNESP) foram consultadas as seções técnicas de comunicação (STCom) dos câmpus, visando à identificação e compilação dos “termos” efetivamente usados na representação da documentação que circulava nas STCom. Completada essa fase, realizou-se pesquisa terminológica visando a padronização de termos (n=1056) e termos não preferidos (n=158), criando-se rede de remissivas. A metodologia utilizada na pesquisa terminológica está descrita na seção 3 desse trabalho.

2 VOCABULÁRIO CONTROLADO COMO INSTRUMENTO DE ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

Os vocabulários controlados atuam como instrumentos que minimizam os ruídos na comunicação entre os sistemas de informação e seus usuários, desse modo são importantes instrumentos auxiliares no processo de organização, representação e recuperação da informação. Considerando-se a perspectiva mediadora dos arquivos, acredita-se que, sem o recurso da padronização terminológica na representação documentária, prejudica-se o fluxo de comunicação. Na ausência de controle terminológico por meio de linguagem documentária, isso é, sem a adoção de procedimentos de desambiguação da linguagem natural, realiza-se um uso pobre da linguagem em si, enquanto elemento capaz de sintetizar e sistematizar conceitos.

Dessa forma, os vocabulários controlados atuam como linguagem documentária e são construídos especialmente para “padronizar e facilitar a entrada e a saída de dados em um sistema de informações” (KOBASHI, 2008, p. 1). Essa padronização visa aumentar a precisão do sistema e, por extensão, sua eficácia e sua confiabilidade.

Os conceitos de “vocabulário controlado” e de “linguagem documentária” estão interligados e chegam mesmo a confundir-se, principalmente para o usuário não

especializado. Uma linguagem documentária é, entretanto, por definição, mais ampla. Lara (2004, p. 1) ressalta que as linguagens documentárias, referem-se:

[...] ao conjunto de diferentes tipos de instrumentos especializados no tratamento da informação bibliográfica (sistema de classificação enciclopédica ou facetada e tesouros) [designando] de modo mais amplo e completo, a linguagem especialmente construída para organizar e facilitar o acesso e a transferência da informação.

Portanto, os vocabulários controlados fornecem ferramentas que visam melhorar a eficácia em relação a representação da informação, o que contribui para que o acesso a essas informações aconteçam de forma fluída, sem entraves informacionais.

Através da utilização dos vocabulários controlados torna-se possível diminuir os problemas que podem existir em relação ao acesso, ao uso e a disseminação da informação porque estes permitem que haja uma comunicação mais clara entre o usuário da informação e o respectivo sistema, uma vez que, os vocabulários controlados colaboram com a padronização terminológica.

Aguiar (2008, p. 220) considera que o vocabulário controlado são:

Concebidos como instrumento documentário que visa facilitar a organização, a representação e a recuperação da informação com a finalidade de otimizar a transferência social da informação; ao promoverem ambientes de recuperação da informação de domínios contextuais específicos com consistência, além de minimizarem os ruídos na recuperação da informação.

Dessa forma, faz se necessário o uso de ferramentas que contribuam para a representação dos documentos, de forma que essa representação permita uma recuperação eficiente, independente de quem seja o usuário dessa informação.

Devido à constante necessidade informacional, o profissional da informação, depara-se com uma nova realidade, isto é, garantir não somente a organização da informação, mas proporcionar a recuperação, o acesso e uso da informação, pois, sem o registro da informação não é possível que essa informação seja reutilizada em um momento posterior, uma vez que a partir desse uso, um novo conhecimento pode ser gerado, propiciando uma nova produção de informação Guimarães (2003). Portanto, fazer uso de ferramentas que contribuam para o processo de organização e recuperação da informação, torna-se fundamental.

Acredita-se que uma instituição, que contemple acervos digitais ou analógicos, e que não utilize nenhum tipo de tratamento informacional corrobora para um caos

informacional, neste sentido, a organização da informação criará mecanismos que garantam a organização dessas informações, através de controles, que perpassam todo o ciclo documental, porque para que uma informação possa ser recuperada, disseminada e socializada é necessário que se considere "todas as questões que perpassam os processos de produção, coleta, tratamento ou organização, recuperação, disseminação e uso da informação", conforme aponta Guimarães (2003).

Todavia para que seja possível difundir esse conhecimento, é necessário refletirmos sobre a maneira pelas quais as informações estão sendo produzidas para que dessa forma, seja possível analisarmos seus impactos em relação ao processo de representação e recuperação da informação.

Sobre essa perspectiva, observa-se juntamente com Narakawa (2011, p. 24), que:

Para que esse conhecimento possa ser compartilhado e socializado, é necessário refletir sobre formas possíveis de organização da informação. Garantir sua disponibilização é de certa forma, potencializar a apropriação dessa informação, permitindo que, assim, as pessoas a utilizem em prol das suas necessidades.

Nesse sentido, espera-se que os vocabulários controlados sejam utilizados de forma cada vez mais expressiva pelas unidades de informação, porque embora existam mecanismos automáticos que atuam também como facilitadores na interação entre os usuários e os estoques de informação, a garantia de qualidade dessa interação ainda ocorre, basicamente, por meio de processos humanos de tratamento temático da informação, em tarefas que envolvem análise criteriosa e procedimentos de inferência e de controle de vocabulário.

3 A PROPOSTA DO VTARQ-UNESP

A criação e a composição da sub-comissão de atualização e manutenção do vocabulário de termos arquivísticos da UNESP foi criada pela Portaria Interna CADA/UNESP nº 01, de 10 de março de 2014. Trata-se, portanto, de uma sub-comissão vinculada à Comissão de Avaliação de Documentos e Acesso (CADA), da UNESP.

O VTArq-UNESP foi construído em função de solicitação do Grupo Gestor do Sistema de Gestão Documental Eletrônica (SIGAD) à CADA.

CADA é a atual denominação da comissão responsável pela avaliação dos documentos de arquivos conforme o Decreto nº 58.052, de 16 de maio de 2012, que regulamenta, no Estado de São Paulo, a Lei de Acesso à Informação - Lei Federal nº 12.527 de 18 de novembro de 2011. O Decreto também criou em todos os órgãos e entidades de administração pública estadual os Serviços de Informação ao Cidadão (SIC), os quais são responsáveis pelo atendimento presencial ou eletrônico ao cidadão e pelo encaminhamento das solicitações aos SICs setoriais.

Na UNESP, a instalação do SIC, em 2012, ocorreu de forma centralizada junto ao Gabinete do Reitor, mas organizada mediante a determinação de que as Seções Técnicas de Comunicação de cada unidade universitária ou unidade experimental, onde se localizam os arquivos correntes, se tornassem unidades locais do SIC central, funcionando como um sistema de arquivos correntes para provimento de qualquer solicitação de informação. Dessa forma, a gestão documental é necessária nos arquivos universitários porque são, no entendimento da CADA, o local de destinação e organização da documentação produzida pelas atividades meio e fins da UNESP.

O SIGAD é um sistema que instrumentaliza e dá visibilidade à gestão documental das Seções Técnicas de Comunicação da UNESP. Para seu completo funcionamento necessita dos instrumentos de organização e representação da informação da área arquivística que são o plano de classificação de atividades meio e o plano de classificação de atividades fins, com tabela de temporalidade compatíveis com o Plano de Classificação e Tabela de Temporalidade do Sistema de Arquivos do Estado de São Paulo (SAESP) que coordena o sistema SIC do Estado de São Paulo. Tais instrumentos estão em estágio de finalização após levantamento de todos os tipos espécies documentais produzidos pela UNESP.

O VTArq-UNESP foi elaborado não em função do Decreto Lei Estadual, que não prevê este instrumento de organização e representação da informação, mas por solicitação das Seções Técnicas de Comunicação, que tinham uma prática comum de fazer a determinação de assuntos em todos os denominados processos que organizam os documentos de uma mesma procedência organizacional. Assim, a origem do VTArq-UNESP está na compatibilização de três listas alfabéticas de termos compilados por Seções Técnicas de Comunicação da UNESP. Essa compilação, sem método ou técnica de elaboração de vocabulários, foi realizada ao longo do tempo de prática dessas Seções, o que determinou a autorização desses termos sob garantia de uso.

A elaboração do VTArq-UNESP, portanto, tem origem nas listas alfabéticas como primeira etapa que constituiu-se da normalização e estruturação lógica e semântica dos termos compilados com uso do software TemaTres, versão 1.6.

A partir das listas alfabéticas enviadas pelas STCom, adotaram-se procedimentos de controle vocabular envolvendo controle de sinonímia, de número e gênero e de forma gramatical, além do estabelecimento de relações de equivalência entre os termos, com vistas a diferenciar os termos dos termos não preferidos.

Para esse procedimento, adotou-se um modelo de ficha terminológica simplificado e adaptado às necessidades do projeto. Esse modelo foi organizado a partir de estudo e compilação das propostas de Pavel e Nolet (2002), Barros (2004) e Lara e Tálamo (2007). Durante a fase de pesquisa terminológica e preenchimento das fichas, para dirimir as ambiguidades, recorreu-se à utilização de dicionários especializados, glossários, vocabulários controlados de outras instituições, consulta a especialistas (servidores e docentes da UNESP) e consulta à legislação da UNESP.

A lista final foi elaborada com o recurso do software TemaTres, um software de código aberto utilizado para o gerenciamento de representações formais do conhecimento (SCORNIK; SOSA, 2007).

O VTArq-UNESP, ainda em fase final de construção, possui 1056 termos e 158 termos não preferidos, necessários à política de remissivas do vocabulário controlado. Apresenta dois tipos de relações conceituais: USE (que remete do termo não preferido para o termo) e UP (usado para; que remete do termo para o termo não preferido). Possui, ainda, 85 notas de escopo, utilizadas para esclarecer o significado atribuído a alguns termos e/ou justificar a adoção de alguns critérios de utilização.

4 CONSIDERAÇÕES FINAIS

Verificou-se que será necessário, em continuidade, avaliar o uso do VTArq-UNESP no atendimento às necessidades da produção e gestão documental dos arquivos correntes da UNESP no Sistema SIGAD.

Atualmente, o VTArq-UNESP está em processo de adequação à terminologia dos Planos de Classificação da UNESP para sua disponibilização final dentro do SIGAD. Essa revisão precede, necessariamente, a do uso uma vez que sua terminologia precisa estar adequada à produção documental estadual e universitária determinantes da

procedência arquivística. A liberação para seu uso somente será possível após essa revisão e adequação e precedida por capacitação.

Em sequência, está prevista a capacitação de todos os responsáveis pelas seções técnicas de comunicação na UNESP para a função e uso do VTArq-UNESP. Nessa oportunidade, será estimulada a criação de um grupo de atualização e manutenção do Vocabulário, que deverá ser permanente e acompanhado pelo grupo responsável pela elaboração.

O Grupo de elaboração será responsável por realizar o monitoramento avaliativo de uso do VTArq-UNESP, com metodologia de avaliação extrínseca e intrínseca aplicadas anualmente além da verificação quantitativa e qualitativa dos termos e sua utilização.

REFERÊNCIAS

AGUIAR, F. L. **O controle de vocabulário como dispositivo metodológico para a organização, tratamento e recuperação da informação arquivística**. Dissertação (Mestrado em Ciência da Informação) - Pontifícia Universidade Católica de Campinas, Campinas, 2008.

BARROS, L. A. **Curso básico de Terminologia**. São Paulo: Edusp, 2004.

GARCÍA MARCO, F. J. Paradigmas científicos en representación y recuperación de la información. In: GARCÍA-MARCO, Francisco Javier (Ed.). **Organización del conocimiento en sistemas de información y documentación**: actas del I Encuentro de ISKO-España, Madrid, 4 y 5 nov. 1993. Zaragoza: Universidad de Zaragoza, 1993.

GUIMARÃES, J. A. C. A análise documentária no âmbito do tratamento temático da informação: elementos históricos e conceituais. In: RODRIGUES, G. M.; LOPES, I. L. (Org.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação**. Brasília: Thesaurus, 2003, v. 2, p. 100-117.

KOBASHI, N. Y. **Vocabulário controlado**: estrutura e utilização. Brasília, Escola Nacional de Administração Pública, 2008. Disponível em: <http://www2.ensp.gov.br/rede_escolas/arquivos/vocabulario_controlado.pdf>. Acesso em: 09 set. 2014.

LARA, M. L. G. Linguagem documentária e terminologia. **Transinformação**, v. 16, n. 3, p. 231-240, set./dez. 2004.

LARA, M. L. G.; TÁLAMO, M. F. G. M. Uma experiência na interface linguística documentária e terminologia. **DataGramZero**, v. 8, n. 5, out. 2007.

NARUKAWA, C. M. **Estudo de vocabulário controlado na indexação automática**: aplicação no processo de indexação do sistema de indexação semiautomática (SISA). Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Marília, 2011.

PAVEL, Silvia; NOLET, Diane. **Manual de terminologia**. Canadá: Departamento de Tradução do Governo Canadense, 2002.

SCORNIK, C.; SOSA, O. **Análisis de software libre para la gestión de lenguajes documentales**: caso de estudio: software TemaTres. Corrientes: Universidad nacional del Nordeste, 2007.