

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-2 – Organização e Representação do Conhecimento

DIRETRIZES PARA AVALIAÇÃO DE DOMÍNIOS DE CONHECIMENTO EM TESAuros: UMA ANÁLISE DA ATUALIDADE TEMÁTICA DO MACROTHESAURUS BRASILEIRO DE DIREITO CONSTITUCIONAL

Joyce Angélica Freire Messa (Universidade Federal Fluminense – UFF)

Maria Luiza de Almeida Campos (Universidade Federal Fluminense – UFF)

GUIDELINES FOR EVALUATION OF KNOWLEDGE DOMAINS IN THESAURI: AN ANALYSIS OF THE THEMATIC CURRENT OF THE MACROTHESAURUS BRASILEIRO DE DIREITO CONSTITUCIONAL

Modalidade da Apresentação: Comunicação Oral

Resumo: Propõe diretrizes para a verificação da atualidade temática de tesauros de recuperação, por intermédio da análise de domínios configurados em sua estrutura sistemática. Por meio de uma pesquisa de natureza aplicada, abordagem qualitativa, objetivo exploratório, e utilizando o levantamento bibliográfico como procedimento técnico, foram buscados estudos de avaliação de tesauros com a intenção de identificar princípios para a análise da atualidade temática desses instrumentos, no que concerne ao seu domínio de conhecimento. Como campo de aplicação para a proposição de tais diretrizes, o Macrothesaurus Brasileiro de Direito Constitucional foi analisado, a partir dos critérios reconhecidos na literatura, e sistematizado, numa primeira fase. Em seguida, ementas da disciplina de Direito Constitucional de cursos universitários federais brasileiros foram resgatadas na internet a fim de identificarem-se termos atuais do domínio de conhecimento em questão. Por fim, por meio da comparação entre a amostra de termos atuais do Direito Constitucional e a parte sistemática do Macrothesaurus, foi possível verificar que a ferramenta se encontra parcialmente atualizada e, a partir dos procedimentos realizados, determinar um conjunto de diretrizes que consideramos fundamentais para a avaliação da atualidade temática de tesauros de recuperação e que visam prover os profissionais da informação de procedimentos metodológicos para a seleção de tesauros.

Palavras-Chave: Diretrizes para Avaliação de Tesauros; Análise de Domínio; Macrothesaurus Brasileiro de Direito Constitucional.

Abstract: We propose guidelines for the verification of the thematic current of retrieval thesauri, through the Domain Analysis configured in its systematic structure. Through a research of applied nature, qualitative approach, exploratory objective and using the bibliographical survey as technical procedure, thesauri evaluation studies were searched with the intention of identifying principles for the analysis of the currentness of the thesauri with regard to its domain. As a field of application for the proposal of such guidelines, the Macrothesaurus Brasileiro de Direito Constitucional was analyzed,

based on the criteria recognized in the literature and systematized, in a first phase. Next, the Constitutional Law subjects of Brazilian federal university courses were retrieved on the Internet in order to identify current terms of the domain of knowledge in question. Finally, by comparing the sample of current terms of Constitutional Law with the systematic part of the Macrothesaurus, through which it was possible to observe that it is partially updated in its conceptual aspects, and based on these procedures, we determined a set of guidelines that we consider fundamental for the evaluation of the thematic current of retrieval thesauri and that aim to provide the information professionals with methodological procedures for the selection of thesauri.

Keywords: Thesauri Evaluation Guidelines; Domain Analysis; Macrothesaurus Brasileiro de Direito Constitucional.

1 INTRODUÇÃO

A utilização de um instrumento terminológico é essencial para sanar dúvidas relativas à propriedade do termo utilizado para representar o assunto de um determinado documento. Assim, a contar da leitura da obra e com auxílio da ferramenta utilizada, o indexador consegue distinguir termos, analisando sua definição e suas relações, escolhendo, entre todos os oferecidos, aquele previamente estabelecido no instrumento que pode melhor servir para traduzir a temática determinada (indexação por atribuição).

Os tesouros pretendem oferecer uma maior consistência na indexação¹ dos documentos, a fim de refletir corretamente o contexto do documento que representa. A construção de tesouros apresenta uma relação de custo-benefício muito alta, pois é uma atividade que demanda tempo e requer mão de obra especializada.

Desse modo, o profissional da informação é levado a adotar tesouros existentes no mercado. Entretanto, uma questão de fundamental importância a ser considerada é que pode haver obsolescência dos instrumentos devido à dinâmica do conhecimento, pois toda representação de um dado domínio reflete o estado vigente dele em um tempo e espaço. Este é sempre um problema para o profissional de informação, pois, com o desenvolvimento do conhecimento, novos assuntos surgem e novos documentos, por conseguinte, e assim por vezes o instrumento se torna obsoleto. Nesse sentido, consideramos que estudos que venham a evidenciar critérios para a verificação da atualidade temática de instrumentos de tratamento e recuperação de informação, como tesouros, são de fundamental importância para o profissional da informação.

¹ Entende-se por “consistência de indexação”, neste trabalho, o conceito definido por Lancaster (2004, p. 68), ou seja, “a extensão com que há concordância quanto aos termos a serem usados para indexar um documento”.

Sendo assim, levantamos critérios existentes na literatura e os validamos a partir da avaliação da atualidade temática do Macrothesaurus Brasileiro de Direito Constitucional (MBDC), chegando à enunciação de diretrizes que podem ajudar profissionais da informação a analisar tesouros de qualquer área. Consideramos que tais diretrizes sejam importantes para os profissionais da informação, pois são eles os responsáveis pela triagem, organização e conservação da informação, além de ser de seu domínio a viabilização de seu acesso (CINTRA et al., 2002).

Ressaltamos que a escolha do termo “diretrizes” para este estudo baseou-se no conceito defendido por Camargo (2010, p. 147), de que “Diretrizes são consideradas recomendações ou atividades a serem realizadas ou checadas”, justamente o que buscamos aqui.

Por possuir objetivos com esses direcionamentos, este trabalho se configura como uma pesquisa de natureza **aplicada**, abordagem **qualitativa**, objetivo **exploratório** e que utilizou o **levantamento bibliográfico** como procedimento técnico.

Frisamos que, para a identificação dos conceitos atuais do domínio do Direito Constitucional, foram utilizadas ementas de cursos universitários de Direito, devido ao fato de que elas apresentam uma perspectiva propedêutica². E, ainda, propõem um conjunto de temáticas, visando preparar o profissional para atuação em um mercado de trabalho que é dinâmico e, em virtude disso, elas devem abordar conteúdos que permitam acompanhar o desenvolvimento do conhecimento no domínio em questão.

O presente artigo trata dessas questões a partir das seguintes seções, além desta introdução: **Avaliação de tesouros** – na qual abordamos os principais esforços em avaliação de linguagens documentárias e tesouros, em âmbito nacional e internacional, com a descrição de diversos princípios a serem observados na análise desses instrumentos conceituais; **Diretrizes para avaliação da atualidade temática de tesouros** – nas quais sistematizamos os procedimentos que sugerimos para avaliação da atualidade temática de tesouros; **O Macrothesaurus Brasileiro de Direito Constitucional (MBDC)** – em que apresentamos o instrumento utilizado como estudo de caso neste trabalho, os aspectos relativos à sua construção e suas principais características; **Avaliação da atualidade temática do MBDC:**

² Perspectiva propedêutica é considerada neste trabalho como uma abordagem temática de ensino, que resulta em “desdobramentos curriculares”, como: Preparar o aluno para exames externos, por exemplo, vestibular; preparar futuros cientistas e preparar o “cidadão do futuro” (AULER, 2003, p. 9).

procedimentos e resultados – cujo intuito principal foi elucidar as etapas obedecidas na avaliação deste instrumento; e, por último, apresentamos as **Considerações finais**.

2 AVALIAÇÃO DE TESAUROS

Muitos dos estudos de avaliação de linguagens se restringem à estrutura e ao desempenho de sistemas de recuperação da informação. Sobre esses estudos, é comum que se avaliem a indexação, os algoritmos de recuperação ou classificação, interfaces de usuário ou técnicas interativas (KELLY; SUGIMOTO, 2013).

Para atingir os objetivos delineados para este trabalho, selecionamos autores que avaliaram linguagens documentárias (LD) e cujos critérios nos interessam aqui devido ao fato de que, neste estudo, um dos aspectos abordados será a avaliação de tesouros.

Sobre a função e a importância da avaliação de linguagens documentárias, Boccato e Fujita (2006, p. 268) expressam que “sob o ponto de vista do indexador e do usuário/pesquisador é fundamental para que se verifique até que ponto o desempenho de um sistema de informação fica comprometido ou não com a sua utilização”.

Percebemos, assim, que a avaliação da linguagem documentária utilizada num determinado sistema de informação é fundamental para certificar a boa comunicação entre o acervo e os usuários, de modo que os termos dessa ferramenta representem adequadamente os assuntos desse acervo. A seguir, abordaremos alguns importantes estudos de avaliação de linguagens documentárias nos cenários nacional e internacional.

2.1 Principais Iniciativas de Avaliação de Linguagens Documentárias

Nesta seção, apresentaremos, de forma cronológica, os autores que discutiram aspectos relacionados à avaliação de linguagens documentárias.

O projeto *Armed Services Technical Information Agency* (ASTIA) comparou o desempenho do Unitermo com uma lista de cabeçalho de assuntos desenvolvidos por eles, a fim de descobrir qual era mais útil. Este teste, realizado em 1953, utilizou o critério de relevância, tendo apresentado seus resultados três anos depois, no artigo “*Seven years of work on the organization materials in the special library*” (SOUZA, 2007, p. 34).

Já o *Cranfield I*, desenvolvido por Cyril Cleverdon, a partir de 1957, analisou 18.000 documentos da Aeronáutica, desde a Classificação Decimal Universal (CDU), Unitermo, um catálogo alfabético de assuntos e um esquema de classificação facetada. O objetivo desse

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

estudo foi aferir qual instrumento seria mais eficiente, em se tratando de revocação, precisão, número de termos de indexação por documento, tempo dedicado à indexação, qualidade dos indexadores e tipos de documentos (SOUZA, 2007).

No *Cranfield II*, dez anos após o primeiro, foram indexados 1.400 termos da área da Aerodinâmica para então avaliarem-se os componentes de 33 linguagens de indexação a partir da revocação e precisão (SOUZA, 2007).

O teste *Aberystwyth*, da *College of Librarianship Wales*, analisou a eficiência e a eficácia de 800 documentos, abordando os critérios de especificidade da linguagem e relações entre termos, exaustividade e especificidade da indexação, método de coordenação, precisão dos instrumentos de partição e operadores relacionais, e provisão de contexto nas buscas nos arquivos (SOUZA, 2007).

Muitos daqueles que avaliaram LD não descreveram a metodologia utilizada. Keen (1973 *apud* SOUZA, 2007) o fez, destacando os seguintes critérios utilizados no *Aberystwyth*: origem (adaptações de outras linguagens, domínio contemplado, histórico), método de uso dos termos (pré ou pós-coordenado e com/sem notação), métodos indicativos de associações hierárquicas e características de representação (se é composta de termos/conceitos, número de termos e a adoção de linha metodológica na elaboração).

O projeto *System for the Mechanical Analysis and Retrieval of Text (SMART)*, elaborado por Salton entre 1965-1968 e desenvolvido na Cornell University, foi definido por Robertson (2008) como o mais significativo trabalho de avaliação, no início dos experimentos em recuperação de informações baseados em computador. Esse trabalho foi precursor em muitas das ideias utilizadas nos motores de busca na Web.

Entre 1966 e 1967, Lancaster avaliou o Sistema Medical Literature Analysis and Retrieval System (MEDLARS), sistema de recuperação de informação também baseado em computador e cuja pesquisa era realizada por meio de termos de indexação da LD na área da Saúde – *Medical Subject Heading (MESH)*. Este estudo analisou a revocação e a precisão do sistema, evidenciando falta de especificidade, coordenações falsas e relações incorretas entre os termos (ROBERTSON, 2008).

Em 1976, Carvalho, Botelho e Paranhos estabeleceram 12 etapas para comparar e avaliar o Unitermo, o tesauro EJC, o tesauro ERIC, a *Library Literature Subject Headings List (LLSHL)* e o tesauro de Ciência da Informação, de Schultz, acerca da eficácia das linguagens, por meio dos critérios de sensibilidade (*Se*) e especificidade (*Sp*). As autoras definiram

sensibilidade como “capacidade do sistema em fornecer ao usuário documentos relevantes”, ao passo que a especificidade era “a capacidade do sistema em não oferecer ao usuário documentos irrelevantes” (CARVALHO; BOTELHO; PARANHOS, 1976, p. 51).

Sager, Somers e Mcnaught (1981) também compartilharam os critérios utilizados em seus estudos de avaliação, sendo eles a estrutura do conhecimento (cobertura, descritores, classificação, hierarquia e notação); a estrutura linguística (termos e descritores, tipos de palavras, seleção de descritores e sua forma e as relações entre termos); a apresentação formal (tipografia, códigos, símbolos, ordenação e ortografia); e a representação computacional.

Lancaster (1986) sugeriu que um tesouro deve ser analisado mediante a sua introdução (na qual apresente a área e as características do instrumento, como relações entre os termos, existência de estrutura alfabética e hierárquica, controle de sinonímia e nota de escopo), assim como a consistência e a especificidade do vocabulário, obediência ao padrão internacional, sua tipografia e *layout*. Lancaster ainda indicou que sejam medidas a equivalência entre descritores e não descritores, a reciprocidade da relação entre os termos, a definição dos mesmos, a flexibilidade e a pré-coordenação (média de palavras por descritor).

Bermejo, Rubio e Rojo (1989) investigaram os 47 tesouros existentes em língua espanhola, apresentando cinco parâmetros que medem a funcionalidade dos tesouros: distribuição temática dos instrumentos (domínios); qualidade informativa da sua apresentação (parte sistemática, alfabética, índices complementares e introdução); consistência sintagmática dos descritores (reciprocidade das relações sintagmáticas e paradigmáticas, forma das palavras, plural/singular); estrutura semântica nas relações hierárquicas e associativas (determinação dos campos semânticos, grau de conexão, definições ou notas de aplicação, taxa de equivalência, nível de pré-coordenação e taxa de enriquecimento – proporção entre as relações com o número de termos) e os aspectos estéticos do conjunto e suas partes.

Quatro anos depois, Lara (1993) analisou a Classificação Decimal de Dewey (CDD), CDU, o Thesaurus multilíngue de população (POPIN), e o Macrothesaurus de Informação Socioeconômica para o Planejamento, avaliando primeiramente os aspectos de identificação, objetivos, princípios de construção e forma de avaliação. Posteriormente, avaliou características de divisão das cadeias hierárquicas, flexibilidade dos sistemas, lógica das relações entre os termos e aspectos de eficiência.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Currás (1995) apresentou técnicas de comparação de tesouros, consideradas por nós critérios de avaliação: especificação, exaustividade, estruturação dos termos (unitermos ou termos compostos), uso de termos preferidos, relações de sinonímia, de quase-sinonímia, de hierarquização e de associação, aplicações e definições, origem dos termos (linguística, zona cultural, nível científico, costume de uso) e idioma.

Gil Urdiciain sugeriu aspectos avaliativos, como composição (parte sistemática, e/ou alfabética e gráfica; se possui índices auxiliares e introdução), tamanho (número de descritores e não descritores), relações de equivalência, taxa de enriquecimento, reciprocidade das relações hierárquicas e associativas, número de notas de escopo, morfologia das palavras (forma, singular/plural, entrada), composição e tipografia (tipo de letra, recursos de destaque) e nível de pré-coordenação (URDICIAIN, 1998).

Também em 1998, Letícia Strehl desenvolveu uma metodologia para analisar instrumentos segundo o número de palavras por descritor, uso de singular/plural, sinônimos, descritores compostos, termos homógrafos ou inconsistentes, rotação de descritores, relação entre assuntos redundantes e de assunto/sua subcategoria, descritores que indicam período histórico, identificadores geográficos e cronológicos (STREHL, 1998).

Correa Uribe propôs, também em 1999, um modelo de avaliação de tesouros, dividido na construção de um protótipo e na qualificação do tesouro com base no protótipo. A primeira etapa é a que mais nos interessa aqui, a “definição de atributos globais”, ou seja, o **planejamento** (objetivos, talento humano, metodologia, dados terminológicos, seleção de formato, processamento de dados, fontes bibliográficas, definição de etapas, prova piloto, publicação e atualização), a **estrutura** (descritor, não descritor, relações de equivalência, hierarquia, associação e notas de escopo), a **organização** (introdução, apresentação, instruções de uso, seções alfabética, hierárquica e gráfica, anexos, índice permutado, lista de nomes geográficos e glossário) e a **atualização** (inclusão de termos novos, supressão de termos em desuso, correção ou troca de termos, periodicidade e metodologia de atualização) (CORREA URIBE, 1999).

Para Van der Laan (2002), numa avaliação de tesouros é imprescindível analisar: **informações gerais** – propósito do tesouro; corpo temático; equipe de trabalho; existência de outros tesouros na área; norma de elaboração de tesouro utilizada; temas periféricos; grau de especificidade cujos temas periféricos foram tratados; política de atualização; número total de termos e subtotais de descritores e não descritores; data de inclusão do último descritor;

informações sobre a estrutura – explicação do uso de abreviaturas e convenções; explicação de sinais de pontuação usados; explicação sobre uso de qualificadores e notas de escopo; **informações sobre os descritores** – método de coleta dos candidatos a descritores; fontes de coleta dos candidatos a descritores; critérios de determinação de descritor preferido; forma de validação dos descritores; especialistas, consultores *ad hoc* e outros para avaliação dos descritores; explicitação sobre as relações de equivalência entre os descritores e não descritores; **informações sobre a organização dos conceitos** – explicação sobre as relações hierárquicas; explicação sobre as relações associativas.

Owens e Cochrane (2004) alertaram que muitas das avaliações de tesouro de sua época se aprofundavam nas qualidades “universal” e “ampla” dos tesouros, mas muito pouca atenção era dada aos tesouros de domínios específicos. Sobre o assunto, elas pontuam:

Sugerimos que, se um tesouro de assuntos específicos está em conformidade com a norma ANSI/NISO, o seu cumprimento deve ser esclarecido através de algum tipo de reconhecimento. Tal sistema de classificação ou "selo de aprovação" pode ajudar os bibliotecários na avaliação da utilidade de um tesouro específico do assunto e sua aceitabilidade para esforços de mudança de vocabulário eletrônico. Além de apoiar a construção de tesouros, os padrões podem ser usados para avaliar a estrutura e rigor de um tesouro existente. Idealmente, [a estrutura de padrões citados por Lancaster] combinada com uma avaliação subjetiva por especialistas no assunto pode confirmar a eficácia do vocabulário controlado na satisfação das necessidades de informação do usuário (OWENS; COCHRANE, 2004, p. 92, tradução nossa).

Owens e Cochrane (2004) abordaram o grupo focal como um método de avaliação de tesouros, por meio do qual a utilização ou não de acrônimos e abreviações de termos, a precisão de termos, o contexto do domínio, o controle de sinônimos e termos superiores, além da categorização de termos são algumas das questões feitas aos usuários. As autoras apresentaram, por fim, os principais propósitos a que servem os tesouros, e, portanto, os critérios para avaliá-los: tradução, consistência, indicação de relações e recuperação (OWENS; COCHRANE, 2004). Entretanto, Souza (2007) questiona se aspectos particulares da prática da indexação, como a consistência e a capacidade de tradução, são critérios para avaliar o instrumento ou a própria representação.

No tutorial “Critérios para avaliação de tesouro documentário”, Campos (2006) delineou os seguintes parâmetros para fundamentar decisões de um serviço de informação quanto à adoção ou não de um tesouro existente: Domínio de conhecimento; Apresentação de introdução redigida claramente; Forma de apresentação: alfabética e/ou sistemática;

Idioma: monolíngue ou multilíngue; Unidade linguística: conceito, palavra ou assunto; Tipos de relações; Aspectos ligados à consistência: entre termos, singular/plural e nível de especificidade; Nota de aplicação/escopo (definição do termo e/ou política de indexação).

Em sua dissertação, Souza (2007) fez um levantamento das principais avaliações de linguagens documentárias publicadas até então, tendo elencado os principais critérios para avaliar um instrumento terminológico esportivo, sendo eles: estrutura dos termos (análise do relacionamento entre os termos e estrutura do vocabulário); forma de apresentação (origem da linguagem, se o instrumento apresenta introdução explicativa, domínio, objetivos, condições de uso e atualizações, fundamentos que nortearam sua elaboração, e se possui parte alfabética/sistemática); campo de abrangência (áreas de conhecimento abrangidas, apontando se dados geográficos e cronológicos compõem os assuntos); forma das palavras (unidade linguística, singular ou plural, palavras no feminino ou masculino, na forma simples ou composta, com entrada direta ou inversa, utilizando expressões em mais de uma língua, variações ortográficas e abreviaturas); estrutura semântica (vínculos lógicos entre os termos/subordinação correta, conceitos empregados, articulação de assuntos e exclusão de ambiguidade); relações de equivalência (controle de sinonímia e plano linguístico adotado); *software* adotado pelo SRI (critério para sua escolha).

Ao averiguar se três tesouros biomédicos do Irã obedeciam ao padrão ANSI/NISO Z39.19-2005 – *Guidelines for the construction, format and management of Monolingual controlled vocabularies*, Hariri, Kazerani e Kiya (2012) consideraram o nível de observância aos seguintes padrões: de relações semânticas, de construção (em que grau estão corretos os recursos homográficos, as notas de escopo, o contraste entre as palavras unitárias x sintagmas nominais, adjetivos, artigos iniciais, termos compostos e soletração); de exibição dos vocabulários controlados e de gestão do sistema.

2.2 Classificação dos Critérios de Avaliação de Tesouros

Após a análise dos critérios de avaliação de linguagens levantados na literatura, identificamos apenas os relativos aos tesouros, os quais sintetizamos do seguinte modo:

a) critérios relacionados à política de planejamento do tesouro:

- origem; objetivo; método de uso; equipe; *software*; Gestão de Sistema;
- apresentação formal: número total de descritores; índices complementares; introdução;

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

- política de atualização/inclusão/ supressão e correção de termos; data de inclusão do último descritor;
 - obediência ao padrão internacional;
 - flexibilidade do sistema;
 - idioma (monolíngue ou multilíngue);
 - explicação do uso de abreviaturas.
- b) domínio de conhecimento que aborda:**
- domínio representado; estrutura do conhecimento; temas periféricos;
 - especificidade;
 - categorização correta dos termos.
- c) forma de apresentação:**
- qualidade informativa das partes sistemática e alfabética;
 - aspectos do conjunto e suas partes.
- d) unidade linguística e sua forma de tratamento:**
- unidade linguística abordada;
 - tipografia: uso do singular ou plural, termos compostos, entrada direta ou inversa;
 - critério de determinação de descritor preferido; método de coleta e validação dos descritores; controle de sinônimos e quase-sinônimos; termos homógrafos e inconsistentes; tradução;
 - descritores, assuntos e/ou lista de nomes geográficos e períodos históricos/cronológicos;
 - aspectos ligados à consistência dos termos.
- e) tipos de relação:**
- relações de associação e equivalência entre descritores e não descritores;
 - relações hierárquicas e características de divisão.
- f) notas de aplicação/escopo:**
- explicação do uso de notas de escopo.

A partir dessa análise, foi possível determinar diretrizes que consideramos importantes serem utilizadas para verificação da atualidade temática de tesouros, que apresentamos a seguir. Essas diretrizes foram aplicadas para a verificação do MBDC, como é possível verificar na seção 5.

3 DIRETRIZES PARA AVALIAÇÃO DA ATUALIDADE TEMÁTICA DE TESAuros

A avaliação de tesauros abrange um rol de ações que devem ser desdobradas ao longo de uma pesquisa. Sabemos, todavia, que, na prática, as etapas podem ocorrer simultaneamente, algumas podem não ser necessárias e outras podem ser acrescentadas, consoante às circunstâncias de cada estudo.

Assim, apresentamos nesta seção as diretrizes que sugerimos para avaliação de qualquer tesouro, ilustrando na seção subsequente como cada diretriz foi utilizada na avaliação do MBDC. Essas diretrizes são fruto dos estudos que realizamos e também foram melhoradas à medida que as aplicávamos no campo empírico deste estudo, ou seja, no âmbito do MBDC. Porém, apesar de serem fruto também da análise do tesouro em questão, consideramos mais didático apresentá-las primeiramente.

1ª etapa – Selecionar fonte segura para levantar termos representativos do domínio: o levantamento dos termos que representam um domínio é a primeira etapa para avaliar um instrumento terminológico específico desse domínio de conhecimento. É fundamental que a fonte escolhida possua autoridade notória, de preferência reconhecida nacionalmente.

2ª etapa – Definir a amostra a ser analisada: uma vez levantada a fonte, é necessário estabelecer parâmetros para definir o corpus da pesquisa. O propósito desta ação deve ser o de ajustar os dados, tornando-os passíveis de manipulação, e, principalmente, esquivar-se de tomar decisões infundadas, haja vista que cada método escolhido deve ser elucidado. Exemplos de variáveis: período de cobertura, idioma e distribuição geográfica.

3ª etapa – Estabelecer a forma de identificação dos termos na amostra: após definir a amostra, deve-se extrair os termos que representam o domínio: estes podem ser identificados por intermédio de palavras-chave, por exemplo, quando se selecionam trabalhos acadêmicos como fonte. O objetivo desta etapa é elencar os termos que designam o conteúdo dos documentos.

4ª etapa – Identificar o conteúdo conceitual dos termos: nesta etapa, devem-se identificar termos sinônimos e homônimos a fim de não manter a duplicação dos descritores. Para tanto, faz-se necessário encontrar definições seguras dos conceitos determinados. Deve-se, também, consultar especialistas para validar a escolha das fontes para a determinação das definições dos termos, como também para que eles possam manifestar o entendimento do conteúdo conceitual dos termos, sugerindo definições, quando for o caso.

5ª etapa – Construir quadro de conceitos e definições da amostra levantada: neste momento, sugerimos a criação e o preenchimento de um quadro com cada um dos termos identificados na amostra em número sequencial, sua definição, fonte de onde foi extraída a definição, sinônimos e observações que se mostrem convenientes.

6ª etapa – Examinar a parte sistemática do tesouro: recomenda-se que sejam detalhadas as áreas e as subáreas que formam a parte sistemática do instrumento. Devem-se observar os conceitos superordenados e subordinados para verificar a consistência nas relações hierárquicas, além da grafia dos termos e sua adequação ao domínio em questão.

7ª etapa – Comparar a atualidade conceitual do tesouro com o conjunto de termos levantado: esta etapa trata da análise prática principal do trabalho, na qual se deve buscar no tesouro cada um dos termos levantados na fonte de representação do domínio. Pode-se realizar uma análise quantitativa para verificar o número de termos do *corpus* levantado que existe no tesouro, de forma percentual, evidenciando sua atualidade. Faz-se importante observar, quando existir, as notas de escopo dos termos do tesouro, a fim de sanar dúvidas acerca de sinônimos não reconhecidos nos dicionários, por exemplo.

8ª etapa – Detectar a(s) razão(ões) para não existência dos termos no tesouro: esta é a etapa final e que responde à pergunta: em que grau de modulação os conceitos apontados como representativos de um domínio estão espelhados no tesouro avaliado? Duas respostas recorrentes à esta questão se relacionam a falhas conceituais na elaboração do instrumento: termos que os autores não utilizaram ou desconheciam, e o período de tempo que passou desde que o instrumento foi criado (e o quanto um domínio tem-se desenvolvido e novos termos têm surgido, ou seja, se o instrumento se tornou obsoleto).

Pensamos que, seguindo-se cada diretriz aqui estabelecida, seja possível que profissionais da informação possam avaliar adequadamente a pertinência temática de tesouros de quaisquer domínios do conhecimento. Não obstante, confiamos que essas diretrizes permitem a avaliação de outras linguagens documentárias e que se prestam para avaliarmos a organização de conteúdo de sites ou qualquer estrutura sistemática.

A seguir apresentamos a análise do MBDC à luz das diretrizes acima apontadas.

4 O MACROTHESAURUS BRASILEIRO DE DIREITO CONSTITUCIONAL

Segundo Diniz (2013), a palavra “direito” não é unívoca e nem equívoca, mas análoga, pois designa realidades conexas ou relacionadas entre si, aplicando-se certas vezes à “norma”,

em outras à “autorização” dada pela norma de ter ou de fazer o que ela não proíbe, e também à “qualidade do justo” em incontáveis situações.

Entretanto, devido ao princípio metódico da divisão do trabalho, há necessidade de se decompor analiticamente o direito, como, por exemplo, com a classificação em Direito Público (que regula as atividades da União, dos Estados, do Distrito Federal, dos Territórios, dos Municípios, das autarquias e das demais entidades de caráter público) e Direito Privado (que regula as atividades dos particulares) (DOWER, 2005). Dower (2005) ainda divide o Direito em disciplinas, sendo o Direito Público dividido em Direito Internacional, Constitucional, Tributário, Processual, Administrativo e Penal; e o Direito Privado em Direito Civil, do Trabalho e de Defesa do Consumidor.

Direito Constitucional é, então, um ramo do Direito Público interno, que, segundo Bonavides (2006, p. 43), “fixa as normas fundamentais da organização jurídica e condiciona, debaixo de seus princípios, os demais ramos do Direito Público, com os quais se relaciona”.

O MBDC foi construído em 1987 para dar suporte aos trabalhos da Assembleia Nacional Constituinte (ANC). Ele é produto da base de dados de mesmo nome, criada para auxiliar a indexação de documentos formais e informais, relativos à Constituição, sendo um recurso informacional fundamental para o trabalho de elaboração da Constituição de 1988.

O MBDC não teve nenhuma atualização, considerando-se que sua criação teve como objetivo único auxiliar a ANC. Entretanto, a Biblioteca do Senado Federal mantém, desde o início da década de 1980, o Vocabulário Controlado Básico (VCB³), que se desenvolveu com o suporte do MBDC (LEMOS, 1986).

Acerca da forma de apresentação do tesouro, salientamos que ele é monolíngue (português/Brasil); possui 919 termos autorizados (descritores) e 126 não descritores; seus termos figuram com a primeira letra maiúscula na parte temática ou *corpus* (parte sistemática do tesouro), na qual os descritores estão dispostos em temas/subtemas (ordenados alfabeticamente), seguidos de descritores hierárquicos e associativos, e onde os termos mais gerais, ou as grandes áreas do Direito Constitucional, possuem uma notação única, denominada “código”; há ainda uma listagem de descritores em ordem alfabética, ou seja, ele segue a linha europeia, apresentando parte sistemática, além da alfabética.

A parte sistemática do MBDC é composta de 10 grandes áreas, a saber:

³ O Vocabulário Controlado Básico encontra-se disponível na internet, na página <http://biblioteca2.senado.gov.br:8991/F/-?func=find-b-0&local_base=sen10>.

- a) 10000 FILOSOFIA DO DIREITO / SOCIOLOGIA JURÍDICA
- b) 11000 TEORIA GERAL DO ESTADO
- c) 12000 PODERES / SEPARAÇÃO DE PODERES
- d) 13000 GOVERNO
- e) 14000 PODER CONSTITUINTE
- f) 15000 CONSTITUIÇÃO
- g) 16000 ORDEM INTERNACIONAL
- h) 17000 LEGISLAÇÃO / JURISPRUDÊNCIA
- i) 18000 NOMENCLATURA DE PAÍSES
- j) 19000 NOMENCLATURA DE REGIÕES GEOGRÁFICAS / ESTADOS-MEMBROS / TERRITÓRIOS / DF / REGIÕES METROPOLITANAS / ÓRGÃOS DE DESENVOLVIMENTO REGIONAL

5 AVALIAÇÃO DA ATUALIDADE TEMÁTICA DO MBDC: PROCEDIMENTOS E RESULTADOS

A fim de verificar a atualidade temática do MBDC, foi necessário analisar o domínio do Direito Constitucional apurando seu panorama atual, investigando como este domínio está organizado nos dias atuais e quais são as grandes áreas temáticas desse ramo jurídico, por tratar-se de um domínio amplo e dinâmico.

Nesse sentido, atendemos a primeira diretriz, ou seja, **selecionar fonte segura para levantar termos representativos do domínio**. Para isso, selecionamos a página do Sistema e-MEC, regulado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP)/Regulação e Supervisão da Educação Superior (SERES), instituição máxima na avaliação de cursos de graduação no Brasil, para definir de quais universidades buscaríamos ementas que nos fornecessem os termos representativos do domínio analisado.

Para executar a segunda diretriz – **definir a amostra a ser analisada**, estabelecemos como primeira regra que apenas cursos de graduação em Direito federais, gratuitos e presenciais seriam escolhidos. Não obstante, decidimos que os cursos que iriam compor nossa amostra deveriam possuir nota de conceito de curso (CC) entre 4 e 5⁴, o que totalizou 24 cursos. Então, selecionamos um curso de cada região do país, usando como critério a disponibilidade das ementas na internet.

A terceira diretriz, **estabelecer a forma de identificação dos termos na amostra**, foi obedecida na individualização de cada um dos termos que constituem cada uma das cinco ementas escolhidas. Ou seja, o método que utilizamos para extrair os termos que

⁴ CC: conceito permanente obtido após avaliação in loco pelo MEC (alunos de universidades com Conceito Preliminar de Curso maior que 3 não são obrigados a serem avaliados pessoalmente, transformando o CPC em CC).

representam o domínio do Direito Constitucional foi mediante a identificação e a classificação de cada termo que compunha cada ementa levantada.

Os procedimentos realizados em seguida compõem o delineado na quarta diretriz, **identificar o conteúdo conceitual dos termos**. Nesse momento, foi necessário nos apoiarmos na literatura do domínio (glossários, dicionários, obras básicas), com o propósito de obtermos definições seguras de cada um dos termos da amostra e de modo a não mantê-la na lista de termos duplicados. Quando não encontradas definições nas obras de referência, consultamos especialistas em Direito Constitucional – bacharéis e pós-graduados em Direito, servidores do Ministério Público do Trabalho. Desse modo, foi possível estabelecer definições validadas, por intermédio de entrevista oral e preenchimento de questionário (quadro de conceitos).

De modo a desempenhar a quinta diretriz recomendada: **construir quadro de conceitos e definições da amostra levantada**, elaboramos um quadro no qual listamos cada conceito da amostra, ordenado por um número sequencial, e então detalhamos, um a um, sua associação institucional (da ementa de qual(is) universidade(s) o retiramos), a definição levantada, o dicionário ou obra da qual ela foi retirada, sinônimos assinalados e observações.

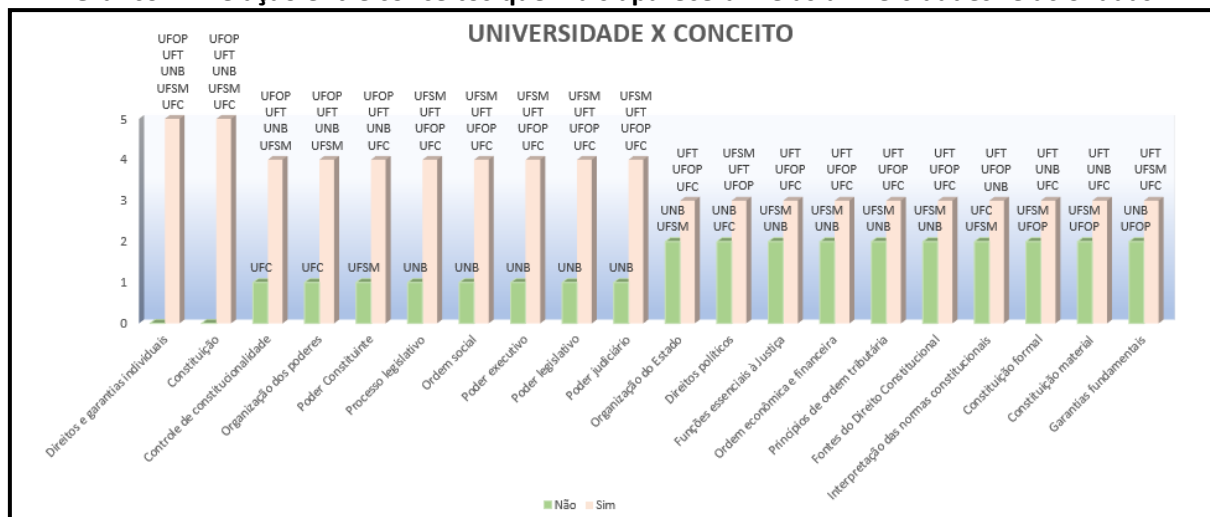
A seguir, descrevemos a parte sistemática do MBDC, expondo suas classes e relações e observando como seus termos estavam grafados, respeitando o indicado pela sexta diretriz – **examinar a parte sistemática do tesouro**.

Isso posto, **comparar a atualidade conceitual do tesouro com o conjunto de termos levantado**, sétima diretriz, tornou-se viável, pois tendo a lista de conceitos que compunham o corpus do MBDC, pudemos verificar nele a existência ou não dos termos levantados nas ementas. Ou melhor, comparamos os conceitos dispostos na parte sistemática do MBDC e as definições dos termos das ementas, considerando os termos com grafia e significado idênticos em ambos, e também termos sinônimos. Acerca da sinonímia existente, foi necessário analisar as notas de escopo dos termos do MBDC, de forma a garantir que eram realmente sinônimos dos termos encontrados nas ementas, como também termos subordinados e relacionados (modulação). Nesta etapa, quando persistiam dúvidas acerca de termos sem nota de escopo, recorreremos aos especialistas que nos auxiliaram.

Esta etapa revelou a necessidade de eles indicarem as correspondências (e as ausências) entre os termos que compõem a parte sistemática do MBDC e as ementas, o que nos possibilitou verificar a pertinência do MBDC como instrumento de recuperação precisa de informações desse campo nos dias de hoje e mapear um conjunto de conceitos que

representam as temáticas tratadas nas ementas das disciplinas do Direito Constitucional: após a verificação dos termos levantados, chegamos ao total de 147 (cento e quarenta e sete) conceitos representativos do domínio (Gráfico 1).

Gráfico 1 : Relação entre conceitos que mais apareceram e as universidades relacionadas



Fonte: Desenvolvido pelas autoras, 2017.

A contribuição e análise dos especialistas considerou diversos aspectos que os fizeram concluir que, por meio de uma comparação quantitativa, o MBDC encontra-se desatualizado, já que possui apenas 92 (noventa e dois) dos 147 (cento e quarenta e sete) termos levantados (38% do total). Entretanto, este fato não inviabiliza o seu uso, posto que ele possui bem-definidas as grandes áreas temáticas, permitindo assim a representação do domínio na atualidade – mesmo que de uma forma mais geral.

Por fim, chegamos à oitava e última diretriz: **detectar a(s) razão(ões) para não existência dos termos no tesouro, e a resposta para isso**. Acreditamos que a idade do instrumento seja uma das explicações da desatualização do MBDC. Outrossim, entendemos que alguns termos que compõem as ementas do MBDC possam não se tratar de conceitos específicos do Direito Constitucional, mas termos jurídicos indispensáveis para que questões relacionadas ao Direito Constitucional possam ser compreendidas.

6 CONSIDERAÇÕES FINAIS

Como já exposto, 38% (trinta e oito por cento) de termos atuais do Direito Constitucional extraídos das ementas não constam no MBDC. Embora não seja um número

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

tão grande, essa constatação fornece subsídios para a continuidade deste estudo, favorecendo a ampliação do tesouro e trazendo benefícios aos usuários deste domínio.

Sabendo-se que a Ciência da Informação é a principal responsável pelo estudo dos processos de análise, coleta, classificação, manipulação, armazenamento, recuperação e disseminação da informação, consideramos que trabalhos como este, que envolvem o mapeamento e a avaliação temática de instrumentos representativos de domínios do conhecimento, sejam de suma importância para o desenvolvimento da área.

Acreditamos que estudos neste segmento possam, também, ser utilizados para a gestão de atividades acadêmicas, possibilitando que pesquisadores conheçam os desdobramentos conceituais dos domínios nos quais pretendem se debruçar, e também na orientação dos investimentos, por parte de agências de fomento à pesquisa.

REFERÊNCIAS

ANSI/NISO. **Z39.19-2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, MD: NISO, 2005. 143 p. Disponível em: <http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf>. Acesso em: 03 mar. 2016.

AULER, D. Alfabetização científico-tecnológica: um novo paradigma? **Revista Ensaio**, v. 5, n. 1, p. 68-83, 2003.

BERMEJO, C. A.; RUBIO, A. V.; ROJO, A. S. Desarrollo de lenguajes documentales formalizados en lengua española: II. Evaluación de los tesauros en lengua española. **Revista Española de Documentación Científica**, Madrid, v. 12, n. 3, p. 283-305, 1989.

BOCCATO, V. R. C.; FUJITA, M. S. L. Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11 n. 2, p. 267-281, maio/ago. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a10.pdf>>. Acesso em: 24 jul. 2017.

BONAVIDES, P. **Curso de direito constitucional**. 19. ed. rev. e atual. São Paulo: Malheiros, 2006. 808 p.

CAMARGO, L. S. A. **Metodologia de desenvolvimento de ambientes informacionais digitais a partir dos princípios da arquitetura da informação**. 2010. 287 f. Tese (Doutorado)– Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

CAMPOS, M. L. A. A indexação e descrição em arquivos: a questão da representação e representação e recuperação de informações. **Arquivo e Administração**, Rio de Janeiro, v. 5, p. 17-31, 2006.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

CARVALHO, M. M.; BOTELHO, T. M.; PARANHOS, W. M. M. R. Linguagens de indexação: uma experiência de análise e avaliação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 51, n. 1, p. 42-59, mar. 1976.

CENDON, B. V. Sistemas e redes de informação. In: OLIVEIRA, M. (Coord.). **Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação**. Belo Horizonte: Ed. UFMG, 2011. cap. 4, p. 59-93.

CINTRA, A. M. et al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002. 92 p. (Palavra-chave, 4).

CORREA URIBE, G. Un modelo para la evaluación integral de tesauros, **Revista Interamericana de Bibliotecología**, Medellín, v. 22, n. 1, p. 139-145, 1999.

CURRÁS, E. **Tesauros, linguagens terminológicas**. Brasília, DF: IBICT, 1995. Disponível em: <<http://livroaberto.ibict.br/handle/1/454>>. Acesso em: 24 jul. 2017.

DINIZ, M. H. **Compêndio de introdução à ciência do direito**. 24. ed. São Paulo: Saraiva, 2013.

DOWER, N. G. B. **Instituições de direito público e privado: de acordo com a nova Lei de falência**. 13. ed. São Paulo: Saraiva, 2005.

HARIRI, N.; KAZERANI, M.; KIYA, R. The evaluation of persian thesauri of health and biomedical sciences based on ANSI/NISO Z39.19 2005 Standart. **Journal of Paramedical Sciences**, Teerã, v. 3, n. 1, p. 34-41, 2012.

KEEN, E. M. The Aberystwyth Index Languages Test. **Journal of Documentation**, v. 29, n. 1, Mar. 1973.

KELLY, D.; SUGIMOTO, C. R. A systematic review of interactive retrieval evaluation studies, 1967-2006. **Journal of the America Society for Information Science and Technology**, New York, NY, v. 64, n. 4, 2006, p. 745-770, 2013.

LANCASTER, F. W. **Indexação e resumos: teorias e práticas**. 2. ed. Brasília: Brique de Lemos, 2004.

LANCASTER, F. W. **Information retrieval systems: characteristics, testing and evaluation**. 2. ed. New York, NY: J. Wiley, 1979.

LANCASTER, F. W. **Vocabulary control for information retrieval**. Virginia, NV: Information Resources Press, 1986.

LARA, M. L. G. Linguagens documentárias, instrumentos de mediação e comunicação. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 26, n. 1/2, p. 72-80, jan./jun. 1993.

LEMOS, M. L. V. Desenvolvimento de um vocabulário controlado na biblioteca do Senado Federal. **Ciência da Informação**, Brasília, DF, v. 15, n. 2, p. 155-58, jul./dez. 1986. Disponível em:

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

<<http://www2.senado.leg.br/bdsf/bitstream/handle/id/201036/436020.pdf?sequence=7>>.
Acesso em: 24 jul. 2017.

MACROTHESAURUS brasileiro de direito constitucional. Brasília, DF: Senado Federal, Secretaria de Documentação e Informação; Rio de Janeiro: Fundação Casa de Rui Barbosa, 1987.

OWENS, L. A.; COCHRANE, P. A. Thesarurus Evaluation. **Cataloguing & Classification Quartely**, Philadelphia, PA, v. 37, n. 3/4, p. 87-102, 2004.

ROBERTSON, S. On the history of evaluation in IR. **Journal of Information Science**, [Thousand Oaks, CA], v. 34, n. 4, p. 439-456, 2008.

SAGER, J. C.; SOMERS, H. L.; MCNAUGHT, J. Thesaurus Integration in the Social Sciences. Part 1. Comparison of Thesauri. **International Classification**, Frankfurt, v. 8, n. 3, p. 133-138, 1981.

SOUZA, J. C. C. E. **Avaliação de linguagem de indexação aplicada à informação jornalística: estudo de caso**. 2007. 156 f. Dissertação (Mestrado em Ciências da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal Fluminense, Rio de Janeiro, 2007.

STREHL, L. Avaliação da consistência da indexação em uma biblioteca universitária de Artes, **Ciência da Informação**, Brasília, DF, v. 27, n. 3, 1998. Disponível em:
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000300011>.
Acesso em: 24 jul. 2017.

URDICIÁIN, B. G. Evaluación semántica y estructural de tesauros. **Revista General de Información y Documentación**, Madrid, v. 8, n. 2, 1998.

VAN DER LAAN, R. H. **Tesouro e terminologia: uma inter-relação lógica**. 2002, 196 f. Tese (Doutorado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. Disponível em:
<<http://www.bibliotecadigital.ufrgs.br/da.php?nrb=000339228&loc=2002&l=27d7844ee48161b8>>. Acesso em: 24 jul. 2017.