

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Departamento de Ciências da Informação

Curso de Biblioteconomia

VINICIUS FERRO

A INDEXAÇÃO E O USUÁRIO: ANÁLISE DE EXPRESSÕES DE BUSCA DO DIREITO
PENAL NO PORTAL LEXML

Porto Alegre

2014

VINICIUS FERRO

**A INDEXAÇÃO E O USUÁRIO: ANÁLISE DE EXPRESSÕES DE BUSCA DO DIREITO
PENAL NO PORTAL LEXML**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do grau de Bacharel em Biblioteconomia, da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

Porto Alegre

2014

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Carlos Alexandre Netto

Vice Reitor: Rui Vicente Oppermann

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Diretor: Ana Maria Mielniczuk de Moura

Vice Diretor: André Iribure Rodrigues

DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO

Chefe: Maria do Rocio Fontoura Teixeira

Vice-substituto: Valdir Jose Morigi

COMISSÃO DE GRADUAÇÃO DA BIBLIOTECONOMIA

Coordenadora: Glória Isabel Sattamini Ferreira

Coordenadora Substituta: Samile Andréa de Souza Vanz

Dados Internacionais de Catalogação

F368iFerro, Vinicius

A indexação e o usuário: análise de expressões de busca do direito penal no Portal LexML. / Vinicius Ferro; professora orientadora: Rita do Carmo Ferreira Laipelt. - Porto Alegre, 2014.
124 f.

Monografia (graduação) – Universidade Federal do Rio Grande do Sul. Faculdade de Biblioteconomia e Comunicação. Curso de Biblioteconomia.

1. Indexação 2. Tratamento da Informação 3. Expressões de busca I. Laipelt, Rita do Carmo Ferreira II. Título

Departamento de Ciências da Informação

Rua Ramiro Barcelos, n. 2705, Bairro Santana

CEP: 90035-007 - Porto Alegre - RS

Fone: (51) 3308 5067

E-mail: dci@ufrgs.br

Vinicius Ferro

**A INDEXAÇÃO E O USUÁRIO: ANÁLISE DE EXPRESSÕES DE BUSCA DO DIREITO
PENAL NO PORTAL LEXML.**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do grau de Bacharel em Biblioteconomia, da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul.

Monografia aprovada em: 02/07/2014.

BANCA EXAMINADORA

Profa. Me. Rita do Carmo Ferreira Laipelt (Orientadora)
Universidade Federal do Rio Grande do Sul

Prof. Dra. Sônia Elisa Caregnato
Universidade Federal do Rio Grande do Sul

Prof. Me. Glória Isabel Sattamini Ferreira
Universidade Federal do Rio Grande do Sul

*Dedico este trabalho
a minha mãe Deonide Chenet Ferro
e à memória de meu pai Narciso Ferro.*

AGRADECIMENTOS

Agradeço a minha mãe por tudo e por ela falar que sou lindo! Mãe, obrigado por sempre estar comigo, auxiliando-me, incentivando-me, apoiando-me e por me dar esse carinho de sempre, amo-te mamãe!

Ao meu irmão, Vilian Ferro, pela grande parceria e amizade.

Ao meu pai, Narciso Ferro, que partira no decorrer desta trajetória, pelo apoio, carinho e exemplo. Pai, hoje sou um pouco de você, teu sangue é meu sangue, amo-te.

A minha Professora orientadora Rita do Carmo Ferreira Laipelt pelos ensinamentos, auxílios e aportes dados, que fundamentalmente contribuíram para a construção deste trabalho.

À Bibliotecária Francine Feldens, pela oportunidade de estágio, paciência e ensinamentos durante os dois anos de estágio num ambiente harmônico, amigável, encorajador e não tradicional, onde pude vivenciar a prática biblioteconômica em todos os seus alcances, desenvolvendo assim meu lado profissional e cidadão. Aprendi muito, obrigado chefe.

Às bibliotecárias da Biblioteca da Escola de Engenharia/UFRGS pelo apoio e instrução no estágio curricular.

Aos colegas da Biblioteconomia Lara Hladovetz, Samara Ayres, Janice Moser, Andressa Balverdu pelo companheirismo e amizade.

Aos professores do Curso de Biblioteconomia que contribuíram de forma substancial na minha formação acadêmica.

À ciência e à tecnologia, sem elas não somos nada.

"No vivimos para cultivar la memoria mirando hacia atrás. Creo que el ser humano tiene que saber cicatrizar sus heridas y caminar en la perspectiva del futuro, pues no podemos vivir esclavizados por las cuentas pendientes de la vida. Es importante no olvidarse de nada, pero pienso que es necesario mirar hacia el mañana. No se vive de recuerdos. Es importante mirar el pasado, pero también es necesario perderle el respeto".

José 'Pepe' Mujica
Presidente do Uruguai (2010-2014)

RESUMO

Demonstra a possibilidade de melhora da consistência da indexação de assunto por meio da análise de expressões de busca dos usuários. Realizou-se um estudo a partir da análise das expressões de busca de usuários do Portal LexML, um sistema com informações jurídicas e legislativas que indexa seus documentos a partir do uso do Vocabulário Controlado Básico do Senado Federal (VCB), uma linguagem de indexação controlada. A pesquisa tem caráter exploratório e uma abordagem qualitativa, iniciada a partir da coleta de *logs* de pesquisa, conjunto de informações que contém as expressões de busca. Seleciona a área do Direito penal como referência e reconhece as expressões de busca que contêm indicativos de falhas no processo de indexação. A partir desta definição e com o amparo de instrumentos de verificação e validação, como o Código Penal e o Código de Processo Penal, a CDDir, o VCB, a pesquisa na jurisprudência e a avaliação de especialista, chega a uma quantidade de trinta e quatro expressões de busca que foram analisadas de forma triangular: pesquisa na jurisprudência (para verificar o uso e a forma verbal usada pelo especialista), pesquisa no VCB (verificar a possível relação da expressão de busca com descritor) e pesquisa no campo assunto/indexação no Portal LexML (verificar o emprego correto do VCB), com a finalidade de reconhecer problemas de inconsistência na indexação que poderiam gerar a não recuperação informacional. Os problemas de consistência da indexação que puderam ser apurados na análise em questão, dizem respeito à etapa de tradução no processo de indexação (LANCASTER, 1993) e principalmente a problemas na representação de assuntos no VCB. As principais inconsistências verificadas, por meio das expressões de busca, estão atreladas à falta de descritores no VCB para representação de assuntos, indicação de uso de descritor mais genérico em detrimento de descritor mais específico, não utilização do VCB como instrumento padrão para a indexação e o não reconhecimento de variantes terminológicas de descritores. Confirma que a análise de expressões de busca pode ser uma ferramenta de auxílio à melhora da consistência da indexação.

Palavras-chave: Indexação. Expressões de busca. Vocabulário controlado. Tratamento da informação.

RESUMEN

Demuestra la posibilidad de mejorar de la consistencia de la indexación de asunto por medio del análisis de expresiones de búsqueda de los usuarios. Se realizó un estudio a partir del análisis de las expresiones de búsqueda de usuarios del Portal LexML, un sistema con informaciones jurídicas y legislativas que indexa sus documentos a partir del uso del *Vocabulário Controlado Básico do Senado Federal* (VCB), un lenguaje de indexación controlada. La investigación tiene carácter exploratorio y un abordaje cualitativo, iniciada a partir de la colecta de *logs* de investigación, conjunto de informaciones que contiene las expresiones de búsqueda. Así, se seleccionó el campo del Derecho penal como referencia y se buscó reconocer las expresiones de búsqueda que pudieran contener indicativos de fallas en el proceso de indexación. A partir de esta definición y con el amparo de instrumentos de verificación y validación, como el Código Penal y el Código de Proceso Penal, a CDDir, o VCB, la investigación en la jurisprudencia y la evaluación de experto, se llegó a una cantidad de treinta y cuatro expresiones de búsqueda que fueron analizadas de manera triangular: investigación en la jurisprudencia (para verificar el uso y la forma verbal usada por el experto), investigación en el VCB (verificar la posible relación de la expresión de búsqueda con descriptor) e investigación en el campo asunto/indexación en el Portal LexML (verificar el empleo correcto del VCB), con la finalidad de reconocer problemas de inconsistencia en la indexación que podrían conducir/generar la no recuperación informacional. Los problemas de consistencia de la indexación que pudieron ser apurados en el análisis en cuestión dicen respecto a la etapa de traducción en el proceso de indexación (LANCASTER, 1993) y, principalmente a problemas en la representación de asuntos en el VCB. Las inconsistencias verificadas, por medio de las expresiones de búsqueda, están ligadas a la falta de descriptores en el VCB para representación de asuntos, indicación de uso de descriptor más genérico en detrimento de descriptor más específico, no utilización del VCB como instrumento patrón para la indexación y el no reconocimiento de variantes terminológicas de descriptores. Se confirmó que el análisis de expresiones de búsqueda puede ser una herramienta de auxilio a la mejora de la consistencia de la indexación.

Palabras-clave: Indexación. Expresiones de búsqueda. Vocabulario controlado. Tratamiento de la información.

LISTA DE FIGURAS

Figura 1 - Forma de apresentação de resultado de pesquisa no Portal LexML...	69
Figura 2 - Forma de apresentação de resultado de pesquisa no Vocabulário Controlado Básico.....	70
Figura 3- Exemplo de parte de um <i>log</i> de pesquisa.....	72
Figura 4-Exemplo de expressão de busca contida em um <i>log</i> de pesquisa.....	73
Figura 5- Pesquisa com a expressão de busca na jurisprudência do Portal LexML.....	76
Figura 6- Pesquisa com a expressão de busca em todo o conjunto de informações (doutrina, legislação, jurisprudência) do Portal LexML.....	77
Figura 7-Presença da expressão de busca no Vocabulário Controlado Básico.	78

LISTA DE QUADROS

Quadro 1 – Expressão de busca conforme jurisprudência.....	80
Quadro2 -Exemplo de expressão de busca da categoria ‘Legislação’	83
Quadro 3 –Exemplo de expressão de busca da categoria ‘Variantes denominativas’	83
Quadro 4 –Exemplo de expressão de busca da categoria ‘Especificidade’	83
Quadro 5 – Expressões de busca da categoria ‘Legislação’	90
Quadro 6 – Expressões de busca da categoria ‘Variantes denominativas’	94
Quadro 7 – Expressões de busca da categoria ‘Especificidade’	103

LISTA DE ABREVIATURAS

ABNT - Associação Brasileira de Normas Técnicas

CDD - Classificação Decimal de *Dewey*

CDDir - Classificação Decimal de Direito

IP - *Internet Protocol*

RVBI - Rede Virtual de Bibliotecas do Congresso Nacional

TCT – Teoria Comunicativa da Terminologia

TGT – Teoria Geral da Terminologia

VCB – Vocabulário Controlado Básico do Senado Federal

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	JUSTIFICATIVA E PROBLEMA DE PESQUISA.....	16
2	OBJETIVOS.....	19
2.1	OBJETIVO GERAL.....	19
2.2	OBJETIVOS ESPECÍFICOS.....	19
3	REFERENCIAL TEÓRICO.....	20
3.1	INDEXAÇÃO.....	20
3.1.1	Análise de assunto e tradução.....	23
3.1.2	Política de indexação.....	26
3.2	O INDEXADOR.....	29
3.3	CONSISTÊNCIA DA INDEXAÇÃO.....	32
3.4	LINGUAGEM DOCUMENTÁRIA.....	36
3.5	TESAURO.....	40
3.6	O USUÁRIO.....	42
4	TERMINOLOGIA.....	48
4.1	TEORIA COMUNICATIVA DA TERMINOLOGIA.....	49
4.1.1	Termo e Conceito.....	53
4.1.2	Variantes terminológicas.....	55
5	METODOLOGIA.....	64
5.1	TIPO DE PESQUISA.....	64
5.2	PORTAL LexML.....	67
5.3	VOCABULÁRIO CONTROLADO BÁSICO.....	69
5.4	COLETA DE DADOS.....	71
5.5	CONSTITUIÇÃO DO CORPUS TEXTUAL.....	73
5.6	VALIDAÇÃO DO CORPUS DE PESQUISA.....	78
5.7	ORGANIZAÇÃO DOS DADOS DO CORPUS DE PESQUISA.....	81
5.8	APRESENTAÇÃO DOS RESULTADOS.....	84
5.9	LIMITAÇÃO DA PESQUISA.....	86
6	ANÁLISE DOS DADOS E RESULTADOS.....	87
6.1	ANÁLISE DOS RESULTADOS DA CATEGORIA LEGISLAÇÃO.....	90
6.2	ANÁLISE DOS RESULTADOS DA CATEGORIA VARIANTES DENOMINATIVAS.....	94

6.3	ANÁLISE DOS RESULTADOS DA CATEGORIA ESPECIFICIDADE.....	102
6.4	IDENTIFICAÇÃO DAS INCONSISTÊNCIAS NO PROCESSO DE INDEXAÇÃO.....	105
7	CONCLUSÃO.....	107
	REFERÊNCIAS BIBLIOGRÁFICAS.....	112
	APÊNDICE A- Seleção inicial de expressões de busca.....	119
	APÊNDICE B - Expressões de busca qualificadas para a análise.....	121
	APÊNDICE C - Comparação entre expressão de busca e descritor representativo do mesmo no VCB.....	123
	APÊNDICE D - Categorias das expressões de busca e descritores respectivos.....	124

1 INTRODUÇÃO

O tratamento e a organização da informação têm como um de seus componentes o processo de indexação, que visa promover uma organização do conhecimento que seja acessível e, ao mesmo tempo, possa representar de forma mais completa possível a informação. Embora, em princípio, a indexação possa ser vista como processo simplificado de representação da informação, ela envolve etapas que demandam certos cuidados para a caracterização mais precisa e fidedigna da informação.

A indexação se desenvolve num caminho triangular produzido pela leitura documental, o indexador e a linguagem de indexação utilizada. Ainda que esta se desenvolva na perspectiva de fornecer acesso à informação, por vezes, não atende às necessidades informacionais dos usuários, visto que pode não considerar fatores que envolvem desde a linguagem de indexação utilizada até a perspectiva do usuário, o que implica na recuperação ou não da informação.

Buscou-se, no levantamento teórico, destacar os aspectos relacionados à organização da informação, por meio de autores como Lancaster, Gil Leiva, Cabré, Fujita, dentre outros que abordam o tratamento da informação e o processo de indexação. Demonstraram-se as etapas e caminhos que conduzem a análise e interpretação de conceitos e suas relações envolvidas na definição de assunto e tradução da informação para uma linguagem documentária.

O processo de indexação pode ser realizado de duas maneiras: a indexação manual, que é realização por atribuição, e a indexação automática, realizada por critérios que envolvem a extração automática de termos presentes no documento por meio de um sistema informatizado. Este trabalho trata apenas do primeiro, visto que do mesmo se podem analisar aspectos intrínsecos que envolvem a subjetividade na identificação da informação e na sua representação por parte do indexador.

A importância do usuário da informação no processo da indexação foi, também, relacionada, visto que, cada vez mais a informação ganha importância na sociedade, nos sistemas de informação e no dia-a-dia, por meio das tecnologias informacionais que desempenham o papel de disseminação da informação. Assim, o usuário a quem esta informação se destina, pode participar do processo de indexação. Conhecendo as formas de pesquisa e as características da busca informacional, tende-se a criar um cenário mais favorável para que a indexação comporte características não apenas oriundas da análise

conceitual do documento, mas também da forma com que o usuário reconhece aquela informação.

1.1 JUSTIFICATIVA E PROBLEMA DE PESQUISA

A crescente produção de informação e armazenamento em sistemas informatizados denota uma necessidade de aperfeiçoamento da indexação, para que a mesma permita uma forma qualificada de recuperação da informação. A indexação caracteriza-se pela descrição de informações de um documento, guiada por instrumento regulador, que visa ser o meio de ligação entre o usuário e a informação. Por vezes, a busca é realizada através de artifícios de pesquisa que visam ampliar o volume de informação recuperada, embora as variantes terminológicas nem sempre estejam compreendidas e indexadas nesse processo.

A indexação, por ser uma das principais formas de representação da informação com vistas à recuperação documental, apresenta diversas características que podem determinar a recuperação ou não do conteúdo informacional de determinado documento. O entendimento de que o processo de indexação consiste no reconhecimento, na compreensão e na ordenação de palavras, com valor ou não de termo, em uma interface de descrição pode denotar uma necessidade de avaliação constante desse processo, a fim de melhorar a indexação realizada, que, por consequência, conduziria a uma melhora da busca informacional.

Para tanto, com a perspectiva de avaliar a consistência da indexação de um sistema que empregasse uma grande quantidade de informação especializada e se utiliza de um vocabulário controlado para a descrição dos assuntos de seus documentos, selecionou-se o Portal LexML. Trata-se de uma base de dados especializada em informação jurídica e legislativa que tem o intuito de reunir leis, decretos, acórdãos, súmulas, projetos de leis, entre outros documentos das esferas federal, estadual, municipal e distrital dos Poderes Executivo, Legislativo e Judiciário de todo o Brasil. Ou seja, quer se consolidar em nível nacional como uma rede de informação legislativa e jurídica que pretende organizar, integrar e dar acesso às informações disponibilizadas nos diversos portais de órgãos ligados ao governo brasileiro na Internet.

Com essa proposta integradora entre as diferentes esferas do governo e da justiça, observa-se no Portal LexML uma necessidade maior ainda de controlar a entrada de dados em seu sistema para possibilitar uma indexação de qualidade. Todo esse controle passa, necessariamente, pela indexação dos documentos, visando evitar problemas de inconsistências na representação temática, que provém da análise documental e da tradução para uma

linguagem documentária controlada. Até mesmo para verificar se a linguagem documentária utilizada atende as necessidades dos usuários, mostra-se necessária uma avaliação/acompanhamento da indexação destes documentos. O Portal LexML emprega como linguagem documentária para a representação dos assuntos dos documentos por ele armazenados, o Vocabulário Controlado Básico (VCB) do Senado Federal. Tal instrumento se configura num sistema de organização do conhecimento que possibilita a representação documental, fornecendo, assim, um controle terminológico dos termos empregados na representação temática dos documentos descritos.

As formas diferenciadas de expressão da informação no processo comunicativo, traduzidas por meio de signos linguísticos, podem vir a afetar a recuperação da informação, ora não conhecidas e representadas pelas linguagens documentárias. Essas diferentes formas de referir-se a um mesmo assunto em um ambiente especializado são conhecidas como variantes terminológicas, que possuem capacidade de representar um mesmo assunto por meio de diferentes signos linguísticos. Deste modo, por vezes, o termo utilizado para representar determinado assunto pode ser insuficiente para atender às necessidades informacionais dos usuários.

Considerando que o Portal LexML possui um vasto conjunto de informações que pertencem às grandes áreas de seu interesse, jurídica e legislativa, e com o intuito de realizar uma análise da indexação realizada nesse Portal, foi delimitada a área do Direito Penal como área de interesse desta pesquisa. Trata-se de reconhecer na linguagem do usuário, representada por meio das expressões de busca, necessidades informacionais que possam não estar sendo atendidas pela indexação realizada.

Em geral, os sistemas de armazenamento e recuperação de informação oferecem parâmetros de pesquisa predefinidos para os diferentes tipos de busca que podem ser utilizados pelos usuários, assim como o Portal LexML, que oferece a possibilidade de busca por assunto e a busca geral. Todo o processo de pesquisa por parte do usuário pode estar carregado de impressões, entendimentos variados e visões sobre determinado tema ou assunto que são manifestados no momento da pesquisa, através das expressões de busca. Além disso, a informação buscada pode estar por vezes associada a determinadas características generalistas ou específicas, em conjunto ou não com outro assunto.

Visto isso, pode-se dizer que as formas com que o usuário realiza suas buscas informacionais podem influenciar diretamente nos resultados das pesquisas. Seja por um motivo relacionado à limitação da compreensão entre o sistema operador e o usuário, seja por um problema relacionado à indexação ou relacionado à terminologia da área expressa na

linguagem documentária, que pode não comportar as diferentes formas de representação de um assunto (variantes terminológicas) ou devido à multidisciplinaridade do assunto.

Uma forma de avaliação/acompanhamento da qualidade da indexação é a análise de *logs* de pesquisa, especificamente a análise das expressões de busca utilizadas pelos usuários. Os *logs* constituem-se em arquivos com dados de acesso, gerados pelo sistema automatizado, comportando as informações das buscas efetuadas pelos usuários no sistema, ou seja, as expressões de busca utilizadas.

Com base no exposto, a questão de pesquisa que se coloca é:

Tendo em vista que a forma com que o usuário realiza suas pesquisas informacionais é uma ação que produz expressões de busca, as quais podem não estar compreendidas pela indexação, perguntamos: a comparação entre os descritores utilizados na indexação de documentos e as expressões de busca dos usuários utilizadas para a recuperação de documentos pode fornecer subsídios para melhora da consistência da indexação?

2 OBJETIVOS

Para a elaboração desta pesquisa, foram definidos o objetivo geral que se pretende alcançar e os objetivos específicos que concorrem para a efetivação do objetivo geral.

2.1 OBJETIVO GERAL

Analisar a consistência da indexação de assunto, a partir das expressões de busca dos usuários, do portal LexML na área do Direito Penal.

2.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos que visam atingir o objetivo geral são:

- a) comparar a linguagem do usuário e a linguagem de indexação (VCB) usada no portal LexML;
- b) verificar em que medida a variação terminológica interfere no processo de recuperação da informação pelos usuários;
- c) discutir a contribuição da possibilidade da análise das expressões de busca para a melhora da indexação.

3 REFERENCIAL TEÓRICO

Abordamos, neste tópico, as referências teóricas que embasam a construção da pesquisa aqui apresentada, através dos seguintes tópicos: a indexação, o sujeito que realiza o processo, ou seja, o indexador; a consistência da indexação e a linguagem documentária.

3.1 INDEXAÇÃO

A organização e o tratamento da informação demandam uma série de ações para que se estabeleça um canal de comunicação eficiente entre a informação e o usuário. A indexação está presente nesse processo de comunicação, servindo como elemento de tradução e representação informacional com vistas à recuperação da informação. Gil Leiva (2008, p.55) diz que o “[...] tratamento da informação tem um objetivo: a difusão da informação”.

O processo de indexação consiste em descrever e caracterizar a informação com a ajuda da representação dos conceitos apresentados num documento, a fim de permitir uma pesquisa eficiente da informação contida num conjunto de documentos (MOREIRO, 2002). Oliveira e Araújo (2012, p. 24) afirmam que a “[...] indexação pode ser ainda o processo de compreensão da essência do conteúdo de um documento e representação deste num modo de expressão em linguagem de indexação”.

A indexação é entendida como um processo que envolve a análise documental, ou seja, uma atividade interativa que depende, também, da familiarização do indexador com o texto, com a área, com o assunto e com a linguagem de indexação utilizada. Segundo Vieira (1988, p. 43), “[...] indexação é uma técnica de análise de conteúdo que condensa a informação significativa de um documento através da atribuição de termos, criando uma linguagem intermediária entre o usuário e o documento”.

A indexação é definida pela Associação Brasileira de Normas Técnicas (ABNT) na norma NBR 12676 (1992, p. 2), que dispõe sobre os métodos para análise de documentos, determinação de seus assuntos e seleção de termos de indexação, como o “[...] ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação”.

Definir indexação é ressaltar a sua função de ser a intermediação entre o usuário e a informação. Ao encontro disso, Bezerra afirma que

[...] a escolha adequada de pontos de acessos que traduzam, de forma clara e precisa, a informação contida nos documentos, identificada no processo de análise documental, fará o diferencial no processo de recuperação da informação (2008, p. 3).

Cintra (1983, p. 23) define a indexação como: “[...] tradução de um documento em termos documentários, isto é, em descritores, cabeçalhos de assunto, termos chave, que têm por função expressar o conteúdo do documento”. A definição dada por Cintra (1983), passa fundamentalmente pela leitura técnica do texto, com o propósito de extrair da linguagem do autor informações que garantam uma tradução da informação para uma linguagem de indexação fidedigna com o que o autor buscou tratar no documento analisado.

Ademais, Lancaster (1993) diz que a indexação se constitui em uma atividade com o propósito de indicar do que trata o documento, ou seja, destacar a temática presente no documento, permitindo a criação de pontos de acesso, mediante os quais um item bibliográfico pode ser acessado. Em todo esse processo de representação da informação, consideram-se:

[...] objetivos da indexação, política de indexação, regras e procedimentos do manual de indexação, a linguagem documentária para representação e mediação da linguagem do usuário e os interesses de busca do usuário, o que significa considerar o sistema de informação e seus serviços de organização e recuperação de informação, bem como o indexador com seu conhecimento prévio profissional e objetivo em situação de interação durante o processamento textual para os objetivos de indexação (FUJITA, RUBI, BOCCATO, 2009, p. 6).

O processo de indexação visa satisfazer a necessidade de representar a informação presente nos materiais informacionais de uma biblioteca, em um sistema de informação, em uma base de dados. Lancaster (1993, p. 8) afirma que “[...] a indexação de assunto possui duas etapas que são a análise conceitual e a tradução”.

A análise conceitual constitui-se na ação de decidir o que é tratado num documento, decidir qual o assunto daquele documento que está sendo representado e os termos a serem utilizados para a indexação. A tradução diz respeito à adoção de uma linguagem controlada para descrição dos assuntos. Segundo lembra o próprio Lancaster (2004), é fundamental que, durante a etapa de análise conceitual, não se tente adequar os assuntos presentes no documento à linguagem de indexação utilizada, pois se criariam, assim, condições para perda da informação ou uma representação equivocada da informação presente no documento.

O conceito de indexação, considerando-a como um processo que envolve critérios técnicos e políticos, abrange muito mais do que o modelo análise-tradução, que pode se limitar tão somente à leitura e representação do documento. Ele se inicia anteriormente à

análise documental e exige uma definição de critérios sobre a utilização de um vocabulário controlado ou de uma linguagem de indexação da informação. O processo de indexação é a representação do conteúdo informacional contido em um documento, tal processo consiste na leitura, compreensão, definição do assunto abordado no documento, para associá-lo a descritores pertinentes autorizados pela linguagem documentária adotada.

A qualidade da recuperação da informação está diretamente ligada à qualidade da indexação; conseqüentemente, uma boa indexação se reflete na precisão e excelente recuperação informacional no sistema no qual ela se realiza.

Segundo Neves, Dias e Pinheiro a indexação envolve ações que vão

[...] além da leitura de documentos, a indexação compreende outras tarefas, dentre as quais se destacam as seguintes: 1) a identificação do conteúdo do documento, ou seja, o assunto ou assuntos de que trata, ou um resumo desse conteúdo; 2) a geração dos respectivos textos desses assuntos/resumos. Na geração desses textos, são muitas vezes utilizados instrumentos como linguagens de indexação (um thesaurus, por exemplo), ou normas técnicas de elaboração de resumos (2006, p.141).

Nessa perspectiva, Dumont (2006, p. 8) diz que “[...] a apropriação do texto pelo leitor implica a produção de sentido, que é onde se imprime a singularidade da leitura”, sendo que esta se caracteriza pela “[...] construção de sentidos e de significados”. Assim, a indexação exige várias competências, que vão desde o domínio de técnicas e métodos relativamente simples até a proficiência na aplicação de complexas linguagens de indexação (DUMONT, 2006).

Sintetiza-se o conceito de indexação como uma técnica de aplicação de análise de conteúdo, por meio da leitura e interpretação/tradução da informação, com a finalidade de extrair do documento informações que levem a representação da informação presente no documento. Necessariamente, a indexação deve ser condizente com a informação relatada no documento e deve estar voltada ao atendimento das necessidades dos usuários, ou seja, descrever a informação que é de interesse da comunidade usuária e fazê-la da forma mais completa, que possibilite sua recuperação.

Na realização da indexação, Lancaster (2004) afirma que todo o processo deve ser conduzido por meio de elementos norteadores que possibilitem ao indexador proceder de forma a garantir qualidade ao processo de indexação e que atenda às necessidades dos usuários. Os elementos norteadores destacados pelo autor referem-se ao indexador, ao vocabulário, ao documento e ao processo.

É recomendável que o indexador tenha algum conhecimento prévio sobre a temática da sua área de atuação, sobre os usuários e a forma como buscam informação. Com relação ao vocabulário, Lancaster (2004) destaca a questão da especificidade e exaustividade da linguagem utilizada, a ambiguidade e a possibilidade de utilização de instrumentos auxiliares na indexação, como dicionários especializados. Quanto ao processo, o elemento destacado é a existência de uma política de indexação, com a definição de padrões de indexação. E, por último, o documento, que atua como elemento introdutório da análise, passando pelo tipo de documento, a sua temática, extensão e características próprias que servirão de elemento para a representação temática.

3.1.1 Análise de assunto e tradução

Como realizar a indexação dentro de um contexto multifacetado, no qual o assunto muitas vezes é multidisciplinar? A indexação deve ser voltada para a recuperação do documento ou para a recuperação da informação? Essas e muitas outras perguntas são bastante pertinentes quando pensamos em indexação. A Associação Brasileira de Normas Técnicas (1992, p. 2) define três estágios para a indexação:

- a) exame do documento e estabelecimento do assunto de seu conteúdo;
- b) identificação dos assuntos presentes no assunto;
- c) tradução desses conceitos nos termos de uma linguagem de indexação.

A análise conceitual busca compreender a faceta do assunto que está sendo tratado no documento. Constitui-se de passos que buscam esclarecer e identificar a abordagem temática do documento. Dentre estes passos, destacam-se a leitura, a interpretação, a associação e observação de fatores que possam contribuir para a indexação com mais clareza e exatidão da informação. Rubi (2009, p. 82) afirma que “[...] a análise de assunto realizada durante a catalogação deverá ser norteada por princípios de política de indexação que devem fazer parte de um manual de indexação da instituição e sobre a qual o bibliotecário deve ter conhecimento e domínio [...]”.

Lancaster (2004) sugere a formulação de perguntas sobre o documento, a fim de extrair informações que identifiquem a temática do mesmo, os aspectos que o tornam de interesse ao usuário e também os motivos pelos quais ele foi incorporado ao acervo. Esse processo extrairá informações das características, ideias e fenômenos associados ao

documento, que podem auxiliar o indexador a destacar aspectos na indexação que tenham mais importância aos usuários do sistema de informação, qualificando assim a recuperação da informação e promovendo a sua disseminação. Na análise conceitual, o indexador busca esclarecer para si que aspectos estão explícitos e implícitos no texto. Esses aspectos dizem respeito à caracterização que deve ser levantada para a indexação que se realizará, visando a obtenção de termos que possam representar de forma específica o documento.

A indexação, por se tratar de um processo que tem o objetivo de difundir a informação, se constrói atendendo as necessidades dos usuários daquela informação e, também, dentro das características da base de dados na qual ela está sendo realizada. Entende-se, assim, que a indexação está diretamente ligada à administração da informação e, mais especificamente, ao contexto da sua organização (CIANCONI, 1990).

Com o aumento considerável da produção informacional em todas as áreas, principalmente no meio técnico e científico, surge a necessidade de melhorar a forma de se indexar a informação dispersa por diferentes bases de dados, repositórios, bibliotecas e pela própria Internet. O condicionamento de recuperação ou não da informação gerado pela indexação, principalmente no meio técnico-científico, evidencia a necessidade de estudo e acompanhamento da indexação, para que a pontualidade, o acesso e a fidedignidade que se espera da informação não sejam perdidos no processo de recuperação da informação.

As diversidades de formas de representação possíveis da informação fazem com que surjam diferentes conjuntos de termos indexadores para representar um mesmo documento. Essa diversidade possível é dependente dos critérios utilizados para se chegar ao conjunto de termos utilizados na indexação e também da destinação deste documento, ou seja, para que público ele se destina (LANCASTER, 1993). Assim, estabelecem-se parâmetros para a realização da indexação, levando em consideração os aspectos que envolvem a base de dados em que ela é realizada, os usuários e o contexto organizacional da instituição. Ao encontro disso, Fujita diz que

[...] a indexação como processo de análise documentária é realizada mais intensamente desde o aumento de publicações periódicas e da literatura técnico científica de modo geral, que impulsionaram a necessidade de criação de mecanismos de controle bibliográfico em centros de documentação especializados (2003, p. 61).

De fato, o processo de indexação levanta aspectos que envolvem a totalidade de etapas do tratamento da informação. Porém, a análise de assunto se destaca como o cerne da indexação, pois dela projetam-se os conceitos que, identificados na leitura e na compreensão,

buscam estabelecer a tematicidade e concepções de termos para representar o documento. Como primeiro passo da análise de assunto para estabelecimento dos conceitos, temos o processo de leitura, que

[...] é subdividido em outros três estágios: compreensão do conteúdo do documento; identificação dos conceitos que representam este conteúdo; e seleção dos conceitos válidos para recuperação (FUJITA, 2003, p. 64).

Para tanto, é necessária a identificação dos conceitos dos assuntos por meio de abordagem sistemática, com finalidade de identificação daqueles conceitos que são os elementos essenciais na descrição do assunto (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2). Através desta síntese inicial, o indexador estabelece a representação por conceitos, que assume função preponderante entre o significado do conteúdo documentário e o termo que o representa. Desse modo, a representação, num sistema de informação, será possibilitada pelo termo, ainda que seja preciso também considerar que a representação durante a análise pode ser realizada por conceitos (FUJITA, 2013).

Fujita afirma que:

[...] na identificação de conceitos, que se caracteriza como o segundo estágio do estabelecimento de conceitos, o indexador, após o exame do texto, passa a abordá-lo de uma forma mais lógica a fim de selecionar os conceitos que melhor representem seu conteúdo (2003, p. 64).

Terminadas essas etapas, ingressa-se na tradução, que é a etapa em que, apropriados os conceitos dos assuntos presentes na linguagem do autor dentro do texto – sendo esta considerada uma linguagem natural –, busca-se compatibilizá-la e representá-la por meio de uma linguagem artificial (linguagem documentária ou de indexação).

Com uso de uma linguagem documentária adotada pelo sistema de recuperação da informação cada termo, então, será procurado e uma vez obtida a confirmação de que é compatível, total ou parcialmente e nesse caso adequado, será considerado um termo autorizado para a representação tanto para a indexação quanto para atendimento às questões de buscas pelo usuário. Portanto, a representação por conceitos na análise de assunto para identificação e seleção de conceitos é diferente da representação realizada na tradução porque, em primeiro lugar, é realizada durante a análise de conteúdo do documento, ou seja, com o documento na íntegra e em segundo lugar, porque utiliza conceitos com os quais são representados os termos e assim identificada a tematicidade intrínseca do documento. (FUJITA, 2013, p.50)

Assim, o processo de indexação, que estimula a interpretação e a determinação dos assuntos de um documento através da análise conceitual, destacando os assuntos presentes nos documentos analisados, completa-se com a tradução. Ou seja, a transcrição dos assuntos

destacados para uma linguagem de indexação que possa representar da forma mais completa a informação presente no documento analisado.

3.1.2 Política de indexação

A constituição de uma política de indexação parte necessariamente de um conhecimento prévio da estrutura de funcionamento de um sistema de informação, levando-se em consideração os tipos de materiais do acervo, a comunidade, o público usuário e o uso documental. A política de indexação deve refletir os objetivos do sistema de informação e deve servir de guia para tomada de decisão no processo de indexação (RUBI, 2012).

Fujita e Redigolo (2006, p. 126) afirmam que a “[...] necessidade de tratar tematicamente a informação de uma forma mais específica é devido às especializações dos acervos [...]”, assim a política de indexação pode contribuir no processo de tratamento temático da informação, desde o estabelecimento de técnicas para tal até os aspectos ligados a especificidade e exaustividade da informação, sempre tendo como norte o usuário. Rubi (2012, p. 91) diz que “[...] a política de indexação deva servir como subsídio para a organização do conhecimento no catálogo, atuando como guia para o bibliotecário no momento da análise de assuntos dos documentos descritos nesses registros”.

A política de indexação pormenoriza o tratamento que será dado à informação presente nos suportes informacionais disponibilizados no acervo. A constituição de uma política de indexação formalizada faz-se necessária para registrar o *modus operandi* de execução da atividade dos indexadores. Atua como uma eficiente ferramenta de indexação, visto que padroniza a operacionalização de tratamento da informação, dando consistência a base de dados utilizada para gerir o acervo. Segundo Carneiro

[...] uma política de indexação, como um guia para tomada de decisões, deve levar em consideração os seguintes fatores: a) características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; b) identificação dos usuários, para atendimento de suas necessidades de informação; c) recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações. São elementos a considerar na política de indexação: cobertura de assuntos, seleção e aquisição de documentos, o processo de indexação (níveis de exaustividade e especificidade, capacidade de revocação e precisão, linguagem), estratégia de busca, forma de saída, tempo de resposta do sistema, avaliação do sistema (1985, p. 221).

Na política de indexação, busca-se a construção de um método de controle de inserção de informações na base de dados, a fim de construir um sistema de recuperação de informação

consistente, que auxilie o usuário na busca informacional no catálogo informatizado. Strehl diz que

[...] os objetivos de uma política de indexação são a definição das variáveis que afetam o desempenho do serviço de indexação, o estabelecimento dos princípios e critérios que servirão de guiana tomada de decisões para otimização do serviço, a racionalização dos processos e a consistência das operações envolvidas (1998, p. 330).

A constituição de critérios para a política de indexação passa necessariamente pela definição de uma linguagem documentária controlada que será utilizada para representação dos assuntos presentes nos recursos informacionais. Conforme apresenta Rubi:

A política de indexação deve ser compreendida como uma decisão administrativa que reflita os objetivos da biblioteca, identificando condutas teóricas e práticas das equipes envolvidas no tratamento da informação da biblioteca para definir um padrão de cultura organizacional coerente com a demanda da comunidade acadêmica interna e externa (2009, p. 83).

Para Carneiro (1985, p. 229) a elaboração de uma política de indexação pressupõe a consideração de aspectos que envolvem decisões relativas a cada fase do processamento temático da informação, são eles:

- a) cobertura dos assuntos: identificação dos assuntos cobertos pelo acervo, com a identificação de áreas que necessitem um tratamento mais específico ou superficial da informação e a identificação dos assuntos centrais e periféricos das áreas de interesse;
- b) seleção e aquisição dos documentos-fonte: indica o estabelecimento de uma política de desenvolvimento da coleção, observando a área de interesse de cobertura do sistema e a qualidade do material bibliográfico no que tange o acervo e, a observância de aspectos voltados ao usuário, como a adequação técnica e intelectual dos documentos ao nível dos usuários, a língua dos documentos e a definição de prioridades de aquisição em relação aos recursos financeiros.
- c) nível de exaustividade: é uma decisão política definida pela administração do sistema de informação e diz respeito a quantidade de termos que serão utilizados para representar um documento em face do seu tipo, cobertura dos assuntos do documento e em relação ao tipo de sistema de informação a qual será aplicada.
- d) nível de especificidade: é a extensão com que se pretende tratar tematicamente um documento, sendo que há uma relação direta entre a especificidade e a capacidade de revocação do sistema, ou seja, quanto mais especificidade aplicada ao

tratamento documental, maior a taxa de precisão e menor a revocação (recuperação).

- e) escolha da linguagem: a linguagem de indexação diz respeito ao desempenho do sistema de recuperação da informação e à qualidade na representação da informação, da linguagem natural para uma linguagem de indexação, que consiga atender as exigências dos usuários, bem como indexar da forma mais precisa a informação.
- f) capacidade de revocação e precisão do sistema: a revocação relaciona-se com a capacidade do sistema de recuperar um bom número de documentos relevantes; a precisão, por sua vez, relaciona-se com a capacidade do sistema em não recuperar documentos que não tenham relevância. Há uma relação inversa entre revocação e precisão, ou seja, quando acontece um aumento na revocação, diminui, assim, a capacidade de precisão do sistema.
- g) estratégia de busca: estabelece a forma de busca que se realizará no sistema.
- h) tempo de resposta do sistema: corresponde ao lapso temporal entre a solicitação da demanda ao sistema e o fornecimento de uma resposta satisfatória. O tempo pode variar em função da precisão e revocação dos sistemas.
- i) forma de saída: corresponde à forma com que serão apresentados os resultados da pesquisa ao usuário. Quanto mais diversificadas forem essas possibilidades, como a separação por autor, ano, tipo de documento, etc., melhor será o desempenho do sistema.
- j) avaliação do sistema: corresponde à satisfação das necessidades dos usuários com relação à aplicação dos demais aspectos enumerados acima. A avaliação do sistema de informação possibilita um acompanhamento da evolução do sistema, do uso, da satisfação do usuário, do atendimento das demandas, enfim, possibilita gerenciar o sistema de informação a partir das avaliações realizadas.

Dentre os aspectos que se mostram consolidados como referência para a constituição de uma política de indexação, encontra-se a escolha de uma linguagem de indexação que tem, por sua vez, a finalidade de ser o instrumento de apoio para a tradução dos assuntos de um texto para uma linguagem padronizada, que visa à consistência da indexação. Por mais completa que a linguagem de indexação seja, cabe sempre ressaltar que a mesma deve comportar o maior número de possibilidades de acesso, incluindo, nessa perspectiva, a questão das remissivas, muito comum nos vocabulários controlados, que, segundo Barros (2004, p. 177), têm por função “[...] indicar ao consulente uma forma mais adequada,

preferível ou usual”. De forma prática, a constituição de uma rede de remissivas tem por função viabilizar as relações semântico-conceituais entre as unidades linguísticas, propiciando que um assunto seja recuperado por todas as suas formas de manifestação, sejam elas oriundas da linguagem especializada ou da linguagem geral.

Assim, os aspectos fundamentais na elaboração de uma política de indexação, conforme Rubi (1985), estão ligados entre si, de modo que a alteração de algum elemento que compõe o sistema afetará os demais como um todo, criando um ciclo de interdependência que deve ser observado por inteiro sempre que se mudar o planejamento de um sistema de recuperação da informação.

De maneira geral, o estabelecimento de uma política de indexação possibilita uma interação entre o sistema e o usuário com o objetivo de assegurar a organização e a transferência da informação armazenada no sistema. Esses critérios prescritivos auxiliam o indexador no processo de tomada de decisão e elevam a padronização na representação da informação.

3.2 O INDEXADOR

A deflagrada explosão informacional, com destaque para os processos de produção e de disseminação da informação e conhecimento, elevou o fluxo informacional. A produção de diferentes tipos de informação projetou uma necessidade de gerenciamento da organização da informação para o atendimento às diversificadas demandas de informações (MOURA, 2006).

No caminho da realização da indexação, o papel dos indexadores transparece fundamental para a finalidade de todo o processo de indexação: prover a recuperação da informação. As etapas que se procedem durante a indexação são constituídas de necessárias habilidades para analisar e decidir qual o melhor conjunto de termos que representam os conceitos expressos nos assuntos dos documentos, tendo o indexador a função de “[...] prever os tipos de pedidos para os quais determinado documento será provavelmente uma resposta útil” (LANCASTER, 2004, p.17).

Visto que o bibliotecário indexador deve ser o mediador da informação, cabe a ele lidar com as três linguagens diferentes que compõe esse processo comunicacional e de representação da informação. A capacidade de lidar com a linguagem do autor, com a linguagem utilizada pelo sistema de informação e com a linguagem do usuário é uma prerrogativa do indexador no processo de representação. Aproximar-se da forma com que o usuário realiza suas estratégias de busca, ter conhecimento prévio sobre a área, vivências e

experiências em indexação e habilidade no uso da linguagem de indexação são ferramentas importantes que facilitam o processo de tratamento e representação da informação.

O indexador, no momento da indexação, busca identificar assuntos familiares ao conhecimento prévio que adquiriu sobre a área temática na qual trabalha, embora deva atentar para as novas temáticas que possam interessar aos usuários do documento. Ao mesmo tempo, essas novas temáticas proporcionam ao indexador ampliar seu vocabulário sobre a terminologia da área (FUJITA, 2003).

Dias e Naves (2007) levantam aspectos sobre a contextualização da indexação como uma etapa subjetiva e afirmam que não se sabe como o indexador decide qual é o assunto de um documento, visto que a percepção e o entendimento dão-se de forma natural. Os autores ressaltam que o processo de indexação realizado pelo indexador deve manter uma coerência e descrever o conteúdo real do documento, a informação que responda as necessidades dos usuários, os conceitos que representam os documentos e os descritores que os representam. Assim, espera-se que a indexação seja um instrumento eficaz de recuperação da informação.

Gil Leiva (2008) trata a indexação como um processo cognitivo de análise documental que visa obter um conjunto de unidades conceituais que representem um documento ou uma necessidade de informação. O processo de indexação envolve ações particulares de leitura, interpretação e compreensão do indexador com a finalidade de fornecer acesso ao usuário da informação presente nos documentos indexados. Assim, o indexador que realiza um processo cognitivo para indexação, percorre desde a variável texto, como objeto de comunicação, até o leitor (usuário), com seu conhecimento prévio necessário para o processo de compreensão (FUJITA, 2008).

A aplicação da leitura técnica é uma das ações necessárias para apropriar-se da informação, reconhecendo nela os assuntos tratados. Dessa forma, para Dias e Naves (2007, p. 44), a “[...] leitura técnica caracteriza-se como uma prática de dar sentido perceptivo e intelectual ao texto”. Nessa perspectiva, a leitura técnica tem a função de fornecer subsídio ao indexador na identificação e na seleção dos conceitos, por meio do reconhecimento dos assuntos. A construção de um significado para o texto é o objeto primordial da leitura técnica. A partir de então, haverá uma integração da informação com o conhecimento próprio do indexador, que se dará desde a construção de um reconhecimento das palavras e seleção de um significado adequado ao contexto, até um reconhecimento global do texto e do discurso (PINTO e GÁLVEZ, 1996).

Segundo Pinto e Gálvez (1996, p. 41, tradução nossa), pode-se distinguir a leitura técnica em três níveis diferentes de processamento da informação:

- a) nível superior: neste nível desenvolve-se o reconhecimento da intenção comunicativa do emissor (autor), por meio de inferência e conhecimento geral do indexador, levando a uma representação global do significado do texto.
- b) nível intermediário: neste nível ocorre o reconhecimento semântico e sintático do léxico, identificação de palavras e a interpretação semântica dos conceitos representados no texto.
- c) nível inferior: neste nível ocorre o reconhecimento de caráter perceptivo, com o reconhecimento, codificação e decodificação de signos.

Assim, da função do indexador exige-se o exercício de uma leitura técnica que o transforma no mediador da informação, exigindo-se do mesmo “[...] respeito aos procedimentos formais da linguagem, para que ele possa construir um objeto coerente [...]” (MOURA, 2006, p.22). Nesse sentido, Neves, Dias e Pinheiro afirmam que:

[...] para compreender um texto, os indivíduos lançam mão de todo o conhecimento prévio armazenado na memória de longo prazo, demandando, inclusive, possíveis esquemas de procedimento existentes na memória semântica. O conhecimento anterior facilita o processamento do texto e a compreensão, por oferecer uma estrutura na qual o conteúdo do material lido possa ser relacionado (2006, p. 142).

No que concerne à observância de fontes próprias do material bibliográfico para a descrição da informação, não existe um padrão próprio para cada tipo de documento. O exame ao documento torna-se fundamental para a identificação de seus assuntos por meio da leitura técnica, caracterizada por Lancaster (2004, p.24) como “[...] um misto de ler e passar os olhos pelo texto”.

Dias e Naves (2007) destacam que a identificação de certas partes do documento e do texto como título, subtítulo, capítulo, bibliografia, resumo, introdução, sumário, prefácio e apresentação podem auxiliar na identificação dos assuntos, corroborando para a determinação do conteúdo do documento. Caracterizar a natureza destas formas básicas de organização da informação, examinar definição e características comuns aos diferentes tipos de texto, bem como relações e funções que compõem essas estruturas pode facilitar a identificação dos assuntos e conceitos integrantes do texto (IZQUIERDO ALONSO, 2004).

O indexador, limitado por seu tempo de realização da indexação e com grande volume de tarefas, fica impossibilitado de aprofundar-se na leitura do documento, devendo sistematizar alguns critérios básicos de identificação dos assuntos, conforme o tipo de documento que possui em mãos.

Embora não se tenha uma regra formalizada de quais partes do texto devem ser analisadas numa estratégia de identificação de assuntos, sabendo-se da impossibilidade de efetuar uma leitura extensa do documento, opta-se por uma leitura técnica das partes que podem conter indicadores do conteúdo do documento. Na medida em que o indexador faz uma análise que lhe dê subsídios suficientes para a identificação dos assuntos, em partes do texto que contenham representatividade da informação, considera-se que o que foi analisado é adequado para a indexação (CINTRA, 1983).

O contexto profissional do indexador, que também influencia no processo de indexação, é composto por uma rotina de contato tanto com a prática da indexação, quanto com a informação, o sistema de recuperação da informação, os usuários e suas demandas. Esses fatores acabam gerando um conhecimento extensivo, uma habilidade própria que lhe permite fazer uma análise mais detalhada, observando e reconhecendo elementos intrínsecos, sutilezas e especificidades não aparentes nos documentos, que podem reforçar e qualificar a indexação. Assim Fujita e Redigolo dizem que:

[...] um fator determinante para a realização do trabalho do indexador é o contexto em que está inserido. O contexto é uma variável importante, porque estabelece e delimita o desempenho do indexador dentro da unidade. Sendo assim, esses catalogadores precisam exercer as suas atividades de acordo com o interesse dos usuários que utilizam o sistema de busca da biblioteca (2009, p. 125).

Com uso de metodologias, ferramentas e diretrizes, é possível qualificar o trabalho e “[...] o conhecimento anterior facilita o processamento do texto e a compreensão, por oferecer uma estrutura na qual o conteúdo do material lido possa ser relacionado” (FUJITA, 2010, p. 91). Assim, a partir da construção da significação da informação por meio da leitura e interpretação, o indexador emprega interpretações que, por possuir um conhecimento prévio, auxiliam no processo comunicacional. A descrição das informações que ele produz possibilita orientar a recuperação da informação através de estratégias de indexação pré ou pós-coordenadas que facilitarão e conduzirão o usuário à informação desejada.

3.3 CONSISTÊNCIA DA INDEXAÇÃO

A consistência é o resultado esperado do bom desenvolvimento da indexação realizada. No entanto, diversos fatores podem influenciar na consistência, os quais se relacionam a aspectos que têm origem nas etapas do processo executadas pelo indexador.

Entende-se, portanto, por consistência da indexação a dimensão de concordância da representação do conteúdo com os termos empregados (DIAS e NAVES, 2007).

Segundo Gil Leiva (2008, p. 72), os elementos que caracterizam o processo, assim como o resultado da indexação, são a exaustividade, a consistência, a especificidade e a correção. Dentro de uma estratégia de busca, Lancaster (1993, p. 75) define a “boa indexação” como aquela “[...] que permite que se recuperem itens de uma base de dados durante buscas para as quais eles sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas úteis”.

Comumente, a indexação é executada como um processo que envolve as etapas de análise de assunto e tradução para uma linguagem de indexação, gerando, assim, uma representação informacional que pode apresentar incoerências, visto que envolve ações interpretativas e entendimentos próprios do indexador. Segundo Strehl (1998, p. 330), a “coerência da indexação consiste na concordância quanto aos termos a serem usados para indexar um assunto tratado em diferentes documentos”.

No contexto da organização e tratamento da informação, é uma exigência a padronização da indexação. Dessa forma, todas as atividades de caráter interpretativo que estejam ligadas ao processo de indexação, como a análise conceitual e a tradução da informação, necessitam que se proceda de forma a objetivar a representação temática fidedigna do seu conteúdo, a fim de facilitar sua recuperação em cumprimento aos requisitos dos usuários (ARAÚJO JÚNIOR, 2007).

No processo de análise conceitual, não se tem critérios objetivos que venham a ditar o como fazer, portanto, busca-se identificar os termos que possam ser mais pertinentes à indexação, mais informativos ou mais relevantes aos usuários. Lancaster destaca as duas falhas envolvidas no processo de análise conceitual:

[...] deixar de reconhecer um tópico que se reveste de interesse potencial para o grupo usuário atendido [...] e [...] interpretar erroneamente de que trata realmente um aspecto do documento, acarretando a atribuição de um termo (ou termos) que seja inadequado (1993, p. 77).

A inconsistência também está ligada ao indexador e aos termos selecionados para representar o documento analisado. O julgamento realizado pelo mesmo, no momento da análise documental, influencia na escolha dos termos, visto que os níveis de especificidade

podem ser diferentes, quando observadas as indexações realizadas de um documento por diferentes indexadores (DIAS e NAVES, 2007).

Na etapa de verbalização, onde a informação condensada e extraída do texto é traduzida para uma linguagem de indexação, Lancaster destaca as duas possíveis falhas de tradução:

[...] deixar de utilizar o termo mais específico disponível para representar um assunto [...] e [...] empregar um termo que seja inadequado para o conteúdo temático devido à falta de conhecimento especializado ou por causa de desatenção (1993, p.77).

Ressalta-se que na etapa de tradução da informação em termos adequados podem ocorrer omissões que não deveriam ser imputadas exclusivamente ao indexador, mas, sim, aos vocabulários controlados utilizados. Esses, na sua construção ou, até mesmo, por não manterem uma atualização constante, nem sempre comportam as variantes terminológicas existentes na área e acabam, dessa forma, dificultando o acesso à informação na medida em que não respondem adequadamente às demandas na recuperação da informação. Incluem-se também como problemas a não representação de assuntos e a falta de descritores mais específicos para representar os assuntos.

Dodebei (2002, p.72) afirma que, na construção das linguagens documentárias, “[...] é necessário certificar-se de que a terminologia levantada possui duas garantias: literária e do usuário”. Assim, espera-se que na linguagem documentária seja considerado um maior número de termos possíveis que representem uma mesma informação, a partir de uma estruturação em que a informação possa ser recuperada tanto através do termo autorizado, quanto pelas suas formas variantes.

No que tange a parte mais objetiva da indexação, o seguimento da política de indexação e o uso de uma linguagem de indexação são premissas da atuação do indexador frente a suas rotinas de indexação. Strehl afirma que a política de indexação envolve objetivamente

[...] a definição das variáveis que afetam o desempenho do serviço de indexação, o estabelecimento dos princípios e critérios que servirão de guia na tomada de decisões para otimização do serviço, a racionalização dos processos e a consistência das operações envolvidas (1998, p. 330).

Como visto até então, a indexação é o processo no qual o indexador estabelece os conceitos, a tradução e os termos que serão utilizados na indexação. Contudo, existem

critérios que norteiam a profundidade de descrição dos assuntos tratados e, também, estabelecem quantidades de termos que devem ser utilizados para que a representação da informação presente no documento possa ser recuperada. São eles: a especificidade e a exaustividade.

Rubi (2009, p.85) diz que “[...] a especificidade está relacionada ao nível de abrangência que a biblioteca e a linguagem documentária permitem especificar os conceitos identificados no documento”. Já a exaustividade diz respeito ao número de termos atribuídos como descritores do assunto do documento, ou seja, em que medida todos os assuntos discutidos no documento são reconhecidos durante a indexação e traduzidos para a linguagem documentária da biblioteca.

Portanto, a ideia central da especificidade e da exaustividade é que com o emprego dessas premissas, o processo de recuperação da informação poderá ser aprimorado, trazendo mais qualidade para o sistema de informação. Rubi (2009, p. 90) afirma que essa situação “[...] apresentará impacto direto na recuperação da informação: quanto mais ou menos exaustivamente a indexação é feita, maior ou menor será a revocação no momento da recuperação”.

Quanto à especificidade, Lancaster (1993) recomenda que se utilizem termos específicos ao invés de termos genéricos, incluindo todos os tópicos que sejam de interesse dos usuários que estejam sendo tratados de forma substancial no documento e que se indexem os tópicos tão especificamente quanto à linguagem de indexação permitir. Assim, quanto maior a especificidade, menor a revocação e maior a precisão; ou seja, um número menor de itens será recuperado, porém, maior será a precisão. A recuperação com níveis de revocação menor produz um índice maior de precisão (RUBI, 2009).

No que se refere à exaustividade, que é a quantidade de termos que serão utilizados na indexação, Lancaster (1993, p. 23) afirma que “[...] a indexação exaustiva implica no emprego de termos em número suficiente para abranger o conteúdo temático do documento de modo bastante completo”. Quanto maior a exaustividade, maior a revocação, criando uma diminuição da precisão, ou seja, mais itens serão recuperados e menor será a precisão do sistema. A recuperação de um grande volume de documentos está ligada à baixa precisão dos termos designados para representar seus assuntos (RUBI, 2009).

A atribuição de termos evidencia a importância que está sendo dada a informação após a análise do indexador. A exaustividade, como uma necessidade de representação da informação por um maior número de termos atribuídos pelo indexador, resulta numa estratégia evidente de dar mais opções de acesso à informação ao usuário, embora quando se

tem no horizonte as premissas da especificidade na indexação, a informação tem um caráter mais qualitativo, onde se opta pelo direcionamento da indexação a um nível mais específico. Para Martínez Tamaio e Valdez (2008, p. 43, tradução nossa) “[...] entre exaustividade e especificidade há uma relação diretamente proporcional, isto é, quanto maior a exaustividade maior é a especificidade e vice-versa”. Tanto a especificidade quanto a exaustividade são elementos de avaliação da indexação e devem fazer parte da política de indexação do sistema de informação, para que se consiga padronizar as formas de representação da informação e também para que se registre o processo, fazendo com que o mesmo se torne rotina dos indexadores, o que traz qualidade e consistência à indexação realizada; pois a capacidade de relevância de um sistema de informação está relacionada ao número de documentos recuperados para atendimento das solicitações encaminhadas pelo usuário.

3.4 LINGUAGEM DOCUMENTÁRIA

Em geral, os sistemas de recuperação da informação possibilitam acessar a informação por meio de diferentes formas de busca, fazendo pós-coordenações com os termos utilizados na busca pelo usuário; sendo, tais termos, oriundos da linguagem natural ou de uma linguagem controlada. Para que se obtenha uma qualidade na interação com o usuário e na recuperação da informação, é necessário o controle do vocabulário utilizado na indexação dos documentos através de uma linguagem documentária.

Bocatto diz que

[...] as linguagens documentárias são linguagens estruturadas e controladas, construídas a partir de princípios e de significados advindos de termos constituintes da linguagem de especialidade e da linguagem natural (linguagem do discurso comum), com a proposta de representar para recuperar a informação documentária (2009, p. 119).

A linguagem documentária pode ser definida como um sistema de esquemas conceituais, a partir de relações lógico-semânticas, onde os elementos do sistema possuem relações hierárquicas entre si e/ou entre categorias, além de representarem uma determinada área do conhecimento (CAMPOS, 2001). A produção de uma linguagem documentária de um domínio de uma especialidade executa uma função de controle terminológico. Cintra et al. (2002, p. 40) afirmam que “[...] por meio das terminologias de especialidade, as palavras passam a ser termos, assumindo significados vinculados a sistemas de conceitos determinados”.

A linguagem documentária apresenta-se como uma forma de controle terminológico tanto para o processo de representação quanto para o de recuperação da informação, assim

[...] visa à organização e à disseminação de conteúdos informacionais de sistemas de informação, tais como as bibliotecas universitárias, que exigem melhor controle da terminologia para um desempenho adequado da recuperação e filtragem de informações (BOCCATO e FUJITA 2001, p. 26).

A linguagem documentária se constitui de um vocabulário controlado, que se firma como um instrumento formal de controle de representação temática informacional por meio de controle terminológico de uma determinada área. Cintra et al. (2002) afirmam que a elaboração dessas linguagens documentárias atende às atividades de indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos destinados a traduzir os conteúdos dos documentos. A linguagem documentária pode ser vista como um instrumento de mediação entre sistemas, conjuntos informacionais e os usuários, exercendo a função de elo entre, ao menos, duas linguagens: a linguagem do sistema e a do usuário (FUJITA e REDIGOLO, 2009).

O vocabulário controlado é a linguagem documentária de controle da terminologia empregada em determinada área. Strehl (1998, p. 330) afirma que “[...] o vocabulário controlado é um instrumento utilizado na segunda etapa do processo de indexação, na qual o indexador traduz os conceitos extraídos do documento para a linguagem utilizada no sistema”.

Algumas das características de um vocabulário controlado são a rigidez, a inflexibilidade e a precisão, necessárias para desempenhar a ação comunicativa, devendo apresentar uma série de qualidades, em que se destacam o funcionamento como código inteligível e fonte para interpretação do sentido, bem como sua caracterização como metalinguagem, além da incorporação do usuário como integrante do processo (FUJITA e REDIGOLO, 2009).

Lancaster (1993, p.14) diz que um vocabulário controlado é “[...] essencialmente uma lista de termos autorizados”, [...] formatado por meio de uma “estrutura semântica”, sendo muito “mais do que uma mera lista”. Ou seja, há, então, três tipos de vocabulários controlados: esquemas de classificação bibliográfica, cabeçalhos de assunto e tesouros. A sua estrutura tem por objetivo:

- a) controlar sinônimos, optando por uma única forma padronizada, com remissivas de todas as outras;
- b) diferenciar homógrafos;

- c) reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si, como as relações hierárquicas e as não hierárquicas (ou associativas).

Rowley (2002) destaca as vantagens e desvantagens da utilização de um vocabulário controlado para a organização do tratamento da informação. A autora estabelece como vantagens da utilização de um vocabulário controlado o fato de ele solucionar problemas semânticos, proporcionar a identificação de relações de gênero-espécie, além de mapear áreas do conhecimento. Entre as desvantagens identificadas pela autora e são o alto custo, a desatualização do vocabulário, a imprecisão de cobertura, os erros de construção e a dificuldade de representar no sistema todas as relações relevantes entre os termos.

O vocabulário controlado tem papel importante também na recuperação da informação, visto que os termos utilizados para representação do assunto de um documento, oriundos de um vocabulário controlado, podem orientar os usuários na “[...] elaboração das estratégias de busca de informações no sistema” (FUJITA; REDIGOLO, p. 128, 2009).

Há diversidade nas formas de busca e expressões que o usuário pode empregar no processo de busca informacional, visto que a linguagem natural é permissiva quanto às construções que comportam todas as intenções do usuário no processo comunicativo de busca de informação. Assim, “[...] neste contexto, o vocabulário controlado torna-se o ponto de convergência entre as linguagens utilizadas por autores, indexadores e pesquisadores, premissa fundamental para comunicação de informações dentro de um sistema” (STREHL, 1998, p. 331).

Bocato e Fujita (2010) afirmam que os indexadores desenvolvem a atividade de tratamento temático da informação por meio da adoção de sistemas de organização do conhecimento. O uso de um vocabulário controlado para a organização da informação amplia o acesso à mesma, visto que o próprio vocabulário controlado acaba por ser um retrato de um conjunto de documentos disponíveis no catálogo de uma base de dados, e ainda delimita conceitos, define termos e remete à informação.

Agrega-se qualidade, com o uso de um vocabulário controlado, ao sistema de informação tanto no processo de indexação da informação como na recuperação, evitando que diferentes termos com mesmo significado venham a representar conteúdos informacionais idênticos, causando uma dispersão dos resultados de pesquisa. Para tanto, é necessário que o vocabulário controlado, que tem por objetivo servir de ligação entre o usuário e o documento por ele representado, funcione também como uma forma de controle terminológico no que diz respeito às variantes terminológicas.

O vocabulário controlado pode conter uma rede de remissivas em sua estrutura, favorecendo, desse modo, a recuperação da informação que possua diferentes formas de representação. O Código de Catalogação Anglo-Americano (2004, p. D-12) afirma que remissiva é a “[...] indicação remetendo de um cabeçalho ou entrada para outra”. Independentemente do termo utilizado pelo usuário ao realizar a busca informacional, a recuperação se dará por qualquer forma utilizada, uma vez que a estrutura da linguagem documentária que possua uma rede de remissivas pode efetuar esta recuperação.

Ao levar em consideração as variantes terminológicas, torna-se possível estruturar um vocabulário controlado que incorpore uma rede de remissivas dentro de um sistema de informação. Isso ampliaria a padronização na representação informacional, além de aumentar o alcance de recuperação da informação por parte do sistema de informação, já que as diversas formas de representação de um assunto podem estar compreendidas na linguagem documentária. A linguagem utilizada pelo usuário em suas buscas informacionais pode ser uma das fontes de coleta de termos para a rede de remissivas, ao passo que pode conter, em sua estrutura, diferentes ideias, entendimentos e usos da informação, gerando, assim, diversificadas formas de representação.

Dessa maneira, o processo de indexação é dependente das etapas de análise e tradução de termos selecionados de uma linguagem documentária. Boccato diz que

[...] o processo de representação mediante linguagem documentária conduzirá o bibliotecário indexador à escolha dos termos correspondentes à especificidade e exaustividade que a linguagem possui e, conseqüentemente, à especificidade e exaustividade do sistema (2009, p. 122).

Algumas linguagens documentárias procuram apresentar os termos ordenados, tanto sistemática quanto alfabeticamente, sendo fundamentalmente um conjunto de termos autorizados. Lancaster (2004, p.19) diz que “são três os tipos principais de vocabulários controlados: esquemas de classificação bibliográfica (como a Classificação Decimal de *Dewey*), lista de cabeçalhos de assuntos e tesouros”. Quando o usuário conhece qual é o vocabulário controlado utilizado pelos indexadores para representar a informação, facilita-se o processo de recuperação da informação. A não coincidência entre as expressões de busca e as unidades indexadoras impede que o processo comunicativo entre usuário e sistema se realize (ALVORCEM; VAN DER LAAN, 2007).

3.5 TESAURO

Como linguagem documentária, o tesauro tem como uma das principais características ser a ligação entre os termos, segundo critérios estabelecidos, de modo que alguns termos conduzam a outros (CURRÁS, 2005). Essencialmente uma linguagem de representação de uma especialidade, o tesauro é baseado numa estrutura conceitual de uma determinada área temática, onde cada conceito deve ser representado por uma única forma verbal (DODEBEI, 2002). De acordo com Moreira e Moura (2006, p.7), “[...] um tesauro é uma linguagem de documentação com a característica específica de possuir relações entre os termos que o compõe”. Assim,

[...] o tesauro surgiu da necessidade de manipulação de grande quantidade de documentos especializados, onde é preciso trabalhar com vocabulário mais específico e uma estrutura mais articulada e integrada do que aquela presente nos cabeçalhos de assunto (remissivas e referências cruzadas tipo 'ver' e 'ver também'). Parte do desejo de uma comunidade de usuários em recuperar documentos de uma área específica, na qual é necessária maior sistematização para a recuperação (MOREIRA; MOURA 2006, p. 7).

O tesauro tem como principal aplicação o controle terminológico – controlar sinônimos, homógrafos e quase sinônimos – a fim de melhorar o processo de indexação e recuperação da informação. Sua previsão de utilização está atrelada a uma política de indexação, que deve esclarecer qual linguagem documentária será utilizada. Barité (2007, p. 6, tradução nossa) diz que a “[...] tendência mais recente aponta que os tesouros e outras linguagens documentárias podem ajustar sua estrutura de conhecimento às estruturas cognitivas dos usuários, a sua linguagem e a sua maneira de buscar informação”.

O tesauro estrutura-se de uma forma a manter relação semântica entre seus termos, que pode ser hierárquica, associativa ou de equivalência. Segundo Currás (1995, p.88) “[...] tesauro é uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos linguísticos que o compõem, termos, simples ou compostos, encontram-se relacionados entre si sintática e semanticamente”. Os tesouros são reconhecidos como linguagens documentárias e são ferramentas de controle da terminologia por meio de termos estruturados que, quando disponibilizados aos usuários, facilitam a recuperação da informação, visto que o usuário estará a par da linguagem que é usada na indexação da informação.

Currás (1995) afirma que o tesauro deve cumprir algumas condições como: ser uma linguagem especializada que permita a introdução ou supressão de termos para manter a

atualidade; servir de conversor da linguagem natural dos documentos, para uma linguagem controlada, que possa representar a informação contida nos documentos; servir como uma ponte entre os documentos e os usuários; é, portanto, a linguagem que tanto o indexador como o usuário utilizarão.

Dentre as principais funções de um tesauro, destacam-se o controle de sinônimos e quase sinônimos e a distinção de homógrafos. O tesauro possibilita, ainda, a facilidade na condução da busca por intermédio dos termos relacionados e das referências cruzadas de maneira a permitir uma consistência melhor da indexação e do transporte da linguagem de busca para a linguagem de indexação. (OLIVEIRA; ARAÚJO 2012, p. 23).

O desenvolvimento dos tesouros parte do pressuposto do controle terminológico. Seu desenvolvimento se dá a partir das linguagens de especialidades das áreas científicas e da linguagem de busca do usuário, com vistas a criar um instrumento que compatibilize a linguagem do usuário com a linguagem utilizada pelo sistema (CURRÁS, 1995). O tesauro indica as diferentes relações, utilizando abreviaturas ou sinais gráficos pré-convencionados. Com base em Currás, (1995) e Dodebei (2002), as relações existentes no tesauro são:

- a) relação hierárquica: incluem as relações gênero e espécie, todo e parte, tipo de e exemplo de, onde dois conceitos diferentes possuem características idênticas, porém, quando comparados entre si, um apresenta uma característica de subordinação ao outro.
- b) relação associativa: apresenta as relações entre termos do mesmo nível hierárquico com proximidade de significado, mas que possuem características diferentes entre si.
- c) relação de equivalência: ocorre quando um conceito é representado por mais de um termo, sendo sinônimos ou quase sinônimos. Também ocorre para indicar preferência de grafia entre os termos que apresentam mais de uma.

A estrutura de um tesauro é que traz ao usuário ou ao indexador a possibilidade de encontrar o termo mais adequado, mesmo desconhecendo o nome específico para a ideia ou conceito. Além do mais, informa ao usuário outros termos, oportunos ou não, apresentados para sua escolha, através da própria estrutura do tesauro (MOREIRA; MOURA, 2006). Portanto, o tesauro é uma linguagem documentária complexa, em vista do relacionamento entre seus termos e possui caráter específico, visando à melhora da representação da informação por meio do controle da indexação; facilitando, também, a recuperação da informação.

3.6 O USUÁRIO

Sanz Casado (1994, p. 19) traz uma definição de usuário da informação como “[...] aquele indivíduo que necessita de informação para o desenvolvimento de suas atividades”. Imprescindível para o processo de indexação, o usuário ocupa papel decisivo em um sistema de informação e, ao mesmo tempo, é a parte envolvida que mais percebe as consequências, tanto positivas como negativas, das decisões que são tomadas na indexação, já que todas as ações realizadas durante o ciclo de tratamento informacional dentro de um sistema de informação, afetarão de alguma forma, a comunicação com o usuário.

Se é verdade que estamos presenciando a mudança de paradigma no que diz respeito a sistemas de informação, de centrado em bases de dados para centrado nos usuários, então é necessário que se pense melhor a organização do acervo e seu sistema de acesso. É necessário que isso se faça de modo a tornar o usuário cada vez mais independente dentro de uma biblioteca ou centro de informação. (GIGANTE, 1995, p.22).

Cada vez mais os sistemas de informação voltam seu olhar e atenções para o usuário, com o intuito de assegurar-lhes serviços que possam atender as suas demandas informacionais. O usuário, ao buscar informação, demonstra um desejo, expressa uma vontade. Assim Sanz Casado diz que a busca informacional é

[...] o processo no qual o indivíduo engaja-se decididamente em busca de informações capazes de mudar seu estado de conhecimento. Ele faz parte de uma atividade social por meio da qual a informação torna-se útil para o indivíduo ou para o grupo. Para uma mesma rede de informação, cada um de nós vai buscar a informação de maneira um tanto diferente, dependendo de nosso conhecimento das fontes, de nossas experiências passadas e assim por diante (1994, p. 48).

No entanto, nos sistemas não automatizados, em que a informação era organizada em fichas e catálogos manuais, o usuário contava com o auxílio direto da figura do mediador para acesso à informação, pois o sistema em si era complexo e melhor compreendido por quem o construía ou lidava com ele diariamente. Ferreira (1996, p. 4) afirma que “[...] tal situação ocorre, porque, historicamente, os sistemas de informação têm sido planejados em função das tecnologias utilizadas para sua implementação, ou do conteúdo da informação a ser inserida”. Ou seja, o usuário não estava sendo observado no processo organizacional da informação e, também, a linguagem documentária não era preponderante na relação de recuperação da informação em vista do usuário, já que o mesmo não lidava diretamente com a busca informacional.

Para resgate histórico, cabe ressaltar que a necessidade de se conhecer a relação do usuário com o sistema de informação sempre foi suprida pelos estudos de usuários, embora seus enfoques ainda não privilegiassem conhecer toda a extensão da relação entre ambos, ressaltando apenas alguns aspectos próprios do sistema. A observação do usuário, como peça chave na gestão do sistema de informação, não ocorre nos modelos tradicionais de pesquisa das décadas de 70 de 80, em que satisfação, necessidade de informação e desempenho não eram avaliados e observavam apenas o sistema de informação (COSTA; SILVA; RAMALHO 2009).

Com o surgimento das novas tecnologias da informação e comunicação, ocorreram mudanças significativas, inclusive nos suportes de registros do conhecimento, o que acabou gerando uma espécie de afastamento do usuário dos sistemas de informação, mas não da informação em si. Assim, o usuário, por meio de ferramentas tecnológicas, pode acessar conteúdos informacionais sem a necessidade de estar fisicamente no local em que se encontra o documento. Com isso, a relação do usuário de informação com o sistema de informação mudou, forçando a mudança da forma com que se avalia a sua relação com o sistema de informação. Fujita afirma que se podem obter por meio de estudos de usuários, as

[...] análises sobre preferências, percepções, dificuldades e procedimentos acerca do uso de serviços de informação para compará-los com os resultados com os resultados de outras pesquisas que visam à compreensão do contexto (2006, p.12).

Por conseguinte, Ferreira (2002) identifica uma abordagem alternativa aos estudos de usuários, em que seu direcionamento visa ao próprio usuário da informação, onde se considera que o comportamento do usuário, a necessidade de informação, o uso da informação e a forma como usuário se relaciona com o sistema de informação passam a constar na forma de abordagem. Assim, começa-se a observar o usuário como sendo construtivo, ativo, orientado situacionalmente, que focaliza aspectos cognitivos, possui individualidades e se orienta de forma qualitativa (FERREIRA, 2002).

A lógica básica por trás dessa perspectiva centrada no usuário é que os sistemas de informação devam ser modelados de acordo com o usuário, com a natureza de suas necessidades de informação e com seus padrões de comportamento na busca e no uso da informação, de modo a maximizar sua própria eficiência (FERREIRA, 2002, p. 6).

Nesse sentido, os sistemas de informação devem estar atentos aos serviços que prestam e à forma com que os executam, principalmente no que envolve a recuperação da informação, visto que o usuário cria expectativas sobre a sua intenção informacional, e uma

recuperação ineficiente acabaria por frustrá-lo. Dodebei (2002, p. 51) afirma que “[...] o processo de comunicação se dá, portanto, quando o indivíduo reduz o que pensa e o que quer a um sistema de convenções comunicativas, ou seja, quando o que pensa e quer é socializado”.

A diferença entre a linguagem utilizada pelo sistema e a utilizada pelo usuário, ao buscar informações, pode ter como resposta um alto índice de revocação com um baixo índice de precisão, sendo recuperados documentos não pertinentes aos interesses do pesquisador, gerando excesso de ruído. Ou então, ao contrário, documentos importantes não são recuperados, resultando no que se denomina de silêncio do SRI (VAN DER LAAN, 2002, p. 16).

Assim, conhecer o usuário e sua linguagem passa a ser uma obrigação dos sistemas de informação, pois ele pode contribuir com melhorias em todo o sistema, posto que, ao passo que utiliza o sistema, acaba por reconhecer todas as intempéries que enfrenta no processo de busca e recuperação da informação. Guinchat e Menou (1994, p. 482) afirmam, portanto, que “[...] o usuário é um agente essencial na concepção, avaliação, enriquecimento, adaptação, estímulo e funcionamento de qualquer sistema de informação”. Bocatto e Fujita dizem que:

No paradigma cognitivo, o usuário se coloca em uma posição ativa que elabora ideias, conceitos, etc., em um contexto individual. Esse cenário é visto pela ótica da dimensão humana, isto é, pela perspectiva centrada no usuário, pois os sistemas de recuperação da informação devem ser modelados de acordo com o usuário, com a natureza de suas necessidades de informação e com seus padrões de comportamento na busca e no uso da informação (2010, p. 29).

No exercício de busca e recuperação da informação, provenientes de uma necessidade informacional, o usuário, por meio de seu entendimento e perspectiva, formula estratégias de busca que possuem características individuais dentro de um determinado contexto. Araújo Júnior (2007, p. 73) define a formulação da busca como a “[...] formulação da questão que deve representar a necessidade informacional [...]” e define busca como a “[...] confrontação das representações dos conteúdos dos documentos com a questão formulada para representar as necessidades de informação”.

Nesse ponto, a linguagem de indexação deve oferecer compatibilidade com a linguagem do usuário, seja por meio da utilização de remissivas, controle das variantes terminológicas ou utilização de termos prediletos ou mais utilizados pelos usuários. Além disso, a linguagem de indexação deve ser o instrumento que padroniza a descrição da informação e dá qualidade à recuperação da informação, por ser o canal de comunicação que será utilizado na busca informacional. Lopes afirma

[...] que o uso da linguagem natural tem dificuldades a serem superadas, da mesma forma que a linguagem controlada apresenta certas desvantagens, pois os termos preferidos pelos indexadores frequentemente não são os termos utilizados pelos usuários em situações específicas de busca, pois, em grandes bases de dados, o processo de indexação envolvendo diversos indexadores provoca, com certeza, inconsistências na identificação de documentos similares (2002, p.47).

A diferença entre os termos utilizados pelos indexadores para a representação temática e os termos utilizados pelos usuários no processo de busca informacional não representa necessariamente uma falha de representação, visto que a indexação realizada no processo de tratamento da informação ocorre com base na utilização de uma linguagem de indexação, que se origina após a análise conceitual. Barité (2007) introduz que o processo de seleção de termos deve estar associado às formulações de busca que os usuários realizam ante um sistema de informação e na forma em que elas se manifestam.

Ocorre que, por vezes, a linguagem de indexação utilizada no sistema de informação é desconhecida do usuário, está inacessível ou a forma de apresentação de sua estrutura não é reconhecida pelo usuário. Para que o processo comunicativo entre o usuário e a base de dados aconteça, a linguagem do usuário precisa ser traduzida para a linguagem controlada da base, pelo próprio usuário ou por um intermediário (LOPES, 2002). Lancaster (2004, p. 12) identifica que “[...] é preciso que os indexadores saibam muito mais do que os princípios da indexação. Devem, em especial, estar inteiramente a par dos interesses da comunidade atendida e das necessidades de informação de seus membros”.

Assim, é fundamental que o usuário esteja a par da linguagem utilizada pelos indexadores no processo de representação temática, caso contrário, o mesmo não atingirá e nem terá sua necessidade de informação atendida. Portanto, além de o usuário ter alguma proficiência na linguagem de indexação, é necessário que o indexador também a conheça com profundidade e saiba utilizá-la. Para o caso de isso não ocorrer, Lopes (2002, p. 47) diz que “[...] as deficiências são inevitáveis, principalmente quando ambos desconhecem as linguagens controladas da base de dados a ser consultada”. Oliveira e Araújo afirmam que

[...] nesse percurso, as linguagens documentárias têm contribuído no atendimento à busca pela informação mediante seus recursos, como os vocabulários controlados, utilizados para possibilitar maior eficácia na recuperação da informação (2012, p. 27).

Na caminhada que o usuário realiza, dentro do processo de busca e recuperação da informação, ele se vale de suas experiências acadêmicas ou profissionais, de seu conhecimento e entendimento próprios sobre determinado assunto, que acabam por

manifestar-se nas estratégias de busca, mais explicitamente nas expressões de busca utilizadas, visto que essas não são passíveis de pré-determinação de critérios, como pode ocorrer numa estratégia de busca com campos pré-definidos como aquela que recupere um tipo específico de informação, por exemplo. Essas ações dos usuários, suas estratégias de busca e expressões utilizadas podem vir a estabelecer uma relação direta com a recuperação ou não da informação, mostrando, assim, a usabilidade da informação e a eficiência do sistema de recuperação da informação.

A partir dessa perspectiva de reconhecimento que o usuário pode ou não estar tendo suas necessidades informacionais atendidas, o que já pauta os estudos de usuários, abre-se uma possibilidade de estudo das expressões de busca dos usuários para a recuperação da informação com o intuito de verificar o nível de adequação das mesmas à linguagem de indexação utilizada pelo sistema de informação. Dentro das possibilidades de busca oferecidas pelo sistema de informação, destacam-se dois tipos mais comuns: a pesquisa básica e a pesquisa avançada. Nas pesquisas básicas, a recuperação se dá por meio de busca a partir da expressão utilizada em todo o registro bibliográfico; já nas pesquisas avançadas, é possível especificar mais a pesquisa, através de múltiplos campos de pesquisa. Nessas pesquisas o usuário emprega estratégias de busca que, conforme seu entendimento, pode conduzi-lo à informação necessitada.

Dessa maneira, o usuário produz estratégias para a recuperação da informação que, com o auxílio de mecanismos da tecnologia da informação, podem vir a ser armazenadas e servirem como fonte de estudo sobre o usuário e as linguagens utilizadas por ele para a recuperação da informação. Surge como uma dessas formas de avaliação/acompanhamento, a análise de *logs*, que consistem em “[...] um arquivo eletrônico gerado pelo sistema de automação de bibliotecas, onde estão guardadas todas as ações feitas pelo usuário durante a utilização do sistema” (PONTES, 2006, p. 42). Em geral, os sistemas de armazenamento de informação oferecem parâmetros de pesquisa predefinidos para os diferentes tipos de busca que podem ser utilizados pelos usuários. Dentro de toda a ação de busca de informação por parte do usuário geram-se dados de pesquisa, definidos por Balby (2002, p. 3) de “[...] arquivos de *log* de transações de busca” com informações que reproduzem a forma com que o usuário pesquisou na base de dados, conhecidos como *logs* de pesquisa, que são passíveis de uma análise e que podem produzir dados para subsidiar uma melhora na indexação.

As expressões de busca dos usuários fazem parte destes *logs* de acesso, visto que neles são armazenadas todas as informações e ações executadas pelo usuário na busca informacional. As expressões de busca se revestem de interesse para pesquisa e, também, para

a prática da indexação, pois comportam a linguagem que o usuário utilizou, independentemente de ela ser especializada ou não, visto que a terminologia de uma área do conhecimento pode ser muito diversificada no que diz respeito aos termos que podem representar o mesmo conceito. Os aspectos que envolvem a terminologia de uma área e que podem influenciar tanto na indexação como na recuperação da informação serão vistos a seguir.

Inicialmente, identificam-se possíveis pontos positivos neste tipo de avaliação que se relacionam ao fato de que o acompanhamento das formas e expressões de busca pelos usuários se dá de forma distanciada, não oferecendo, assim, possibilidade de interferência por parte do pesquisador. Cada vez mais a relação do usuário com o sistema de informação se torna remota, fato gerado pela disponibilização da informação na Internet em um contexto geral. Nesse sentido, Pontes (2006, p. 41) diz que “[...] os usuários podem visitar a biblioteca sem sair de casa e assim não dispensar tempo em ir pessoalmente à biblioteca e não encontrar a informação que procura”.

Essa possibilidade, no entanto, remete a certa preocupação, ao passo que o usuário torna-se cada vez mais independente no processo de busca informacional, distanciando-se da biblioteca física e do bibliotecário. Não se pode considerar isto um problema, mas é inegável que dificulta uma relação mais próxima com o usuário, bem como o reconhecimento de eventuais problemas de comunicação entre ele e o sistema de informação.

4 TERMINOLOGIA

O desenvolvimento científico de uma determinada área promove a ampliação do campo comunicacional da mesma por meio do incremento de novos elementos terminológicos, com a função de descrever os avanços científicos. O reconhecimento e o estudo desses fenômenos estão ligados intimamente ao estabelecimento de uma terminologia, ou seja, à produção de um conjunto de termos especializados que compõe seus domínios, fato que se desenvolveu com a Terminologia, a partir de estudos sobre a expansão linguística e sobre o comportamento de aspectos comunicativos das linguagens especializadas.

A Terminologia (grafada com T maiúsculo), enquanto ciência, trata especificamente de um campo do conhecimento que tem por particularidade o léxico especializado, ou seja, analisa as unidades terminológicas que são concebidas com o avanço dos estudos e conhecimentos de uma determinada área. Já a terminologia (grafada com t minúsculo) constitui-se em um conjunto de termos provenientes de uma determinada área científica que compõe os instrumentos de comunicação profissional (KRIEGER e FINATTO, 2004).

No século XX, surge a Terminologia, ciência delimitada a partir de Eugen Wüster, que, por meio de seus estudos sobre a comunicação técnico-científica, desenvolveu a Teoria Geral da Terminologia (TGT), que tinha por objetivo eliminar as ambiguidades que existiam na comunicação científica através do estabelecimento de metodologia prescritiva e da normalização dos termos. Através de seus estudos, o autor mostrou a necessidade do desenvolvimento de uma ciência capaz de compreender a terminologia como uma prática necessária à ampliação da comunicação do conhecimento técnico-especializado por meio da formalização de instrumentos que representassem os conceitos dos termos utilizados por especialistas de uma determinada área (CABRÉ, 2005).

A Terminologia concebida por Wüster comportava âmbitos sociopolítico, acadêmico e científico, visto que o estudo de uma área deveria ser amparado pelo seu reconhecimento e uso por parte de especialistas, sendo, portanto, “[...] objeto de demanda de profissionais dedicados à resolução específica dessas necessidades” (CABRÉ, 2005, p. 2, tradução nossa). O conhecimento da terminologia decorria da estruturação de conceitos de uma determinada área para a sua representação por meio de um signo linguístico, trabalho este que poderia apenas ser realizado por especialista da área, únicos possuidores do conhecimento necessário para tal. Cabe ressaltar, que, na TGT, o estabelecimento de conceitos não seguia regras, era fruto de uma convenção entre especialistas e era avaliado apenas dentro do domínio especializado, assim, por meio de onomasiologia e com caráter normativo, seria selecionado o

termo que representaria o conceito idealizado. Esta forma de estruturação da linguagem especializada poderia não atender às necessidades reais de comunicação, que é uma prerrogativa de qualquer área, visto que não representaria, necessariamente, todas as formas de ocorrência da informação e era compreendida apenas dentro do contexto especializado a que se referia.

As principais críticas feitas a TGT estão no posicionamento uniformizador da terminologia, não comportando dentro de si as variantes terminológicas, a dimensão comunicativa dos termos e a incapacidade de explicar a comunicação científica no todo. Wüster, além de ponderar a terminologia para o desenvolvimento de instrumentos normalizadores e prescritivos, não comportou na TGT aspectos linguísticos, cognitivos e sociais da interdisciplinaridade da terminologia falada, visto que afastou a mesma da linguagem natural, direcionando-a para o conhecimento (linguagem) especializado (CABRÉ, 1999).

Em oposição à Teoria Geral da Terminologia, os estudos de M. Teresa Cabré trazem uma nova perspectiva sobre a abordagem terminológica nos processos comunicativos, a partir de uma nova teoria, a Teoria Comunicativa da Terminologia (TCT) que será debatida a seguir e respaldará a construção desta pesquisa.

4.1 TEORIA COMUNICATIVA DA TERMINOLOGIA

Pretendendo estabelecer uma linha de análise terminológica para o âmbito da comunicação especializada, mais especificamente para a área do Direito Penal, adotaremos a Teoria Comunicativa da Terminologia (TCT) (CABRÉ, 1999) como base para a construção desta investigação, visto que se pretende buscar na linguagem do usuário especializado ou não (público em geral) as variantes terminológicas que possam agregar ao processo de indexação maior consistência, sendo que estas variantes terminológicas são reconhecidas pela TCT. Além disso, busca-se a definição de termos, conceito e de variantes terminológicas que se fazem necessárias para contextualizar a abordagem metodológica.

A Teoria Comunicativa da Terminologia, segundo (Krieger e Finatto, 2004, p.35) “[...] articula-se baseada na valorização dos aspectos comunicativos das linguagens especializadas em detrimento dos propósitos normalizadores, bem como na compreensão de que unidades terminológicas formam parte da linguagem natural e da gramática das línguas”.

Tanto o conhecimento especializado quanto os textos especializados, como unidades terminológicas podem ocorrer em diferentes níveis de especialização e serem

descritas em diferentes níveis de representação. Só assim a terminologia do desejo passa a ser a terminologia da realidade (CABRÉ, 1999, p.126).

Ressaltando o entendimento de que a preocupação primordial da Terminologia é o termo como componente lexical característico das comunicações especializadas, que permite a representação e transferência de conhecimento, a constituição de produtos terminológicos (dicionários, tesouros) pode constituir um conjunto de unidades terminológicas de uma área especializada (CABRÉ, 2001b). Assim, a Terminologia constitui-se numa área de ampliação do conhecimento sobre o próprio conhecimento de uma área especializada, estudado por meio da linguagem utilizada nas suas trocas informacionais e comunicacionais necessárias para o seu funcionamento.

Toda e qualquer área do conhecimento, ao constituir-se como especializada, produz um conjunto terminológico de seus saberes que são utilizados como componentes das comunicações especializadas, com a finalidade representacional e comunicacional de fixação e circulação do conhecimento. Esse conjunto é conhecido como a terminologia da área, que é encontrada, fundamentalmente, nos textos especializados utilizados na comunicação técnico-científica, sendo considerada pelos especialistas como o reflexo formal da organização conceitual de uma área especializada e como um meio de expressão e comunicação profissional (CABRÉ, 1999).

Muitas são as peculiaridades que circunscrevem a terminologia de um conhecimento científico e que podem influenciar nos processos de representação e comunicação do conhecimento. Essas peculiaridades estão ligadas às ambiguidades, às polissemias, às variantes terminológicas e a outros aspectos relacionados à denominação do conceito do termo enquanto componente linguístico. Esse comportamento é determinado pelo caráter interdisciplinar e transdisciplinar que o conhecimento especializado adquire em função da sua constituição e uso. A terminologia, enquanto conjunto de termos do discurso especializado, é necessária para a comunicação especializada, devido a sua dupla função, a de representar o conhecimento especializado e de transferência deste conhecimento (CABRÉ, 2001b).

A Teoria Comunicativa da Terminologia possui como principal característica o reconhecimento do caráter comunicativo do termo, de sua poliedricidade e de sua natureza de elemento pertencente à linguagem natural. Esse entendimento, que compreende a terminologia de uma área especializada, permite ressaltar a perspectiva comunicacional do termo e sua constante evolução num âmbito linguístico. O princípio do caráter comunicativo da terminologia é uma das características da TCT evidenciado por sua interdisciplinaridade,

abrangendo, principalmente, aspectos da Linguística, quando trata da produção de instrumentos de referência especializados.

Assim, para Cabré (2001a, p. 28) o termo, dentro da TCT, tem caráter descritivo

[...] são unidades com dupla função, representativa e comunicativa. Desde o ponto de vista de sua capacidade de representação permitindo referir-se a uma realidade especializada, com diferentes níveis de especialização [...]. Desde o ponto de vista de sua capacidade de transferência, permitindo a comunicação em diferentes níveis, de maior ou menor especialização [...].

Kostina (2000, p. 20, tradução nossa) afirma que “[...] os termos ou unidades terminológicas em primeiro lugar são unidades linguísticas que possuem todas as particularidades da linguagem natural, inclusive participando da polissemia, da sinonímia e da homonímia [...]”.

A Teoria Comunicativa da Terminologia é constituída de princípios que embasam o tratamento terminológico e que são aplicados na pesquisa realizada, visto que tal teoria admite as variações terminológicas e considera o termo como uma unidade denominativo-conceitual, ou seja, como uma unidade de conhecimento com dimensão textual e discursiva de ocorrência no âmbito especializado.

Segundo CABRÉ (1999), a Teoria Comunicativa da Terminologia é fundamentada nos seguintes princípios:

- a) possui caráter interdisciplinar, que inclui aspectos linguísticos, cognitivos e sociais, caracterizando, assim, a poliedricidade do termo;
- b) observa o conhecimento geral e o especializado sem dissociá-lo na competência do falante especialista, devendo considerar as diferenças do conhecimento especializado e a forma com que é compreendido cognitivamente pelo falante;
- c) deve considerar a interdisciplinaridade dos termos, seu caráter poliédrico em função das diferentes aplicações terminológicas em face a sua atribuição representativa em diferentes áreas especializadas;
- d) deve considerar como um conceito pode formar parte de uma estrutura de diferentes áreas, observando a utilização conceitual. Este princípio aporta a definição de que nenhum termo pertence de forma natural a nenhum âmbito, mas são usados em determinados âmbitos (situações), embora reconheça a existência da criação de termos dentro de âmbitos especializados.

- e) considera o caráter polissêmico das unidades léxicas, em que se deve descrever as características especializadas que, em determinada situação, atribuem um caráter de termo a unidade léxica;
- f) admite a sinonímia como um fenômeno natural da comunicação especializada, devendo estabelecer critérios para distingui-las. Deve-se observar que a variação denominativa será menor quanto maior for a especialização do texto e que as unidades terminológicas em relação à sinonímia podem ter valor similar ou bastante distinto, dependendo das características pragmáticas aplicadas a cada variante;
- g) admite que as unidades terminológicas (termos) se dão de forma natural no discurso especializado e, em consequência, têm uma projeção sintática para além dos seus limites denominativos e variam como uma função do uso. A descrição do termo não pode limitar-se a sua categoria, âmbito temático, definição e equivalentes de uma língua, devendo-se ampliar para a associação de uso, até sua estrutura de argumentos, contexto de aparição e dentro da fraseologia.
- h) deve considerar a variação do discurso e definir as variáveis relevantes que descrevem esta variação na comunicação em geral e na especializada, principalmente. Assim, poderá se considerar a perspectiva temática, a abordagem que se está dando a um tema, o grau de formalidade, o propósito, o tipo de discurso, etc.

Ainda, CABRÉ (1999) define os fundamentos da Teoria Comunicativa da Terminologia, que são importantes para contemplar os princípios estabelecidos e promover a Terminologia como um campo interdisciplinar. Esses fundamentos se constituem de três teorias:

- a) teoria do conhecimento, que explica como é conceituada a realidade, podendo estabelecer como pode constituir-se a relação dos conceitos entre si e com suas possíveis designações.
- b) teoria da comunicação, que descreve as correlações entre tipo de situação e tipo de comunicação, com todos os critérios que podem ampliar, diversificar e explicar as características, possibilidades e limites dos sistemas de expressão de um conceito e de suas unidades.
- c) teoria da linguagem, que trabalha com as unidades terminológicas dentro da linguagem natural, dando conta de seu caráter terminológico, inclusive no processo comunicativo.

4.1.1 Termo e Conceito

Caracteriza-se como o objeto de estudo da Terminologia as unidades terminológicas que formam a linguagem de especialidade de determinada área. O conjunto terminológico é formado por termos, descritos como unidades denominativo-conceituais com capacidade de referência (BARROS, 2004).

O termo é definido por sua dimensão conceitual, ou seja, a unidade lexical passa a ser termo quando é integrante de uma determinada área científica. O seu comportamento, como portador de elementos especializados, confere ao termo uma qualificação de unidade de conhecimento. A consolidação do conceito no âmbito especializado faz com que o seu conteúdo seja relativo a situações de uso no discurso especializado e, assim, represente um conhecimento do domínio em que se situa. Barros (2004, p. 40) diz que o termo é “[...] uma unidade lexical com conteúdo específico dentro de um domínio especializado”.

O conceito, como unidade de conhecimento especializado, se apresenta como a compreensão mental da unidade de significação especializada, ou seja, é a formalização de um conhecimento por meio de uma definição correspondente a um conteúdo distinto e individualizador. Kostina (2000, p. 19, tradução nossa) afirma que “[...] um conceito é a representação mental de um termo que se estabelece sobre a base de oposição nítida entre conceitos”.

O caráter distinto que o termo pode estabelecer está ligado à função de contexto e situação de ocorrência. Assim, os termos são unidades léxicas que se compõem de forma ou de denominação e significado de conteúdo adequado ao tipo de ocorrência e determinados pelo âmbito do assunto (CABRÉ, 1999).

Os termos são unidades de forma e conteúdo em que o conteúdo é simultâneo à forma. Um conteúdo pode ser expresso com maior ou menor rigor por outras denominações do sistema linguístico, e constituem uma nova unidade linguística de conteúdo especializado relacionado semanticamente com a primeira ou de outros sistemas simbólicos e forma uma unidade linguística de conteúdo especializado. O conteúdo de um termo nunca é absoluto, mas sim relativo segundo cada âmbito e situação de uso (CABRÉ, 1999, p. 123).

Seguindo os fundamentos estabelecidos por Cabré (1999), os conceitos de um mesmo âmbito especializado mantêm entre si relações de diferentes tipos, constituindo, assim, a estrutura conceitual de uma área. Devido ao fato de que cada âmbito especializado pode ser estruturado de formas distintas, os conceitos podem participar em mais de uma estrutura com o mesmo ou diferente valor. Assim, Cabré (1999, p. 124) afirma que “[...] os termos não

pertencem a um âmbito, mas sim são usados em um âmbito com valor singularmente específico [...]”.

O trabalho terminológico dentro da Teoria Comunicativa da Terminologia tem a função de descrever as características das unidades léxicas que podem vir a constituir uma terminologia especializada, explicando as relações entre a constituição de uma estrutura conceitual de uma área e permitindo a representação e a transferência do conhecimento especializado. A comunicação especializada, assim como a comunicação geral, se manifesta por meio da linguagem, linguagem esta construída pelo conhecimento científico manifesta por meio de unidade linguística, o termo, que designa um conceito (KRIEGER e FINATTO, 2004).

Cabré (2008, p. 3, tradução nossa) define o termo como “[...] uma unidade de forma e conteúdo que indissociável entre si que representam, no plano da verbalização ou expressão, um conceito”. A função do termo é, ao mesmo tempo, designativa, pelo fato de apontar a uma referência; denominativa, por dar nome a um conceito e significativa, pois o termo em si possui informações sobre o conteúdo que transmite (CABRÉ, 2008).

O termo enquanto elemento fundamental de análise da Terminologia e componente de um discurso especializado é compreendido de forma poliédrica pela TCT, ou seja, possui uma dimensão linguística, uma cognitiva e outra social, como já mencionado. A implicação dessa proposição trata da dimensão social que é estabelecida pela formação do objeto e das classes de objetos; a dimensão linguística é evidenciada pelo plano semiótico, que abrange a construção de cada conceito e uma ou mais unidades terminológicas da língua natural, ou até mesmo a criação de unidades terminológicas artificiais; e a dimensão cognitiva diz respeito à formação do conceito e seu entendimento, compreensão mental (CABRÉ, 2008).

O termo pode ser compreendido pela associação direta entre um conceito e suas representações, isto é, abre margem para o tratamento das variantes no plano terminológico. Esse entendimento é provocado pelo fato da forma de um termo e o seu conceito se darem de modo independente, ou seja, o conceito é independente da sua denominação (CABRÉ, 2008). O termo pode ser constituído de uma palavra apenas ou também por meio de um sintagma terminológico que tem a mesma função representacional e comunicativa. Os sintagmas terminológicos constituem uma única unidade de sentido, a partir da junção de palavras estruturadas em uma determinada ordem sintática (VAN DER LAAN, 2002).

Kostina (2000, p. 19, tradução nossa) diz que “[...] o termo é uma unidade léxica específica da linguagem especializada que denomina os conceitos especializados [...]”. Assim, a comunicação especializada usufrui de elementos próprios de sua terminologia para

promover o processo comunicativo, compreendendo, por conseguinte, a relação do termo com o vocabulário de natureza científica expresso nos textos de especialistas.

As variações terminológicas são compreendidas na TCT como uma ocorrência natural oriunda do processo comunicacional, em face da origem da unidade terminológica direcionada à comunicação especializada em diferentes níveis. O fato de o léxico especializado ter origem na linguagem natural e de que a unidade léxica somente passa a ter condição de termo quando é empregada em textos de área especializada gera a possibilidade de ocorrência de mais de um signo linguístico com capacidade de representação conceitual.

4.1.2 Variantes terminológicas

A partir da perspectiva poliédrica atribuída pela Teoria Comunicativa da Terminologia ao termo, onde uma das dimensões é a linguística, é possível a perspectiva de ocorrência da variante terminológica. O registro de um conceito e sua apropriação pela comunidade científica não inibe o surgimento de mais de uma unidade lexical diferente para representar o mesmo conceito, abrindo caminho para a ocorrência da variante terminológica.

Segundo Suárez e Cabré (2002, p. 1, tradução nossa) “[...] a diversidade linguística se manifesta mediante diferentes fenômenos, entre os quais se destaca a variação das unidades linguísticas em todos os níveis da gramática”. A variação terminológica contempla a diversidade linguística determinada pelas características e necessidades da comunicação, mesmo no contexto especializado. A variação terminológica no processo comunicativo pode ser explicitada devido à ocorrência de mais de uma maneira de denominar um conceito, que seria o caso das sinonímias ou das situações de polissemia.

A Teoria Comunicativa da Terminologia de Cabré (1999) abriu caminho para a superação da ideia da univocidade e do *status* normalizador da Terminologia, no que diz respeito à forma de denominação de um conceito. Para Suárez e Cabré (2002, p. 2, tradução nossa) a comunicação científica manifestada por meio de textos especializados adquire formas distintas para referir-se a um mesmo conceito “[...] com a finalidade de evitar a repetição de uma mesma forma no discurso [...]”, pois “[...] uma vez introduzido o conceito [no contexto especializado], é comum referir-se ao mesmo por meio de expressões alternativas [...]”. Tais expressões alternativas são conhecidas como variantes terminológicas e podem ser formadas por unidades monoléxicas ou sintagmas terminológicos, que possuem valores semânticos equivalentes.

Freixa (2002) entende que os textos especializados divergem quanto aos níveis de especialização, níveis sociais e comunicativos, utilizando, para tal, uma terminologia que atenda e se adapte ao contexto, de modo que, para tanto, se faça uso no discurso especializado de diferentes denominações para tratar de um mesmo assunto de uma especialidade, resultando, assim, na variação terminológica. O resultado mais evidente dessa postura comunicativa é a terminologia alternativa que se cria para atender as necessidades de cada situação comunicativa, onde vários termos podem corresponder mais ou menos ao mesmo conceito, criando uma espécie de competição terminológica, ou que o mesmo termo parece corresponder a diferentes denominações, no todo ou em parte, ou seja, ocorre uma terminologia polissêmica.

A ocorrência das variantes terminológicas se dá no contexto da comunicação científica, visto que é nesse contexto que o processo comunicativo ocorre, ou seja, é nesse contexto que a conversão dos resultados da investigação científica se torna um novo conhecimento. Por isso, o processo comunicativo configura-se em um sistema de comunicação composto por três elementos básicos, que são o emissor, o canal de comunicação e o receptor (STUMPF, 2000).

Weitzel (2006, p. 88) diz que a “[...] comunicação científica pode ser entendida como um processo que envolve a construção, comunicação e uso do conhecimento científico para possibilitar a promoção de sua evolução”. A natureza especializada da comunicação científica lhe concede a propriedade de manifestação diversa de seus discursos com base no direcionamento que lhe é dado. Quando da comunicação entre pares, é tendência que o texto especializado adquira características de uma linguagem mais particularizada, com uma terminologia que se adapte ao direcionamento que lhe é dado. Ou seja, a unidade léxica utilizada no discurso especializado logicamente possui características de especialização, o mesmo podendo não ocorrer quando a comunicação que se deseja fazer é voltada para o público leigo, por exemplo.

É na linguagem especializada, mais especificamente nas situações comunicativas, e em função das próprias necessidades comunicativas que se manifestam as variantes terminológicas; em decorrência da possibilidade gramatical oferecida pela língua ou linguagens de especialidade que compõem um conjunto de meios linguísticos próprios de um campo do conhecimento (BARROS, 2004). Cabré (1993, p. 125, tradução nossa) diz que “[...] uma língua é um sistema complexo e heterogêneo de subsistemas inter-relacionados, em que cada um dos quais é suscetível de ser descrito em distintos níveis [...]”. Para tanto, Cabré (1993) afirma que os elementos mais importantes de caracterização de uma língua são aqueles

que também compõem o léxico especializado, e que se manifestam por meio de diferentes níveis de descrição linguística, a saber:

- a) nível fonológico – fonema;
- b) nível morfológico – morfema;
- c) nível léxico – palavra;
- d) nível sintático – oração;
- e) nível discursivo – texto.

Embora estabelecido em forma de níveis, para Cabré (1993, p. 125, tradução nossa), o sistema linguístico “[...] não se limita aos distintos níveis descritivos: mas se manifesta também em uma série de modalidades denominadas variedades dialéticas e funcionais [...]”. Portanto, não há limitação quanto às unidades alternativas substituírem às básicas dentro de um conjunto linguístico, ou até mesmo unidades que compõem mais de um nível de descrição, como é o caso dos sintagmas terminológicos ou dos fragmentos do discurso. Assim, o nível de descrição gramatical está para Cabré (1993, p. 126, tradução nossa) “[...] condicionado pelas características das situações comunicativas”.

O contexto da comunicação científica é manifesto por meio do texto especializado, definido por Barros (2004, p. 44) como os “[...] discursos orais ou escritos produzidos por uma área do saber [...] que [...] têm por objetivo maior transmitir uma informação [...]”. Sendo esse o principal objetivo do texto especializado, é de entendimento que o mesmo usufrua de uma linguagem que aborde aspectos próprios da terminologia da área.

Segundo Barros (2004, p. 44) “[...] os discursos especializados também subjazem normas discursivas próprias [...], por tanto, a eficácia do discurso especializado também está ligada à argumentação, com a utilização do léxico especializado em diferentes níveis”. A propagação do conhecimento se dá por meio do texto especializado, que se utiliza de unidades léxicas específicas do domínio em questão e que são pertinentes no nível do discurso. Além do mais, adaptam-se a quem este texto se destina e ao canal de comunicação utilizado. Pode-se assim identificar a ocorrência da variação terminológica, uma vez que a destinação da informação e o meio de comunicação utilizado no âmbito profissional podem requerer um nível maior ou menor de especialização do texto científico.

A linguagem de especialidade ou especializada, manifestada por meio da comunicação especializada ou científica, campo de estudo da Terminologia, é considerada por Barros (2004, p. 43) como “[...] as manifestações e potencialidades de um sistema linguístico [...]”, conhecida também como “tecnoleto”, um subsistema de uma língua geral que possui na linguagem natural suas bases constitutivas. Portanto, o termo adquire valor terminológico

devido ao seu uso por um domínio especializado “[...] em toda sua variada amplitude, temática, perspectiva, nível de especialização, propósito comunicativo, propósito funcional, tipo de texto, tipo de discurso, etc...” (CABRÉ, 1993, p. 120). Assim, para Barros (2004, p.59) “[...] a sinonímia, a homonímia, a polissemia e a variação linguística (léxica) de diferentes tipos passam a ser previstas, aceitas e tratadas em um estudo terminológico de perspectiva comunicativa”.

Por conseguinte, cabe um recorte referencial de separação da língua geral e da língua especializada (linguagem de especialidade), para que se possa caracterizar a variante terminológica. Cabré (1993) ressalta que determinadas características de uma linguagem são o elo de estabelecimento de sua especialização e, conseqüentemente, de suas peculiaridades enquanto texto comunicativo especializado. Assim, para Cabré, uma linguagem de especialidade é

[...] um conjunto de possibilidades determinadas pelos elementos que intervêm em cada ato de comunicação: os interlocutores (emissores e destinatários, com todo o conjunto de características que lhes são próprias), as circunstâncias comunicativas e os propósitos ou intenções que se propõe o ato de comunicação (1993, p. 129).

As diferenças que produzem uma delimitação entre a linguagem especializada, onde se encontram as variantes terminológicas que se tem por objetivo analisar, e a linguagem geral são caracterizadas pelo contexto de ocorrência. Entende-se aqui como sendo parte deste contexto o texto de especialidade, o canal de comunicação e os indivíduos que participam deste processo.

Cabré (1993, p. 135, tradução nossa) atribui como característica de uma linguagem especialidade os seguintes elementos:

- a) trata-se de um conjunto especializado, por sua temática, por sua experiência, âmbito de utilização ou por seus usuários;
- b) apresenta-se como um conjunto com características inter-relacionadas e não como fenômenos isolados;
- c) mantém a função comunicativa como predominante sobre outras funções complementares.

A característica de especialidade de uma linguagem está ligada à temática representativa da mesma e a sua especialização, no que diz respeito aos aspectos comunicativos de intercâmbio de informação. Para tanto, a ocorrência desta comunicação está

muito ligada à situação de contexto da mesma, que no discurso textual tem sua principal forma de veiculação (CABRÉ, 1993).

Cabré (1993, p. 139-140) estabelece os critérios de definição de uma linguagem especializada, pressuposto para a análise terminológica, caracterizada por três variáveis: a temática, os usuários e as situações de comunicação:

- a) considera-se especializada a linguagem temática que forma o conhecimento geral dos seus falantes e que, portanto, é objeto para aprendizagem especializada;
- b) considera que os usuários de uma linguagem de especializada são os especialistas, os quais participam do processo comunicativo como produtores e receptores de informação especializada;
- c) a situação comunicativa pressupõe a utilização de códigos linguísticos apropriados, ou seja, utiliza-se um subconjunto da língua regulados por critérios profissionais ou científicos, impondo uma formalidade para a comunicação;
- d) são considerados como subconjuntos especializados de uma língua, que se regulam por critérios das situações comunicativas, apresentando características específicas para o tipo linguístico (unidades linguísticas utilizadas e suas regras) e para o texto, em função de seu tipo e tipo de documento;
- e) entende-se que a linguagem especializada apresenta uma variedade alternativa de expressões em função do uso e características comunicacionais, emergindo, assim, as variantes terminológicas. Estas, por sua vez, têm sua ocorrência ligada à temática, aos destinatários da informação e à função comunicativa que o emissor se propõe no texto, abrindo espaço para a utilização de termos com a mesma função em textos especializados, porém, com propósitos comunicativos diferentes. Além disso, podem atuar como fatores na ocorrência de variantes terminológicas os dialetos geográficos, históricos e sociais, além do estilo pessoal do autor para a construção do texto;
- f) as linguagens especializadas possuem características da linguagem natural, podendo, então, afirmar que a mesma é um subconjunto da linguagem natural e;
- g) considera-se que a linguagem especializada mantém uma relação próxima com a linguagem natural, inclusive existindo uma intersecção entre a linguagem natural e mais de uma linguagem especializada, onde as mesmas compartilham características, unidades e convenções de uso.

Cabré (1993, p. 142, tradução nossa) afirma que “[...] dentro de cada linguagem especializada determinada pela temática, podemos verificar distintos graus de abstração, que

conduzem a diferentes níveis ou estilos discursivos”. Esses níveis discursivos são responsáveis pela ocorrência das variações terminológicas que indicam diferentes denominações para um mesmo conceito. Sua ocorrência acontece devido à necessidade de adequar o texto à situação comunicacional proposta.

Adentrando na questão da ocorrência das variantes terminológicas, Freixa (2002) compreende tais variantes como parte do processo linguístico, ou seja, analisa-as dentro dos pressupostos teóricos que embasam a Teoria Geral da Terminologia de Cabré (1993). Freixa (2002, p. 20, tradução nossa) observa que a variante possui um caráter linguístico integrado a um fenômeno de “[...] compreensão e descrição do funcionamento de uma linguagem”.

O campo de representação de um objeto possui variações atribuídas às situações de uso de uma linguagem por parte dos usuários. Segundo Freixa (2002, p. 21 tradução nossa), as variantes possuem características que se atribuem em função das situações de uso da língua, sendo elas “[...] características geográficas, sociais ou históricas dos falantes [...]”.

Com base na Teoria Comunicativa da Terminologia, que incorpora o fenômeno das variantes terminológicas ao seu contexto de estudo, Freixa (2002) evidencia que a terminologia possui uma capacidade descritiva e explicativa para a integração das variações a partir de suas funções no discurso comunicativo e na distinção de uso por parte dos usuários.

Freixa (2002) estabelece três tipos de variantes linguísticas: as variantes terminológicas, as variantes denominativas e as variantes conceituais. Para tanto, a autora afirma que as variantes terminológicas possuem como característica a questão linguística implícita em seu contexto. Assim, a variação terminológica pode ser entendida como a ocorrência de uma variação funcional empregada no discurso especializado, em comparação com o emprego da mesma expressão fora do discurso especializado. Dessa forma, entende-se que o uso da unidade linguística no discurso especializado visando o processo comunicativo serve de premissa para caracterização da variante terminológica.

A variação denominativa é reconhecida pelas diferentes formas possíveis pela quais se pode fazer o reconhecimento de um conceito por mais de uma denominação. Assim se pode, aqui, evidenciar o fenômeno das sinonímias, ou seja, fenômeno pelo qual mais de uma expressão representa o mesmo conceito.

E, como terceiro tipo de variação terminológica compreendida por Freixa (2002, p. 55 tradução nossa), há a variação conceitual, na qual “[...] um mesmo conceito pode ser abordado de diferentes maneiras e, portanto, podem aparecer diferentes denominações”. Este caso de variação ocorre pelo fato da não delimitação do conceito dentro de uma determinada área, pelo contexto de ocorrência e pelas questões semânticas envolvidas na construção do mesmo.

Para fins desta pesquisa, consideraremos apenas a ocorrência da variação denominativa e suas características, visto que nosso objeto de análise comporta apenas as expressões de busca dos usuários do Portal LexML, da área do Direito Penal. Isso impossibilita uma análise conceitual das variantes conceituais, bem como a observação das variantes terminológicas definidas acima, com base em Freixa (2002), tendo em vista nosso recorte de pesquisa.

A variação denominativa é definida por Freixa (2002, p. 54 tradução nossa) como um conceito que possui correspondência com “[...] diversas denominações”. Assim, pela diversidade das línguas em manifestar diferentes fenômenos linguísticos. Suárez e Cabré (2002, p. 1, tradução nossa) entendem por variação denominativa “[...] a presença de formas distintas para referir-se a um mesmo conceito”.

A ocorrência da variação denominativa possui uma relação direta com as situações comunicativas de sua manifestação, ou seja, a variação de um conceito sob diferentes formas está ligada ao texto especializado, ao conjunto de níveis de comunicação. Para Freixa (2002), a aparição de sinônimos ocorre tanto no discurso especializado, como na utilização da informação pelo usuário, que se apropria do conhecimento apresentado e desenvolve uma compreensão cognitiva sobre o mesmo e, posteriormente, o reproduz. Caracteriza-se, assim, uma das marcas da terminologia, que é o emprego de aspectos cognitivos no ordenamento e conceitualização de um conhecimento (CABRÉ, 1993).

As características das variantes denominativas, segundo Cabré (1993), que se apresentam como formas variadas de um mesmo termo representativo de um conceito, possuem uma perspectiva linguística quanto à grafia, fonética, fonologia, estrutura morfosintática e quanto à aplicação de regras linguísticas para o estabelecimento de um termo significante. Assim, para a terminologia que é avalizada pela perspectiva linguística como um conjunto terminológico de uma área especializada e utilizada na comunicação entre especialistas, a grafia das unidades léxicas, que “[...] tem uma importância capital, já que os processos de normalização não atuam sobre a pronúncia dos termos, mas precisamente sobre sua forma escrita” (CABRÉ, 1993, p. 172, tradução nossa).

Dentre os tipos de variações denominativas, destaca-se a variação de pronúncia e grafia, ou seja, o termo condicionado às regras fonéticas de pronúncia e gráficas, de acordo com as convenções ortográficas. Cabré (1993, p. 173, tradução nossa) afirma que “[...] em terminologia, a denominação ou forma de um termo se considera uma sequência de sons ou letras que representam uma forma de base fonológica, de acordo com as regras que regem sua fonética, ou gráfica, de acordo com as convenções ortográficas de cada língua”.

Evidencia-se, aqui, a questão da pronúncia e da grafia que podem manter uma diferenciação entre a forma com que o termo é pronunciado e a sua forma escrita. Do ponto de vista terminológico, Cabré (1993) explica que, por vezes, os termos não possuem características próprias para diferenciá-los das palavras em geral, visto que a grafia está condicionada às regras gerais da língua.

A variação de morfologia está ligada à possibilidade de decomposição do termo em unidades distintivas e significativas, conhecidos como morfemas. Cabré (1993, p. 174, tradução nossa) diz que o “[...] morfema é uma unidade pequena do sistema linguístico que cumpre dupla condição formal e semântica, para além de ser também uma unidade de distinção no interior da palavra”. Dessa forma, o fonema cumpre sua função distintiva, porém não possui significação própria quando sozinho. Este tipo de variação é caracterizado pela ocorrência de uma base léxica de constituição do termo, conhecida também como raiz, onde são acoplados sufixos ou prefixos com a intenção de constituir uma forma derivativa de um termo com base na estrutura morfológica da linguagem natural (CABRÉ, 1993).

Segundo Cabré (1993, p. 176, tradução nossa), os termos podem ser classificados por “[...] distintos critérios que se agrupam em torno de quatro aspectos diferentes: a forma, a função, o significado e a procedência”. A autora aponta os critérios de pertencimento dos termos a cada grupo, de acordo com a forma:

- a) segundo o número de morfemas;
- b) segundo os tipos de morfemas que intervém na formação dos termos, podendo eles serem derivados, formados por agregação à base léxica se sufixo ou prefixo, ou compostos formados por combinações de bases léxicas, podendo, inclusive, comportar sufixos ou prefixos;
- c) segundo a sua complexidade, determinada pela combinação de palavras a uma determinada estrutura sintática, ou seja, essas estruturas são conhecidas como sintagmas terminológicos, que fazem uso da combinação entre palavras para a representação de um assunto e;
- d) segundo sua formação de origem, sendo o caso aqui das siglas, dos acrônimos, das abreviaturas e formas abreviadas.

A partir do ponto de vista da função no discurso especializado, Cabré (1993, p. 180, tradução nossa) afirma que os “[...] termos podem classificar-se em vários grupos ou funcionalidades: substantivos, adjetivos, verbos e advérbios”. Porém, a quantidade de substantivos no léxico especializado é muito maior frente a adjetivos e verbos.

A partir do ponto de vista do significado, os termos podem ser classificados segundo os conceitos que denominam. Cabré (1993, p. 180, tradução nossa) afirma que os conceitos agrupam-se em classes e subclasses em função de suas características e da “[...] interação entre os conceitos derivam as distintas relações, processos e estados, que também são conceitos”.

Por último, a partir do ponto de vista de procedência linguística, Cabré (1993, p. 181, tradução nossa) diz que “[...] os termos podem ter sido criados ou construídos por aplicação de regras do próprio código linguístico ou tomados emprestado de outro código ou subcódigo”. Ou seja, a linguagem pode ter se apropriado de códigos linguísticos provenientes de outras línguas e culturas com o propósito de formação de uma unidade léxica com a capacidade de representar o conceito de um domínio especializado.

Esse pressuposto teórico embasa a questão principal de estudo desta pesquisa, no qual ocorre a análise das expressões de busca dos usuários sobre se as variantes terminológicas podem ser empregadas na indexação ou nas linguagens documentárias, como forma de melhorar a indexação aplicada no Portal LexML, que comporta informação jurídica e legislativa, ou seja, informação especializada que pertence a áreas diferentes do Direito.

5 METODOLOGIA

Na metodologia apresentada a seguir, foi definido o tipo de pesquisa que será desenvolvida quanto aos objetivos, o tipo de abordagem e a natureza da pesquisa. Ainda, buscou-se apresentar o Portal LexML¹, que servirá de base para a análise da pesquisa, a forma de coleta e tratamento dos dados para posterior análise e discussão dos resultados.

5.1 TIPO DE PESQUISA

A pesquisa realizada tem a tipologia básica, visto que se propõe a elucidar a possibilidade de se levar em consideração a linguagem do usuário para a melhora da indexação. Buscou-se conhecer a possibilidade de utilizar *logs* de pesquisa, que contém as expressões de busca dos usuários como aporte para a melhora das etapas da indexação por assunto no Portal LexML, ou seja, para a melhora da consistência da indexação. A pesquisa tem um escopo qualitativo quanto à abordagem utilizada, pois se buscou estabelecer um comparativo entre a linguagem de pesquisa do usuário e a linguagem de indexação utilizada para representar a informação.

O desenvolvimento da pesquisa quanto aos objetivos pré-estabelecidos possui um caráter exploratório, pois foram investigadas, dentro dos *logs* de acessos, as expressões de busca dos usuários como forma de contribuição para a melhora da indexação entendida, nesta pesquisa, como as etapas de análise de assunto e tradução para uma linguagem controlada dos assuntos tratados nos materiais de informação. Gil (2004, p. 41) diz que as pesquisas exploratórias têm “ [...] por objetivo proporcionar maior familiaridade com o problema, com vista a torná-lo mais explícito ou a construir hipóteses”. A aproximação com o problema de pesquisa, no caso exploratório, visa a sua compreensão com a finalidade de aprimorar os conhecimentos sobre o mesmo, possibilitando, portanto, que se seja mais maleável com relação ao planejamento da pesquisa, para que se possam levar em conta os aspectos que influenciam sua ocorrência (GIL, 2004).

¹ Disponível em: <<http://www.lexml.gov.br/>>.

Segundo Gibbs (2009, p. 9), na pesquisa qualitativa “[...] os pesquisadores qualitativos estão interessados em ter acesso a experiências, interações e documentos em seu contexto natural [...]”. Assim, o pressuposto de análise desta pesquisa, a verificação da interação do usuário com o sistema de informação por meio de *logs* de acesso, apoia-se na premissa de que a pesquisa qualitativa examina “[...] interações e comunicações que estejam se desenvolvendo. Isso pode ser baseado na observação e no registro de práticas de interação e comunicação, bem como na análise desse material (GIBBS, 2009, p. 8)”. A investigação a partir de documentos, que são registros em seu contexto natural, oferece a oportunidade de análise dos dados qualitativos em sua ocorrência singular. Gibbs (2009, p. 17) considera como dados qualitativos todos aqueles oriundos de “[...] qualquer forma de comunicação humana [...]” que necessitam de procedimentos analíticos para seu processamento e posterior análise.

O excludente de pesquisa quantitativa ampara-se no que afirmam Rosental e Frémontier-Murphy (2002, p. 17), de que a “[...] característica qualitativa é uma propriedade que não é passível de ser medida numa escala numérica”. A reunião dos dados desta pesquisa não possui características que possam ser medidas de forma quantitativa. Ao se avaliar a correspondência entre a linguagem do usuário e a linguagem documentária, que é um dos objetivos da pesquisa, busca-se encontrar na linguagem do usuário, com a análise de expressões de busca, características que possam colaborar com a melhora da indexação e não quantificar essa correspondência ou não.

Além do mais, na análise de expressões de busca não há como prever determinadas variáveis, visto que o objeto da pesquisa é o resultado da interação do usuário com a interface de busca do sistema. Presume-se que o usuário faça uso de terminologia variante, tanto da linguagem especializada como da natural, visto que isso é uma característica da comunicação, para evitar a repetição ou até mesmo por preferências linguísticas. Portanto, a identificação das variantes terminológicas ocorre durante a análise dos dados e, apenas nesta etapa, será possível reconhecer características que possam auxiliar na melhora da indexação e que sejam, também, possíveis de agrupamento.

O fato de uma variante ocorrer uma vez dentro do conjunto de dados e de outra variante contar com mais aparições não atribui uma característica quantitativa à pesquisa, pois a mera contagem não responderia por si só ao problema de pesquisa. É necessário, além disso, verificar se a variante possui correspondência na linguagem documentária e se é utilizada nos documentos da área em questão, o que é uma garantia de confiabilidade de que aquela variante usada pelo usuário é também utilizada pelo especialista da área.

A análise das expressões de busca contidas em *logs* de acesso inclui-se nesses critérios de definição da pesquisa como qualitativa, visto que os mesmos representam uma forma de comunicação humana, mediada por meio do uso de um sistema informatizado. É da interação do usuário com o Portal LexML que surgem os *logs* de pesquisa, que comportam os dados de interesse, as expressões de busca. Os dados armazenados pelos *logs* são o conjunto de informações que individualizam a pesquisa do usuário no Portal LexML, que oferta documentos indexados por meio de uma linguagem documentária, que nem sempre pode ser a mesma utilizada pelo usuário para realizar as pesquisas. Portanto, a característica qualitativa da pesquisa fica preservada ao passo que se justifica uma análise manual, amparada por instrumentos de consulta, como um Código Penal e a CDDir (Classificação Decimal de Direito), além da validação do especialista e a verificação da utilização da possível variante terminológica na jurisprudência da área do Direito Penal, escolhida para análise como formas de garantia de pertencimento a área em questão e de utilização pelo especialista.

O delineamento metodológico da pesquisa ressalta os procedimentos técnicos que foram utilizados para a coleta e análise dos dados da pesquisa. O objetivo é analisar as expressões de busca informacional dos usuários que utilizam o Portal LexML de informação jurídica e legislativa, a fim de que, a partir da análise e interpretação dos dados possa-se fornecer subsídios para a melhora da indexação das informações disponibilizadas por esse Portal. Assim, optou-se por uma pesquisa documental, visto que Gil (2004, p.45) afirma que a mesma “[...] vale-se de materiais que ainda não receberam tratamento analítico, ou podem ser reelaborados de acordo com os objetos da pesquisa”.

Por si só, as expressões de busca na pesquisa informacional são informações que estão em seu estado bruto de ocorrência e que ainda não receberam nenhum tipo de tratamento analítico, constituindo-se praticamente num relatório representativo do uso do Portal LexML por parte dos usuários, no que se refere às formas de busca utilizadas pelos mesmos. Embora as características da pesquisa documental remetam à questão do suporte da informação em papel, as expressões de busca coletadas se constituem num documento gerado eletronicamente por um sistema informatizado de coleta de dados que não exige contato direto com os sujeitos da pesquisa, evitando uma coleta manual das expressões de busca. O tipo de investigação científica realizada e sua natureza, abordagem e desenvolvimento possuem respaldo em Gil (2004) e Gibbs (2009).

Para garantir a clareza e compreensão da pesquisa, elaboramos a definição de alguns termos empregados na mesma. Por mais que o conceito de determinados termos, aqui utilizados, ainda gerem discussões quanto a sua amplitude de representação, estabelecemos os

seguintes termos e suas definições visando à compreensão plena de seu emprego nessa pesquisa:

- a) **expressão de busca** – texto utilizado pelo usuário do Portal LexML para realizar a sua pesquisa e que foram extraídas dos *logs* de pesquisa;
- b) **descriptor** – termo autorizado e utilizado pelo VCB para representar um assunto;
- c) **logs de pesquisa** – registro eletrônico de dados de pesquisas de usuários realizadas no sistema de buscas do Portal LexML;
- d) **termo não autorizado** – termo reconhecido pelo VCB e que não deve ser utilizado em detrimento de outra forma definida e padronizada.

5.2 PORTAL LexML

O Portal LexML é um sistema de fornecimento de informações de caráter jurídico e legislativo em âmbito do Governo Federal, estados, municípios e Distrito Federal. A proposta do Portal LexML é agregar a informação, fornecendo acesso a diferentes tipos de documentos de vários órgãos do Poder Executivo, Poder Legislativo e Poder Judiciário, em cumprimento a prerrogativa presente na Constituição Federal que estabelece o acesso à informação como direito fundamental. O Portal LexML se constitui na reunião do catálogo de quatorze órgãos do Governo Federal e da Justiça Federal, com as informações de todos os âmbitos de atuação de cada componente, além de fornecer acesso a acervos de legislação de várias Câmaras Municipais de municípios brasileiros e de outros órgãos de nível federal. Dentre os principais órgãos participantes encontram-se o Senado Federal, o Ministério da Justiça e a Câmara dos Deputados.

O Portal LexML disponibiliza informações em formato digital e permite seu acesso a distância, além de disponibilizar *links* internos que permitem acompanhar todas as mudanças sofridas pela legislação ou informação disponibilizadas no decorrer do tempo. O objetivo geral do Portal LexML é se tornar uma referência em sua área de atuação, além de promover a preservação da informação por meio da disseminação digital da mesma. Por princípio, o acesso às informações presentes no seu acervo pode ser realizado indeterminadamente e de forma gratuita, assegurando, assim, o propósito de constituição do mesmo.

Os documentos indexados pelo Portal LexML se enquadram como informação jurídica, definida por Passos como

Toda unidade do conhecimento humano que tem a finalidade de embasar manifestações do pensamento daqueles que lidam com a matéria jurídica, quando

procuram estudar ou regulamentar situações, relações e comportamentos humanos, ou ainda quando interpretam e aplicam dispositivos legais(1994, p. 363).

Assim, Nascimento e Guimarães (2004, p. 33) definem documento jurídico como “[...] o conjunto de espécies documentais geradas pelo e/ou para o Direito [...]”, que “[...] diz respeito às relações jurídicas existentes entre os indivíduos ou destes para com o Estado e vice-versa [...]” com o objetivo de alcançar a “[...] necessária convivência social, validando fatos e atos de natureza jurídica por meio da proteção à integridade dos mesmos, que retratam uma manifestação da vontade”. Enquadram-se como documento jurídico os oriundos de três fontes documentais características do Direito: a legislação, a jurisprudência, e a doutrina.

Entende-se por legislação, segundo Diniz (1998, p. 75) como o “[...] conjunto de leis de um país, de um estado-membro ou município”. Por jurisprudência, entende-se o “[...] conjunto de decisões uniformes de juízes e tribunais sobre uma dada matéria” (DINIZ, 1998, p. 28). Já a doutrina é entendida, ainda conforme Diniz (1998, p. 244) “[...] como estudo de caráter científico que os juristas realizam a respeito do direito”.

O Portal LexML tem por diretriz, não sendo exatamente uma obrigatoriedade, o fato de que as informações por ele disponibilizadas sejam indexadas com base no Vocabulário Controlado Básico do Senado Federal (VCB). Tal medida visa assegurar a padronização na representação da informação indexada pelo Portal, dar uma maior consistência a indexação realizada e também possibilitar que a utilização do VCB seja um instrumento de recuperação da informação por ele indexada.

Portanto, a partir do Portal LexML foram extraídos os *logs* de pesquisa, que comportam as expressões de busca dos usuários, o objeto de análise deste estudo. A especialização do Portal LexML vai ao encontro das demais premissas desta pesquisa, além do fato de o mesmo utilizar uma linguagem de indexação controlada, o VCB. O Portal LexML oferece resultados de pesquisa que indicam a subárea de pertencimento do assunto dentro da área do Direito, o tipo de documento (artigo, livro), a categoria do documento (jurisprudência, doutrina, legislação), os assuntos que foram usados pelo indexador e outros detalhes, como autor, data, idioma e biblioteca de localização/origem. Além do que, é fornecida a quantidade de resultados e a classificação na CDDir – aspectos primordiais para a avaliação da correspondência da linguagem do usuário com a linguagem de indexação. Tem-se, a seguir, um exemplo de resultado de pesquisa com o descritor “delinquente juvenil” do VCB:

Figura 1 – Forma de apresentação de resultado de pesquisa no Portal LexML.

The screenshot shows the LexML search interface. At the top left is the LexML logo. The search bar contains 'delinquente and juvenil' with a search icon. Below the search bar, it indicates 'Resultados: 429 itens' and 'Ordenar por: Relevância' with a dropdown menu and an 'Ok' button. On the right, there are navigation links: 'Cesta de Itens', 'Modificar Pesquisa', 'Página Anterior', 'Página Inicial', and 'Pesquisa Av'. A 'Visualiz' link is also present. Below the search bar, there are filters for 'Categoria do Documento' (Doutina (425), Legislação (4)), 'Localidade' (Brasil (4)), 'Autoridade' (Federal (4)), and 'Doutrina' (Classificação (CDDir)). The main results area shows a single result (1) with the following details: Tipo: Artigo de revista; Autor: Herzog, Jacques Bernard; Título: [Adolescentes e jovens adultos delinquentes](#); Data: 1963; Assuntos: [Delinquente juvenil](#), [\(Delinquente juvenil\)](#), [Menor delinquente](#), [Menor infrator](#), [Delinquente juvenil](#), [\(Delinquente juvenil\)](#), [Menor delinquente](#), [Menor infrator](#), [menos \[x\]](#); Classificação: DIREITO PÚBLICO - 341.:DIREITO PENAL - 341.5.:Criminologia e antropologia, sociologia e tecnologia criminais - 341.59.:[Delinquentes](#) ou criminosos e suas espécies - 341.591.:Menores [delinquentes](#). Infância criminosa. Criminalidade [juvenil](#) - 341.5915. A 'Adicior' link is visible at the top right of the result box.

Fonte: adaptado pelo autor, através do Portal LexML, disponível em: <<http://www.lexml.gov.br/>>. Acesso em: 11 maio 2014.

5.3 VOCABULÁRIO CONTROLADO BÁSICO

A priori, pensou-se na utilização do Tesouro do Senado Federal como instrumento de linguagem documentária que serviria como base para a análise realizada, visto que o mesmo é de reconhecimento público sobre sua utilização pelas bibliotecas ligadas ao Senado Federal e por se configurar numa estrutura que é característica de uma linguagem documentária tradicional, que é o tesouro, conforme exposto no referencial teórico. Porém, o mesmo já não vem sendo utilizado desde 2008 e nem está mais disponível na página de Internet da Biblioteca do Senado Federal.

Atualmente, vem sendo utilizado o Vocabulário Básico Controlado para a indexação da informação documental das Bibliotecas ligadas ao Senado Federal. Este Vocabulário possui descritores nas áreas de Ciências Sociais, com ênfase no Direito. Sua atualização terminológica é desenvolvida, desde 2002, por meio de um Grupo de Estudo da Rede Virtual de Bibliotecas (RVBI) do Congresso Nacional, que se constitui numa rede de bibliotecas cooperantes coordenada pela Biblioteca do Senado Federal, que agrega recursos bibliográficos, materiais e humanos de quatorze bibliotecas da Administração Pública Federal e do Governo do Distrito Federal, dos Poderes Legislativo, Executivo e Judiciário.

O Vocabulário Básico Controlado possui 9.500 descritores de todos os campos do conhecimento científico que refletem as áreas de atuação das bibliotecas cooperantes da Rede Virtual de Bibliotecas. Sua estruturação interna é parecida com a de um tesouro, oferecendo o

descriptor (termo autorizado), os não descritores (termos não autorizados) e os termos relacionados. O exemplo abaixo se refere ao descriptor “delinquente juvenil”.

Figura 2 – Forma de apresentação de resultado de pesquisa no Vocabulário Controlado Básico.

The screenshot shows the interface of the Vocabulário Controlado Básico. At the top, there is a navigation bar with links: Login, Encerrar Sessão, Usuário, Preferências, Bases de Dados, and Pedido de EEB. Below this is a search bar with options: Busca por Lista, Busca por Palavras, Resultados, Buscas Anteriores, and Favoritos. The main content area displays the following information:

Registro Completo

Formato: [OPAC](#) [Ficha](#) [Reduzido](#) [Nomes MARC](#) [Campos MARC](#)

Registro 1 de 1 ◀ Registro Ant.

N. Sistema	000002514
Descriptor	<input checked="" type="radio"/> Delinquente juvenil
Não Use	<input checked="" type="radio"/> Menor delinquente
	<input checked="" type="radio"/> Menor infrator
Termo relacionado	<input checked="" type="radio"/> Delinquência juvenil
	<input checked="" type="radio"/> Liberdade assistida
	<input checked="" type="radio"/> Medida socioeducativa
Ocorrências	++
Base	VCB
Sub-base	341.59 - Criminologia
Classif.	341.5915 z

Formato: [OPAC](#) [Ficha](#) [Reduzido](#) [Nomes MARC](#) [Campos MARC](#)

Fonte: adaptado pelo autor, através do VCB, disponível em: http://biblioteca2.senado.gov.br:8991/F/696SC99YAE1TSBRRFBXY8YAFUMIQVFXG7LJGDYX3RN15M62F4U-07645?func=find-b-0&local_base=sen10. Acesso em: 11 maio 2014.

O acesso ao Vocabulário Básico Controlado é permitido e de uso livre. A pesquisa pode ser realizada dentro de uma base de dados exclusiva para o VCB, disponível na página da Rede Virtual de Bibliotecas, dentro do *menu* “Base de dados”. Pode-se realizar uma busca a partir de um termo, palavra-chave ou expressão livre e encontrar os descritores utilizados na indexação de determinado assunto e os descritores não autorizados (as variantes terminológicas ora já reconhecidas pelo mesmo).

Como o Vocabulário Controlado Básico é o mais recente instrumento utilizado pelo Portal LexML para indexar os materiais de informação por ele disponibilizados, o mesmo foi escolhido como um dos instrumentos de análise da indexação e comparação da linguagem do usuário; linguagem essa representada pelas expressões de busca existentes nos *logs* de pesquisa.

5.4 COLETA DE DADOS

A coleta dos dados, ou seja, a coleta de *logs* de pesquisa dos usuários do Portal LexML se deu de forma automatizada, executada pelos profissionais responsáveis pela tecnologia da informação do Portal. A execução desta coleta por parte dos responsáveis pelo Portal LexML está ligada ao fato de que a coleta de dados faz parte da pesquisa de doutorado da Professora Rita do Carmo Ferreira Laipelt, da Universidade Federal do Rio Grande do Sul, e que também foi usada para esta pesquisa, entretanto, com diferente enfoque. Isso justifica o fato de que a coleta de dados tenha sido realizada anteriormente ao início da pesquisa, porém ressalta-se que a coleta por si só não gera resultados conclusivos e nem o fato de ter sido executado por terceiros influenciaria nos resultados, já que foi realizada de forma automatizada.

A coleta ocorreu por meio da utilização de *software* que registra as formas com que o usuário pesquisou no Portal, compondo assim os *logs* de pesquisa. Os dados compilados trazem informações sobre o tipo de pesquisa executada, a expressão de busca utilizada e os dados que individualizam a pesquisa executada pelo usuário, além de informações adicionais de data, hora, localização geográfica. Com o intuito de viabilizar a pesquisa, foi definido um período de tempo para delimitar a abrangência da coleta de dados, ora já realizada pelos responsáveis técnicos pelo Portal LexML. O período de tempo definido foi do dia seis de julho a cinco de agosto de 2013, totalizando assim um montante de 31 dias de coleta, de modo que o registro dos *logs* de acesso também foi efetuado aos finais de semana. Os *logs* de pesquisa estão alocados em planilhas do programa *Excel*, *software* editor de planilhas produzido pela *Microsoft Office*®, que organiza os dados em linhas e colunas. Dessa forma, o arquivo de *logs* de pesquisa de um dia de buscas é disposto em doze colunas e em linhas que variam em torno de duas mil linhas nos menores *logs* e chegando a mais de 11 mil linhas nos maiores *logs* diários. Cada linha corresponde a uma pesquisa realizada, ou seja, após o usuário digitar o texto na caixa de busca e clicar no comando “encontrar”, oferecido pelo Portal LexML, gera-se um *log* de pesquisa unitário e a soma dos *logs* de pesquisa unitários, ao longo do dia, gera o *log* de pesquisa diário, portanto, representa uma coleta.

A coleta dos dados realizada foi um procedimento que não deixou de observar a privacidade do usuário do Portal LexML, pois não permitiu a atribuição de nenhuma coleta a um usuário em específico. Isso se deve ao fato de que a comunicação gerada no processo de busca é intermediada por computadores, em que apenas os dados de comunicação entre as

máquinas ficaram salvos, não havendo permissão da identificação de nenhum usuário e nem a coleta de informações pessoais.

A importância deste tipo de coleta de dados reside no fato de que o usuário se expressa livremente no momento da pesquisa. Assim, ele deixa marcas, expressões e estratégias de busca que podem identificar características da sua linguagem, fornecendo subsídios para a pesquisa que se realizou. O processo de armazenamento de *logs* (dados) utilizou como parâmetro o endereço de protocolo de Internet (IP), responsável por parte da comunicação entre computadores, que pode ser identificado como uma sequência numérica que identifica e individualiza um computador (dispositivo) empregado como meio para a busca informacional.

Esta forma de coleta de dados contrapõe-se a estratégia anteriormente aplicada para este tipo de pesquisa, que era o acompanhamento presencial do usuário no momento da busca. Esta possibilidade é oriunda dos avanços tecnológicos que, além de fornecerem acesso a inúmeras informações instantaneamente, também possibilitam o acompanhamento das formas com que o usuário executa suas pesquisas, transformando-se assim numa ferramenta de gestão do sistema de informação. A seguir, apresentamos um exemplo de *log* da pesquisa:

Figura 3 – Exemplo de parte de um *log* de pesquisa.

Colunas											
1	2	3	4	5	6	7	8	9	10	11	12
3353	.82.63.73	2013	7	12	13	23	20	Brasil	"menor infrator"	""	"SUCESSO"
3354	.82.63.73	2013	7	12	13	23	20	Brasil	"menor infrator"	""	"SUCESSO"
3355	.82.63.73	2013	7	12	13	23	20	Brasil	"menor infrator"	""	"SUCESSO"
3356	.82.63.73	2013	7	12	13	23	20	Brasil	"menor infrator"	""	"FALHA"
3357	.82.63.73	2013	7	12	13	23	20	Brasil	"menor infrator"	""	"SUCESSO"
3358	.82.63.73	2013	7	12	13	23	20	Brasil	"menor infrator"	""	"SUCESSO"
3359	.82.63.73	2013	7	12	13	23	20	Brasil	"menor infrator"	""	"SUCESSO"
3360	.82.63.73	2013	7	12	13	23	22	Brasil	"menor infrator"	""	"SUCESSO"
3361	.82.63.73	2013	7	12	13	23	22	Brasil	"menor infrator"	""	"SUCESSO"

Fonte: adaptado pelo autor, do *log* de pesquisa do dia sete de julho de 2013.

A composição do *log* de pesquisa é explicada abaixo com o intuito de demonstrar de que forma a interação do usuário com o sistema de informação Portal LexML foi registrada. Cabe ressaltar que os três primeiros dígitos do *log* foram suprimidos por questões legais de manutenção da privacidade do usuário:

- a) **Primeira coluna:** número da linha no *Excel* em que ocorre a expressão de busca dentro dos *logs* de pesquisa, correspondentes a um dia de buscas, após a reunião dos dados por número de IP;
- b) **Segunda coluna:** IP do computador que foi usado para realizar a pesquisa;

- c) **Terceira, quarta, quinta, sexta, sétima, oitava colunas:** dados de data e hora da pesquisa;
- d) **Nona coluna:** país de origem da pesquisa;
- e) **Décima coluna:** expressão de busca utilizada pelo usuário para realizar a pesquisa.
- f) **Décima primeira coluna:** campo de verificação de delimitação da pesquisa básica. As opções oferecidas pelo Portal LexML são: “tudo”, “legislação”, “jurisprudência”, “proposições legislativas” e “doutrina”. No exemplo em questão, o usuário não selecionou nenhuma opção, embora a pesquisa, por definição do sistema, foi realizada com busca em todas as áreas, ou seja, no campo “tudo”; representada apenas pelos sinais gráficos de aspas. Caso o usuário selecionasse uma delimitação para a pesquisa, que não a opção “tudo”, a mesma apareceria nesta coluna entre aspas.
- g) **Décima segunda coluna:** resultado da pesquisa indicado pelas palavras “sucesso” e “falha”. A qual não se reveste de importância para a pesquisa, pois se refere à comunicação entre o servidor (informática) e o computador do usuário.

Abaixo é apresentada a parte do *log* de pesquisa que contém a expressão de busca do usuário que foi utilizada, com a finalidade de responder ao problema de pesquisa e atingir aos objetivos do estudo. A figura 4 representa a décima coluna da composição do *log* de pesquisa. Os demais dados não foram considerados devido à visão que se está atribuindo ao objeto, considerando-se que o intuito não é analisar a estratégia de busca do usuário e, sim, se a expressão de busca (assunto, variante terminológica) utilizada é compreendida ou não pela indexação realizada pelo Portal LexML.

Figura 4 – Exemplo de expressão de busca contida em um *log* de pesquisa.

A imagem mostra a expressão de busca "menor infrator" em uma fonte colorida e estilizada, com cada letra tendo uma cor diferente (verde, amarelo, laranja, vermelho, roxo, azul, verde). A expressão está entre aspas simples.

Fonte: adaptado pelo autor, do *log* de pesquisa do dia sete de julho de 2013, linha 3353.

5.5 CONSTITUIÇÃO DO CORPUS TEXTUAL

A coleta de dados realizada gerou uma quantidade enorme de registros, então, buscou-se estabelecer alguns critérios para restringir o número de dados que foram analisados e, assim, tornar viável a pesquisa, tanto em função do tempo para sua realização, quanto pela questão de que a análise pretendida pode ser feita por meio da utilização de parte destes

dados. A área do Direito escolhida para se tornar a base de orientação para a análise das expressões de busca foi a área do Direito Penal.

A partir desta etapa, todos os procedimentos executados foram realizados de forma manual, sem o emprego de qualquer mecanismo de auxílio. Tal forma é uma exigência imposta pelas características da pesquisa qualitativa realizada em função do objeto e dos objetivos pretendidos. Procedeu-se com a identificação, dentro da expressão de busca utilizada pelo usuário no processo de recuperação da informação, de possíveis termos, ou partes de termos, da área do Direito Penal, que permitam um comparativo entre a linguagem do usuário e a linguagem e indexação utilizada pelo Portal LexML. No âmbito desta pesquisa, utilizamos a locução “expressão de busca” no decorrer da mesma, para identificar o termo, parte de termo ou sinônimo, variante terminológica, sigla, acrônimo, número de lei, artigo de lei que estiver sendo empregado pelo usuário no processo de busca informacional e que estiver presente nos *logs* de pesquisa analisados.

A constituição do *corpus* para a pesquisa se deu a partir dos dados coletados, ou seja, a partir da seleção das expressões de busca que possuíam ligação com a área do Direito Penal utilizadas pelos usuários do Portal LexML. Os dados utilizados na pesquisa compreendem as expressões de busca empregadas pelos usuários do Portal LexML no momento da ação de busca e recuperação da informação. Essas expressões de busca são parte integrante do *log* de pesquisa do usuário junto ao Portal LexML e constituem o registro das ações dos usuários no processo de busca informacional dentro da base de dados mencionada.

Deste modo, a análise das expressões de busca foi amparada por instrumentos especializados que possibilitaram observar a correlação do possível termo com a área de interesse da pesquisa. Um desses instrumentos é a Classificação Decimal de Direito (CDDir)², uma expansão da classe 340 da tradicional Classificação Decimal de Dewey (CDD), desenvolvida em 1948 pela bibliotecária e bacharela em Direito Doris de Queiroz Carvalho para a aplicação na biblioteca do Ministério da Fazenda. A expansão é voltada para a aplicação em obras jurídicas nacionais, pois agrega as especificidades do Direito Pátrio, tanto que é considerada até a atualidade, já estando em sua quarta edição como uma obra de referência para a área e é utilizada, inclusive, pelo Senado Federal como base em suas linguagens documentárias controladas.

Visto a impossibilidade de se colocar a obra em anexo, devido a sua extensão, optou-se por referenciá-la em nota, sendo que a obra está disponível na íntegra e *on-line* no site do Centro de Estudos Jurídicos da Presidência da República².

Os outros instrumentos que foram utilizados são: Código de Processo Penal e o Código Penal Brasileiro, ambos em suas versões eletrônicas disponibilizadas pelo Portal da Legislação do Governo Federal³. Além do mais, a expressão de busca foi submetida à pesquisa no Portal LexML, como forma de garantir que existam documentos em que tal expressão possa ser encontrada e, também, verificar sua inserção dentro da área do Direito definida: a de Direito Penal.

A seleção inicial foi realizada seguindo o roteiro apresentado a seguir. Tal roteiro é voltado para a elucidação de como se procedeu a seleção das expressões de busca dos usuários no Portal LexML. As premissas básicas da primeira etapa, para que a expressões de busca fossem selecionadas inicialmente, são:

- a) o *log* de pesquisa deve pertencer ao período de seis de julho a cinco de agosto de 2013;
- b) ter relação com a área do Direito Penal, ou seja, estar presente na CDDir, ou no Código Penal, ou na jurisprudência penal ,disponibilizada pelo Portal LexML; ou seja, ser utilizada pelos especialistas da área, já que a jurisprudência contém o texto e a linguagem do especialista;
- c) ser utilizada no Portal LexML, ou seja, ao se utilizar a expressão para pesquisa, a mesma deve recuperar documentos dentro do Portal LexML.

A partir dessa etapa prescritiva, a segunda etapa buscou selecionar expressões de busca **pertinentes** (*grifo nosso*) ao Direito Penal, visando reconhecer nelas possíveis variações, como: variantes terminológicas, assuntos que não são cobertos pelo Vocabulário Controlado Básico, termos genéricos ou termos específicos para descrever subáreas do Direito Penal, a legislação e os assuntos ligados à linguagem atual sobre os temas da área. A fim de ilustrar este passo a passo, foram incluídas as figuras 5, 6 e 7 para demonstrar a seleção, na prática, onde se utilizou a expressão de busca “menor infrator” como exemplo, seguindo-se as seguintes etapas:

² CARVALHO, Doris de Queiroz. **Classificação decimal de direito**. 4. ed. rev. e atual. Brasília: Presidência da República, 2002. 174 p. Disponível em:<<http://www4.planalto.gov.br/centrodeestudos/publicacoes/classificacao-decimal-de-direito-1> >. Acesso em: 08 nov. 2013.

³ Disponível em:<<http://www4.planalto.gov.br/centrodeestudos>>. Acesso em: 08 nov. 2013.

- a) validação da expressão de busca na jurisprudência do Portal LexML por meio de pesquisa no próprio Portal; assim, tem-se a garantia de uso por parte do especialista da área, conforme **Figura 5** abaixo;

Figura 5 – Pesquisa com a expressão de busca na jurisprudência do Portal LexML.

The screenshot displays the LexML search interface. At the top, the search query is 'menor and infrator' with a callout box labeled 'Item a' pointing to the search terms. Below the search bar, the results are categorized by 'Jurisprudência em tipoDocumento'. The search results section shows a list of documents, with the first one highlighted. This document is from the 'Tribunal de Justiça do Distrito Federal e dos Territórios. 2ª Turma Criminal' and is titled 'Acórdão nº 412390 do Processo nº20090130005077ape'. The document's subject matter (Assuntos) includes 'MENOR INFRATOR', 'RELEVÂNCIA', and 'RELATÓRIO'. A callout box labeled 'Item a' points to the word 'MENOR INFRATOR' in the subject matter.

Fonte: adaptado pelo autor, através do Portal LexML, disponível em: <<http://www.lexml.gov.br/>>. Acesso em: 12 maio 2014.

- b) verificação do pertencimento da expressão de busca ao Direito Penal, a partir da certeza de que a mesma é utilizada na jurisprudência, tal verificação é conferida por meio da classificação atribuída ao documento pelo Portal LexML e possível de ser visualizada conforme **Figura 6** a seguir;
- c) verificação da utilização da expressão de busca como assunto usada para indexar documentos no Portal LexML e/ou identificação do descritor do VCB que é utilizado para indexar documentos recuperados a partir da expressão de busca. Isso possibilitou a confirmação sobre se a expressão de busca diz respeito ao descritor e pode ser enquadrada como uma variante terminológica ou reconhecer se o VCB trata o assunto de forma mais genérica ou específica, conforme **Figura 6**;

Figura 6 – Pesquisa com a expressão de busca em todo o conjunto de informações (doutrina, legislação) do Portal LexML.



[Cesta de Itens](#) | [Modificar Pesquisa](#) | [Página Anterior](#) | [Página Inicial](#) | [Pesquisa Avançada](#)

Pesquisa: **menor and infrator**

Resultados: 2325 Itens [Visualizar Tudo](#)

Ordenar por: Relevância Página: 1 2 3 4 5 ... [Próximo](#)

Categoria do Documento

- [Doutrina](#) (222)
- [Jurisprudência](#) (2079)
- [Legislação](#) (13)
- [Proposições Legislativas](#) (11)

Localidade

- [Brasil](#) (441)
 - [Distrito Federal](#) (1653)
- [Estados](#) (1)
- [Municípios](#) (8)

1 [Adicionar](#)

Tipo Artigo de revista

Autor Oliveira, Maria Beatriz Bogado Bastos de

Título [O menor infrator](#)

Data 1998

Assuntos Delinquente juvenil, Brasil. {Delinquente juvenil, **menor delinqüente**; **Menor infrator**} Direitos do **menor**. **Menor**, responsabilidade penal. Delinqüente juvenil, Brasil. {Delinqüência juvenil, Delinqüência de **menor**} [menos \[x\]](#)

Classificação DIREITO PÚBLICO - 341::DIREITO PENAL - 341.5::Criminologia e antropologia, sociologia e tecnologia criminais - 341.59::Delinqüentes ou criminosos e suas espécies - 341.591::**Menores** delinqüentes. Infância criminosa. Criminalidade juvenil - 341.5915

Item b (pointing to 'menor delinqüente')

Item c (pointing to 'Menor infrator')

Fonte: adaptado pelo autor, através do Portal LexML, disponível em: <<http://www.lexml.gov.br/>>. Acesso em: 12 maio 2014.

Como resultado destas duas etapas iniciais, chegou-se ao resultado de 61 expressões de busca, presentes no APÊNDICE A, que atendem as premissas apresentadas acima – ou seja, são utilizadas tanto pelos especialistas como pelos usuários –, pertencem ao Direito Penal, foram validadas por especialista da área, enquadram-se como assuntos para indexar documentos ou recuperam documentos no Portal LexML.

A terceira etapa é a verificação, a partir do instrumento de indexação utilizado pelo Portal LexML, o VCB, a fim de descrever os documentos por ele compilados, conforme figura abaixo:

- a) verificação se a expressão de busca apresentada é compreendida ou não pelo Vocabulário Controlado Básico, conforme **Figura 7**;

Figura 7 – Presença da expressão de busca no Vocabulário Controlado Básico.

Formato: [OPAC](#) [Ficha](#) [Reduzido](#) [Nomes MARC](#) [Campos MARC](#)

Registro 1 de 1

N. Sistema	000002514
Descritor	Delinquente juvenil
Não Use	Menor delinquente
	Menor infrator
Termo relacionado	Delinquência juvenil
	Liberdade assistida
	Medida socioeducativa

Fonte: adaptado pelo autor, através do Vocabulário Controlado Básico do Senado Federal.

Para exemplificar essa terceira etapa, utilizou-se uma expressão que está compreendida pelo VCB, a fim de ilustrar de forma clara quando a expressão de busca é um termo variante já reconhecido. Com outras expressões de busca essa etapa pode não ser possível, quando a expressão de busca for uma variante terminológica não reconhecida pelo VCB, contudo, a expressão de busca é, sim, utilizada pelo usuário e também pelo especialista da área, como se pode identificar nas etapas anteriores. Este fato é considerado nos resultados da pesquisa, a fim de analisar cada caso das 34 expressões de buscas analisadas, com a intenção de fornecerem subsídios para a melhora da indexação.

Nos casos em que a expressão de busca não foi identificada no VCB, mas é utilizada pelo usuário e especialista, realizou-se uma aferição nos Códigos de Justiça Penal para identificar como o assunto (expressão de busca) está sendo disposto em lei. Isto permitiu identificar os casos de omissão do VCB em relação a assuntos da área e também possibilitou identificar se o assunto é tratado de forma mais genérica ou específica pelo VCB.

5.6 VALIDAÇÃO DO *CORPUS* DE PESQUISA

A realização das etapas anteriores resultou no APÊNDICE A, que é tomado por base para a validação do *corpus* de pesquisa com o qual se trabalhou. Dentro das 61 expressões de busca utilizadas pelos usuários, procurou-se a identificação de palavras, expressões ou termos que possuíam significado de análise e que pudessem ser utilizados para observação da

linguagem do usuário, em comparação com a linguagem de indexação utilizada pelo Portal LexML.

Como forma de validação de possível variante terminológica, assunto que não esteja sendo coberto pelo Vocabulário Controlado Básico, termo genérico, termo específico e assuntos ligados à linguagem do usuário, à legislação, ao cotidiano ou atualidades, a partir das expressões de busca dos usuários, foram utilizadas três formas aferidoras, a saber:

- a) análise de especialista da área, no caso, profissional com graduação em Direito;
- b) utilização do Código Penal⁴, do Código de Processo Penal⁴, além da CDDir;
- c) utilização da expressão de busca para pesquisa tanto na jurisprudência quanto nas demais categorias (doutrina, legislação), e para pesquisa no campo assunto/indexação do Portal LexML.

O especialista da área tem a capacidade de informar se o possível termo pertence ou não à área de foco da pesquisa. A utilização de Códigos de Justiça Penal e da CDDir tem por objetivo auxiliar na identificação e na localização das expressões de busca dentro da estrutura do Direito Penal. Já a utilização da expressão de busca para pesquisa na base de jurisprudência do Portal LexML torna-se uma eficiente medida de aferição do pertencimento do mesmo à área pré-selecionada, da sua utilização pela comunidade especializada e verificação de seu uso como assunto para indexar documentos.

A jurisprudência consolida-se como uma fonte de informações produzidas por especialistas da área, que se valem de sua condição para produzir documentos legais destinados a dar seguimento às demandas processuais que lhe são de responsabilidade. Para tanto, podem utilizar tanto uma linguagem mais técnica, característica da área, ou a utilização de uma linguagem mais acessível, visto que este documento, classificado como jurisprudência, tem por objetivo informar as partes demandantes sobre o resultado final do processo judicial. O reconhecimento do possível termo na jurisprudência, além de garantir que o mesmo é aceito e utilizado pelos especialistas da área, pode ser uma forma de observar se o mesmo é ou não compreendido pela linguagem documentária, mesmo que seja por meio de remissiva; do contrário, pode ser um modo de servir de argumento para que o mesmo venha a ser incorporado pela linguagem documentária utilizada pelo sistema de informação, já que é uma forma de pesquisa utilizada pelo usuário.

⁴ Disponível em: <<http://www4.planalto.gov.br/legislacao>>. Acesso em: 12 maio 2014.

A verificação na jurisprudência também permitiu realizar uma limpeza nas expressões de busca, ou seja, retirar partes que não influenciariam na sua validação, pois se manteve a forma com que as mesmas ocorrem na jurisprudência pesquisada para dar seguimento a pesquisa. Um exemplo desta limpeza pode ser verificado abaixo:

Quadro 1 – Expressão de busca conforme jurisprudência.

Expressão de busca usada pelo usuário:	Forma de ocorrência na Jurisprudência do LexML:
‘constrangimento ilegal insignificância’	‘constrangimento ilegal’

Fonte: autor.

Como premissas de limpeza e seleção da parte da expressão de busca que seria utilizada na pesquisa, foram definidos dois critérios aplicados a todas as expressões. O primeiro diz respeito à pesquisa na jurisprudência para visualizar a forma com que o especialista da área a utiliza, seguindo a noção quantitativa de ocorrência para a escolha. O segundo critério diz respeito à possível forma de ocorrência da expressão de busca nos Códigos de Justiça Penal.

No exemplo acima, ao se pesquisar a expressão de busca na jurisprudência, foi possível constatar que a mesma continha um qualificador que comportava o foco da necessidade específica do usuário. A expressão de busca ‘constrangimento ilegal insignificância’ é recuperada com maior número de ocorrências sob a forma ‘constrangimento ilegal’. Isto indica que ‘constrangimento ilegal’ é um termo e que ‘insignificância’ é outro termo com valor de qualificador, ou seja, indica que o usuário busca por documentos sobre ‘constrangimento ilegal’, porém que tratem com mais especificidade sobre ‘insignificância’. Ao se verificar a ocorrência da expressão de busca nos Códigos de Justiça Penal que serviram de ferramentas de auxílio à seleção, também se constatou a ocorrência da expressão de busca ‘constrangimento ilegal’ na sua estrutura sem o qualificador ‘insignificância’.

A vantagem de se trabalhar com uma expressão de busca reduzida e que, ao mesmo tempo, demonstra-se presente na linguagem do especialista, facilita o trabalho de comparação da mesma com o descritor utilizado no VCB para representar o assunto que a mesma comporta; podendo, assim, verificar se o usuário emprega, na sua linguagem de pesquisa, expressões de busca que representem sua necessidade informacional mais específica.

Com a validação do *corpus* de pesquisa e a limpeza das expressões de busca, com base nas premissas supracitadas, chegou-se ao número de 34 expressões de buscas dos usuários que são apresentadas no APÊNDICE B. As demais expressões de busca foram retiradas do

corpus validado, pois não preencheram os requisitos estabelecidos, como a validação pelos especialistas da área, pertencimento ao Direito Penal ou utilização na jurisprudência.

5.7 ORGANIZAÇÃO DOS DADOS DO *CORPUS* DE PESQUISA

Por meio da validação das expressões de busca, foi gerado um conjunto de 34 expressões (APÊNDICE B) com as quais se trabalhou, a fim de solucionar o problema de pesquisa e atingir os objetivos propostos pela mesma. A identificação dessas características aconteceu a partir da pesquisa com as expressões de busca no VCB, na jurisprudência do LexML e nos Códigos de Justiça Penal, a partir das seguintes etapas metodológicas aplicadas às expressões:

- a) pesquisa no Portal LexML para identificação de sua ocorrência na linguagem do especialista e a quais assuntos do VCB a expressão de busca estava atrelada;
- b) pesquisa específica na jurisprudência do LexML, para encontrar a ocorrência das mesmas expressões de busca do usuário na linguagem do especialista, de modo a compreender a quais assuntos do VCB a expressão de busca estava atrelada;
- c) pesquisa no VCB para reconhecimento se a expressão de busca seria uma possível variante terminológica;
- d) identificação de determinado assunto (presente nas expressões de busca com o auxílio dos Códigos de Justiça Penal) que não seja compreendido pelo VCB ou que possua representação genérica e não específica.

Essas etapas de análise elucidaram se as expressões de busca possuem correspondência com um descritor do VCB. Em caso negativo, apontaram o problema da não correspondência da linguagem do usuário com a linguagem de indexação. A partir desta comparação chegou-se ao conjunto de informações do APÊNDICE C. O resultado desta comparação das expressões de busca com o VCB possibilitou relacionar as expressões de busca com descritores presentes no próprio VCB, reconhecer falhas no tratamento de determinados temas do Direito Penal e verificar a atualização do VCB. Além de reconhecer, principalmente, formas de tratamento desiguais para áreas temáticas de mesmo nível apresentadas nos códigos penais, ou seja, por vezes o VCB é mais genérico, por vezes, mais específico para áreas temáticas de mesmo nível.

Contudo, Lancaster (1993) recomenda que se utilizem termos específicos ao invés de termos genéricos, incluindo todos os tópicos que sejam de interesse dos usuários, que estejam sendo tratados de forma substancial no documento. No mesmo sentido, Rubi (2009, p. 85) diz

que “[...] a especificidade está relacionada ao nível de abrangência que a biblioteca e a linguagem documentária permitem especificar os conceitos identificados no documento”. Pode-se, assim, entender que, mesmo seguindo a premissa apresentada por Lancaster (1993) de se representar os termos específicos, muitas vezes, podem ocorrer limitações com relação ao nível de especificidade da linguagem documentária utilizada.

Foi possível identificar de antemão características de ocorrência dessas expressões de busca que possibilitaram um entendimento maior das mesmas, após a validação e a comparação no VCB, separando-as por características de ocorrência. Criaram-se então, três categorias com características definidas para encaixar as expressões, levando-se em consideração as características destas expressões de busca, frente à forma com que são tratadas pelo VCB. Como parte das expressões pode ser classificada como variantes terminológicas e parte como expressões que denotam uma necessidade informacional mais específica, optou-se por trabalhá-las em separado, visando apresentar com mais detalhes as suas características e como elas podem influenciar na indexação.

A sustentação das expressões de busca em três categorias apoia-se nas particularidades de ocorrência das mesmas na linguagem dos usuários, ou seja, não necessariamente a expressão de busca é uma variante terminológica reconhecida ou não pelo VCB; por vezes a expressão de busca do usuário comporta uma necessidade informacional que a indexação documental utilizada no Portal LexML não atende. As categorias criadas para encaixar as expressões de busca, conforme suas características são: Legislação, Variantes Denominativas e Especificidade.

A seguir definem-se as características de cada uma das três categorias, para que as mesmas possam comportar as expressões de busca dos usuários e possibilitar um entendimento do motivo pelo qual cada expressão é atribuída a uma categoria. Isso possibilita uma análise centrada não apenas em cada caso individual das 34 expressões de busca, mas, também, nas características que abrangem mais de uma expressão de busca e que possam, por si só, identificar possíveis problemas de correspondência da linguagem do usuário com a linguagem de indexação utilizada. Reforça-se o aspecto de que nem todas as expressões de busca analisadas são variantes terminológicas, algumas são termos mais específicos de áreas temáticas tratadas de forma mais genérica pelo VCB.

- a) legislação: incluem-se aqui todas as expressões de busca que remetam a leis, códigos e artigos de lei. Embora a maiorias das variantes terminológicas deste grupo não possuam descritor correspondente no VCB, elas tratam de assuntos que

já possuem uma estrutura conceitual consolidada na área do Direito Penal, visto que integram leis, ou Códigos de Justiça Penal.

Quadro 2 – Exemplo de expressão de busca da categoria ‘Legislação’.

Expressão de busca usada pelo usuário (vide Apêndice B, item 34):	Forma de ocorrência no VCB (descriptor):
“código penal art. 26”	“Não possui representação”

Fonte: autor.

- b) variantes denominativas: incluem-se aqui as variantes terminológicas identificadas como sinônimos, variações por meio de substituição de partes das variantes por sinonímias ou por meio de adjetivação ou substantivação dessas partes.

Quadro 3 – Exemplo de expressão de busca da categoria ‘Variantes denominativas’.

Expressão de busca usada pelo usuário (vide Apêndice B, item 9):	Forma de ocorrência no VCB (descriptor):
“drogas”	“entorpecente”

Fonte: autor.

- c) especificidade: incluem-se aqui expressões de busca que possuam um tratamento mais genérico ou mais específico por parte do VCB em comparação com áreas temáticas de mesmo nível. Incluem-se também expressões de busca que remetem a uma linguagem mais atual sobre um tema, que nomeiam assuntos antes conhecidos por características de ocorrência. Esta categoria visa apresentar os problemas relacionados à questão da atualização do VCB e ao nível de especificidade do tratamento da informação.

Quadro 4- Exemplo de expressão de busca da categoria ‘Especificidade’.

Expressão de busca usada pelo usuário (vide Apêndice B, item 30):	Forma de ocorrência no VCB (descriptor):
“pirâmide financeira”	“Crime contra a economia popular”

Fonte: autor.

O pertencimento de uma expressão de busca a uma categoria não é requisito excludente para o pertencimento às demais, o que se procurou fazer foi uma separação a partir

de uma lógica em que o agrupamento de algumas características da expressão de busca fosse suficiente para a vinculação a uma categoria. O agrupamento em categorias é realizado em função da possibilidade de se identificar fatores que influenciam diretamente na indexação. A inclusão das 34 expressões de busca nas categorias acima apresentadas pode ser consultada no APÊNDICE D.

5.8 APRESENTAÇÃO DOS RESULTADOS

Os dados estão apresentados em forma textual e tabelados a fim de demonstrar todo o processo de investigação realizado a partir dos mesmos, ressaltando os resultados obtidos. Primeiramente, apresenta-se o *corpus* de pesquisa com as 34 expressões selecionadas (APÊNDICE D) que atenderam aos critérios mencionados na metodologia anteriormente descrita, identificando as categorias de ocorrência. Posteriormente, os dados coletados, ou seja, as expressões de busca foram comparadas e analisadas com os descritores utilizados pelo Portal LexML para a indexação da informação disponibilizada ao usuário, tendo por base o Vocabulário Controlado Básico do Senado Federal. O mesmo foi o instrumento aferidor para verificar se a indexação atende às necessidades dos usuários, se estão sendo utilizados seus descritores, se existe o reconhecimento das variantes terminológicas, se abrange todas as áreas possíveis, com qual especificidade ou generalidade são tratados os assuntos por ele descritos e, inclusive, verificando se os assuntos pesquisados possuem descritores no VCB.

Foram apresentados, também, os resultados da recuperação de documentos no Portal LexML, comparando o descritor e a expressão de busca, nos casos que foram tomados por base para uma análise mais detalhada. Devido ao fato de alguns itens possuírem características de ocorrência comuns a outros dentro da mesma categoria, apenas um será usado para demonstrar a diferença na recuperação de documentos. Isto objetiva demonstrar que o problema que ocorre na indexação relacionada à determinada expressão de busca pode afetar por igual outros assuntos que possuam mesmas características de ocorrência.

As análises em que se pesquisou no Portal LexML, como forma exemplificativa de apresentar os problemas da indexação encontrados, foram comparando a recuperação do descritor com a recuperação da expressão de busca, para demonstrar o impacto na própria recuperação de documentos. Os dados quantitativos da recuperação desses documentos foram auferidos até a data de 31 de maio de 2014.

A seguir, inclui-se o roteiro que foi seguido para se avaliar a validade do uso das expressões de busca dos usuários, como objeto de verificação da consistência da indexação

realizada no Portal LexML. Este roteiro tem por objetivo demonstrar as etapas que foram realizadas para analisar as expressões de busca selecionadas, a fim de responder ao problema de pesquisa e atingir os objetivos propostos. Além disso, buscou-se demonstrar que a análise de expressões de busca dos usuários de um sistema de informação, recolhidas por meio de *logs* de pesquisa, pode ser uma ferramenta de análise da consistência da indexação realizada.

Utilizam-se as expressões ‘pesquisa geral’ e ‘pesquisa campo assunto/indexação’ no decorrer deste roteiro, com as seguintes denominações, respectivamente: pesquisa livre na tela inicial do Portal LexML sem a seleção de qualquer tipo de documento em específico; e pesquisa avançada em que se procurava a expressão de busca ou descritor apenas como assunto indexado para representar um documento. A ‘pesquisa geral’ é uma forma de busca que desempenha o papel de investigar a expressão de busca ou o descritor em qualquer parte do documento. Já a busca ‘campo assunto/indexação’ pesquisa apenas a expressão de busca ou o descritor presente no campo usado pelo indexador do documento para atribuir termos ao documento indexado, ou seja, limita o sistema a apenas um campo de pesquisa que recorrentemente é usado para indexar os documentos a partir de uma linguagem de indexação.

As etapas realizadas foram aplicadas a cada caso específico de ocorrência ou representadas por um item de mesma forma de ocorrência, dentro da mesma categoria, de modo que nem todas as expressões de busca da mesma categoria se encaixaram em todas as etapas propostas, pois ou não possuíam representação no VCB, ou não eram tratadas com especificidade pelo VCB, ou eram variantes terminológicas já reconhecidas. Portanto, a análise da expressão de busca se adéqua a uma ou mais etapas de cada categoria presente no roteiro abaixo, com o intuito de demonstrar na prática as inconsistências da indexação realizada no Portal LexML. Esta avaliação reconheceu aspectos presentes na linguagem do usuário que possibilitaram estabelecer quais as origens da problemática que impede que o descritor recupere a mesma quantidade de documentos que a expressão de busca que faz correspondência ao mesmo de forma clara e objetiva. Seguem as etapas:

- a) verificar se ocorre no VCB algum tipo de descritor relacionado à legislação, especificamente que cite leis ou parte das leis;
- b) verificar se a expressão de busca já está relacionada como variante terminológica de descritor do VCB;
- c) verificar a recuperação de documentos, utilizando-se a expressão de busca que não possui descritor no VCB, tanto na ‘pesquisa geral’ como na pesquisa pelo ‘campo assunto/indexação’ do Portal LexML;

- d) comparar a recuperação de documentos no Portal LexML, tanto por meio da ‘pesquisa geral’, como por meio da pesquisa pelo ‘campo assunto/indexação’, utilizando-se a expressão de busca reconhecida como termo variante e o descritor respectivo;
- e) comparar a recuperação de documentos no Portal LexML por meio da pesquisa no ‘campo assunto/indexação’, utilizando a expressão de busca e o descritor respectivo, se for o caso de ela não ser reconhecida como variante terminológica no VCB;
- f) discutir as inconsistências de indexação encontradas.

Debate-se, também, o uso de remissivas como uma forma de reconhecer que a expressão de busca é uma variante terminológica e pode ser incluída no VCB, além da criação de descritores para representar demandas informacionais dos usuários ainda não previstas e possíveis aperfeiçoamentos do VCB com relação ao nível de especificidade de tratamento de alguns assuntos da área em questão.

Como o objetivo do trabalho é analisar a possibilidade de melhora da indexação através da análise das expressões de busca dos usuários, foi verificado em que medida a indexação realizada está atendendo às necessidades dos usuários, ou seja, se a linguagem do usuário (expressões de busca) está sendo absorvida e compreendida pelo sistema de recuperação de informação do Portal LexML no processo de busca informacional. Por fim, demonstrou-se como a análise das expressões de busca pode contribuir para a melhora na indexação da informação.

5.9 LIMITAÇÃO DA PESQUISA

A limitação da pesquisa diz respeito à questão da linguagem técnica da área e da abrangência da pesquisa com relação ao problema de pesquisa. Embora tenha se utilizado de mecanismos aferidores para reconhecer os vocábulos da área específica do Direito Penal, teve-se dificuldade para interpretar e reconhecer os vocábulos, principalmente quando o usuário utiliza na expressão de busca termos mais específicos de uma área e o VCB os trata de forma genérica. Sobre a questão relacionada à abrangência da pesquisa com relação ao problema de pesquisa, no que diz respeito à consistência da indexação, pode-se apenas avaliar aspectos da indexação de uma parte relacionada ao Direito Penal, não sendo possível tirar dados conclusivos sobre o todo ou afirmar que os mesmos problemas da falta de consistência de indexação desta área também ocorrem e afetam outras áreas.

6 ANÁLISE DOS DADOS E RESULTADOS

A análise dos resultados apresenta o detalhamento sobre de que forma as expressões de busca podem ser objeto de avaliação da consistência da indexação realizada com base em um vocabulário controlado. A utilização de um vocabulário controlado é, sem dúvida, uma premissa para que se faça uma análise de expressões de busca, visto que atua como uma referência para sustentar uma avaliação triangular comparada entre os elementos da linguagem de indexação, da linguagem do usuário e os resultados da recuperação da informação.

O objeto da pesquisa, constituído pelos *logs* de pesquisa dos usuários, comporta indicativos de que é possível analisar outros elementos oriundos das necessidades informacionais dos usuários, que não apenas os relacionados à indexação, dando-se destaque às estratégias de busca adotadas pelo usuário. Por meio delas, podem-se verificar as relações presentes, as formas de pesquisa que possam conduzir a medidas para a melhoria da educação de usuários e a relação do mesmo com o sistema, sempre com a expectativa de qualificar o serviço prestado. Os problemas de consistência da indexação que puderam ser apurados na análise em questão dizem respeito à etapa de tradução, dentro do processo de indexação e, principalmente, a problemas na representação de assuntos no VCB. As expressões de busca analisadas detalhadamente a seguir são o resultado da metodologia empregada e constituíram o APÊNDICE D, que foi transportado em partes para a análise, segmentado à categoria que se está analisando.

Centrando-se na análise das expressões de busca, com o intuito da condução em prol de uma melhoria na consistência da indexação, a mesma demonstrou-se eficaz para a solução deste problema de pesquisa, visto que a linguagem do usuário pode evidenciar demandas informacionais que não estejam sendo compreendidas na linguagem de indexação (VCB) e, assim, prejudicam a recuperação de documentos. Essa ocorrência de falha na comunicação vai ao encontro do que Vieira (1988, p. 43) afirma ser a indexação: “uma técnica de análise de conteúdo que condensa a informação significativa de um documento através da atribuição de termos, criando uma linguagem intermediária entre o usuário e o documento”.

Como em todas as áreas do conhecimento, as mudanças e avanços nos saberes empregam a linguagem oral e a escrita como forma de se comunicar com outrem e estabelecer-se como um avanço na sua área de domínio, não sendo diferente na área do Direito, que, ainda assim, possui mais uma peculiaridade importante que influencia no seu contexto comunicativo: a jurisprudência.

Por vezes, os avanços dessa área ocorrem primeiramente fora de seu instrumento próprio de comunicação – a legislação –, de modo que parte de seus avanços surge na forma jurisprudencial, ou seja, do posicionamento da justiça sobre determinado assunto, como resultado de demandas e imposições inadiáveis da sociedade, que ainda não estão previstas em lei ou regulamento próprio. A análise dos resultados apresentou indícios desse fenômeno, pois demonstra que nas expressões de busca dos usuários ocorre esse tipo de fato (novos assuntos e terminologias) e isso só valida a análise das expressões de busca como uma ferramenta de auxílio na melhora da consistência da indexação.

Sendo assim, se a elaboração de uma linguagem documentária da área levar em consideração apenas a temática já consolidada em lei, pode deixar de fora as demandas informacionais de temas ainda não consolidados em lei, mas que são utilizados pelos usuários e pelos especialistas. Dodebei (2002, p. 56) conceitua linguagens documentárias como “[...] linguagens de comunicação entre a informação documentária e o usuário que dela necessita”. Cada vez mais se ressalta a importância da linguagem documentária em observar o usuário nos seus processos, visto que ele é o destinatário da informação.

Além do mais, a análise das expressões de busca pode conduzir a uma avaliação das etapas da indexação, mais precisamente a etapa de tradução. Verificou-se que a demanda do usuário expressa na busca informacional não é atendida como se espera, pois o Vocabulário Controlado Básico do Senado Federal deixa de atender por igual áreas de mesmo nível. Por exemplo, alguns capítulos do Código Penal têm seus assuntos descritos com maior abrangência, primando pela especificidade, e outros, de mesmo nível, são tratados de forma mais genérica. Também se verificou a desatualização, anão utilização de descritores para representar as leis e seus artigos e a falta de reconhecimento de variantes terminológicas. A inconsistência da indexação é deflagrada pelo uso de termos não autorizados pelo VCB na representação de documentos no campo assunto/indexação. Esse problema afeta a qualidade da indexação e, conseqüentemente, a recuperação da informação; pois tendo em vista que

“[...] a escolha adequada de pontos de acessos que traduzam, de forma clara e precisa, a informação contida nos documentos, identificada no processo de análise documental, fará o diferencial no processo de recuperação da informação (BEZERRA, 2008, p. 3)”.

Completando a análise triangular, no que cabe aos resultados da pesquisa no Portal LexML a partir da utilização comparada de descritores, variantes terminológicas, e expressões de busca, constatou-se que o VCB não é utilizado obrigatoriamente como instrumento de representação informacional controlada. Pode-se verificar que, ao se pesquisar com

expressões de busca, ou até mesmo com o descritor que a representa, encontrou-se diferenças nos resultados no Portal LexML. Também se pode verificar que são usados termos que não estão presentes no VCB para indexar documentos da área do Direito Penal. Porém, para a consistência da indexação é necessária a utilização do vocabulário controlado, pois ele “[...] torna-se o ponto de convergência entre as linguagens utilizadas por autores, indexadores e pesquisadores, premissa fundamental para comunicação de informações dentro de um sistema” (STREHL, p. 331, 1998).

Esses termos usados como assunto, que não os do VCB, são oriundos provavelmente da interpretação do indexador do documento, podendo os mesmos termos serem variantes terminológicas, parte de um descritor ou até mesmo parte de uma variante, que dificultam a recuperação da informação, já que não são provenientes da linguagem de indexação utilizada pelo Portal LexML. Isto é reflexo de uma indexação inconsistente, que não segue a política adotada, qual seja, o uso do VCB como linguagem de indexação.

As expressões de busca da área do Direito Penal selecionadas para a análise, conforme a metodologia, apresentaram características de ocorrência que possibilitaram elencá-las em categorias, sendo que seis expressões de busca pertencem à categoria ‘Legislação’, dezenove expressões de busca pertencem à categoria ‘Variantes terminológicas’ e nove expressões de busca pertencem à categoria ‘Especificidade’. Pode-se, assim, visualizar semelhanças que se enquadram em categorias não apenas por serem variantes terminológicas, muito comuns quando se trabalha com a terminologia de uma área, mas também por conterem necessidades informacionais dos usuários que se caracterizam pela especificidade na construção da expressão de busca e que não são absorvidas pelo VCB. Destacaram-se, neste quesito, as expressões de busca ligadas à legislação, com expressões que buscavam apenas partes dos Códigos de Justiça Penal (exemplo: ‘artigo 20 do Código Penal’) e expressões que continham a representação nominal de leis (exemplo: ‘Lei Maria da Penha’).

Nas expressões de busca que se referem ao nível de profundidade no tratamento de assuntos da área, identificou-se a desigual utilização da premissa da especificidade, sendo que alguns temas são trabalhados com maior especificidade, com a identificação dos seus assuntos, já outros temas são representados por descritores mais genéricos, que não identificam os assuntos que são abrangidos por ele, prejudicando, assim, a recuperação de informações quando o usuário realiza uma pesquisa mais específica. Conforme Lancaster (1993, p. 77) “[...] deixar de utilizar o termo mais específico disponível para representar um assunto [...]” configura uma falha na tradução, umas das etapas da indexação.

Esse tipo de problema se torna mais preocupante ao passo que ambos os temas estão apresentados nos Códigos de Justiça Penal de forma igual, porém o VCB os trata de forma desigual, ou seja, é mais genérico no tratamento de determinados temas do que em outros, nos quais imprime uma especificidade maior na representação do assunto.

Já nas etapas metodológicas de reconhecimento das expressões de busca para a seleção inicial de um conjunto que pudesse ser validado, pode-se identificar que o VCB não atende a todas as necessidades dos usuários, pois peca na identificação dos assuntos, na representação de leis e artigos e no reconhecimento de variantes terminológicas. Esse fato nos levar a crer que, independentemente da área selecionada, poder-se-á encontrar deficiências na indexação, devido ao fato de que o VCB possui limitações em sua estrutura. Por outro lado, esse dado só reforça a necessidade de que deva existir, sim, uma padronização da representação descritiva dos documentos por meio de uma linguagem documentária que atenda aos anseios não só do indexador, mas também do usuário. Nota-se que a análise conceitual somente pode se concretizar com plenitude, ao passo que a tradução, etapa posterior, seja realizada seguindo-se uma política de indexação que represente os objetivos do sistema de informação e, indiscutivelmente, utilize um vocabulário controlado para indexar.

6.1 ANÁLISE DOS RESULTADOS DA CATEGORIA LEGISLAÇÃO

Essa categoria também apresenta problemas causados pela variação terminológica, apenas nela se destacaram os problemas referentes à legislação da área do Direito Penal, especificamente. No quadro 5, apresentam-se as expressões de busca pertencentes à categoria “Legislação” e os respectivos descritores no VCB (quando houver) que estão sendo debatidas e analisadas.

Quadro 5 – Expressões de busca da categoria ‘Legislação’.

ITEM	EXPRESSÃO DE BUSCA	DESCRITOR NO VCB ¹
1	Lei Maria da Penha	Não existe um descritor
8	novo código penal	Não existe um descritor
13	181 código penal	Não existe um descritor
25	lei anticorrupção	Corrupção administrativa
31	art 20 3º código penal	Não existe um descritor
34	Código penal art. 26	Não existe um descritor

Fonte: adaptado do APÊNDICE D.

A ocorrência de expressões de busca que referenciam a legislação em si são reflexos da intenção do usuário em transmitir para o sistema de informação uma necessidade

informacional específica que regula com as características da área do Direito, como é o caso dos temas abrangidos na legislação penal. O usuário tem a sua necessidade informacional centrada em partes de uma lei ou código, porém o VCB não identifica isso, como pode ser verificado a seguir.

No **item 1**, ocorre uma busca informacional por meio do nome pelo qual ficou conhecida a lei que trata de violência doméstica. Muito embora não seja exclusividade desta lei, pois algumas leis ficam conhecidas por um nome diferente do padrão que, via de regra, é representado por um número seguido pela data de sua criação. Ao se realizar a pesquisa no VCB constatou-se que a lei citada na expressão de busca não possui um descritor no catálogo de assuntos. A lei é apresentada no VCB como uma entrada do catálogo de autoridades, conforme a seguinte notação: Lei n. 11.340, de 7 de agosto de 2006, sendo ela uma entrada não permitida e que remete à entrada autorizada: Brasil. [Lei Maria da Penha (2006)].

O fato de a lei em questão ser reconhecida como uma entrada autorizada no catálogo de autoridades para autor, não exclui a necessidade de representá-la enquanto um descritor de assunto, visto que o Código de Catalogação Anglo-Americano (2004), usado para a catalogação, trabalha com diferentes catálogos para identificar uma entrada por autoridade e uma entrada secundária por assunto. Isso demonstra a necessidade de incluir no catálogo de autoridades para assuntos descritores que façam referência a leis, fato este que pode ser identificado nesta pesquisa, pois representa uma demanda informacional específica do usuário.

Caso semelhante ocorre com o **item 25**, visto que o mesmo comporta uma expressão de busca que difere em muito do descritor utilizado pelo VCB para representar o assunto, porém mantém correspondência direta com o mesmo. O usuário informa ao sistema de informações que busca documentos relacionados à expressão “lei anticorrupção”, citando o nome pelo qual a lei ficou conhecida no meio jurídico, porém a expressão não é reconhecida como variante terminológica do descritor, o que, com toda a certeza, influenciará na recuperação de documentos no Portal LexML. Segundo Cabré (1993), as variantes se apresentam como formas variadas de um mesmo termo representativo de um conceito. Este tipo de variante terminológica apresentada é conhecida como variante denominativa, assim entendida por Suárez e Cabré (2002, p. 1, tradução nossa) como “[...] a presença de formas distintas para referir-se a um mesmo conceito”. Ela ocorre em situações comunicativas, ou seja, a variação de um conceito sob diferentes formas está ligada ao texto especializado. Freixa (2002) indica que a aparição de sinônimos tanto do discurso especializado como sua

utilização pelo usuário, que se apropria do conhecimento apresentado, é uma das marcas da terminologia.

Esta ocorrência demonstra que o VCB precisa definir políticas mais objetivas em relação à descrição de leis, fato que se comprova com o **item 25** em comparação ao item apresentado acima. Ambos os temas possuem uma lei que os descreve dentro da área penal, mas apenas a expressão de busca “Lei Maria da Penha” tem uma descrição no VCB em formato relacionado à legislação, não ocorrendo o mesmo com a expressão de busca “lei anticorrupção”, que possui o descritor ‘Corrupção administrativa’ que não comporta o nome pelo qual a lei ficou conhecida e nem a sua numeração como termos não autorizados na rede de remissivas.

Esta distorção no tratamento de assuntos de igual tipologia demonstra uma necessidade de atualização do VCB, com a definição de políticas que evitem este tipo de representação diferenciada para assuntos idênticos. Lancaster (2004) afirma que todo o processo deve ser conduzido por meio de elementos norteadores que possibilitem ao indexador proceder de forma a garantir qualidade ao processo de indexação e que atenda às necessidades dos usuários. Ou seja, é necessária a utilização de políticas que definam a construção, uso e atualização da linguagem de indexação utilizada pelo sistema.

A constatação que se possibilitou fazer, a partir da análise da expressão de busca do usuário, foi que o VCB não apresenta os descritores para leis que não ficaram conhecidas por um nome. Neste quesito, obteve-se uma surpresa negativa quanto à expectativa, pois se presumia que as leis fossem também representadas, porém o VCB não possui descritores que representem leis, ou seja, o indexador do Portal LexML não possui a opção de utilizar um descritor formalizado e padronizado para representar uma lei.

Tomando-se por base a análise da expressão de busca e do descritor acima, que apresentou uma grande diferença na recuperação de documentos – mesmo tendo um descritor formalizado – entende-se que, quando não é dada essa opção, a possibilidade de que ocorram diferentes formas de representação do mesmo assunto aumenta e, provavelmente, o número de documentos recuperados também varie em função da inconsistência na indexação. O importante é que se estabeleçam critérios que serão utilizados para se chegar ao conjunto de termos utilizados na indexação e também na destinação deste documento, ou seja, para que público ele se destina (LANCASTER, 1993).

O **item 8** da categoria legislação, foi a expressão de busca que mais apresentou dúvida sobre a necessidade informacional do usuário, visto que não se possui uma clara identificação da intenção da busca. Porém, como dentro do Direito Penal existe apenas um Código Penal,

acredita-se que o usuário faz menção a uma possível atualização, reforma ou acréscimo de conteúdo ao mesmo, que pode ser identificada pelo adjetivo “novo” presente na expressão de busca. Dentro do VCB, identificamos que o mesmo possui um descritor que representa uma possível atualização, reforma ou acréscimo de conteúdo ao Código Penal por meio do descritor “Reforma Penal”; entretanto, não possui nenhuma variante reconhecida. Portanto, entende-se que VCB deva reconhecer este tipo de expressão de busca como uma variante denominativa do descritor apresentado, pois, ao se realizar uma pesquisa geral, percebe-se que a expressão de busca recupera 10149 documentos, sendo que muitos deles possuem o descritor “Reforma Penal” relacionado no campo assunto, dando a entender, assim, que a expressão trata mesmo de uma forma variante do descritor autorizado e merece o seu reconhecimento, pois é utilizada tanto pelo usuário quanto pelo especialista.

Já nos **itens 13, 31 e 34** ocorre à busca por partes específicas do Código Penal, que pertence ao Direito Penal. Porém, o VCB não possui descritores que representem partes deste Código, assim como não possui um descritor que represente uma lei em específico. Isso indica uma incoerência do VCB no trato da legislação, pois possui, na sua estrutura, descritores que representam os Códigos de Justiça, porém não possui descritores para leis e seus artigos ou artigos destes próprios Códigos.

Tomando-se por base a expressão no **item 34**, “Código penal art. 26”, para confirmar que a necessidade do usuário deve ser atendida, buscou-se demonstrar que a mesma, quando aplicada na busca campo assunto/indexação não recupera nenhum documento; contudo, quando utilizada na pesquisa geral, recupera 1908 documentos. O que se ressalta no resultado da pesquisa geral é que a expressão de busca é encontrada principalmente nos documentos que compõe a jurisprudência disponibilizada pelo Portal LexML. Isso é identificado devido à forma com que o Portal apresenta os resultados da busca, separando-os em categorias (jurisprudência, doutrina, legislação), por autoridade, por localidade, por biblioteca, dentre outras separações. Em tempo, pode-se identificar que a ocorrência da expressão está na ementa dos documentos jurídicos, que é o resumo de assuntos que eles contêm e que são extraídos do próprio texto, ou seja, são oriundos da linguagem do especialista.

Esta recuperação corrobora para a ideia de que o VCB precisaria, sim, criar descritores que representem partes de uma lei ou código. Isso ocorre em função de que essas expressões de busca comportam uma necessidade informacional do usuário e também são utilizadas pelos especialistas da área, os operadores do Direito, para representar assuntos da própria área. Além do mais, percebeu-se, na pesquisa, que são utilizadas diferentes formas para representar a mesma expressão. Tomando, ainda, por base o **item 34**, visualizou-se nos resultados da

busca que o indexador, por não possuir um descritor padronizado, representa o assunto de diferentes formas, o que gera uma recuperação desigual quando se utiliza uma dessas variações, embora sua significação como assunto seja a mesma.

A análise dos demais itens, por comportarem as mesmas características de ocorrência do **item 34**, conduziria para o mesmo resultado, apontando para a necessidade da criação de descritores que representem parte de leis ou códigos, visto que são demanda do usuário e são empregados nos textos especializados.

Os assuntos tratados nesta categoria de legislação possuem correspondência direta com uma representação numérica característica da sua forma de apresentação. Ou seja, um assunto da área pode ser referenciado tanto pela sua forma nominal, como pela forma numérica, oriunda da estrutura da legislação a que pertence, mais especificamente, os artigos da lei ou código em questão. Isto ficou evidenciado na análise das expressões de busca dos usuários conforme tabela acima, onde se pode constatar que os usuários realizam suas pesquisas usando expressões que claramente citam os artigos do Código Penal. Porém, o VCB não possui descritores que referenciam os artigos dos códigos penais e nem de qualquer outro tipo de lei.

6.2 ANÁLISE DOS RESULTADOS DA CATEGORIA VARIANTES DENOMINATIVAS

Abaixo estão alocadas as expressões de busca da categoria “Variantes denominativas” e seus respectivos descritores no VCB (quando houver), após seu reconhecimento na linguagem do usuário. A análise visa convalidar os objetivos da pesquisa e, assim, apontar os problemas na consistência da indexação realizada no Portal LexML.

Quadro 6 – Expressões de busca da categoria ‘Variantes denominativas’.

ITEM	EXPRESSÃO DE BUSCA¹	DESCRITOR NO VCB¹
2	Lei penal	Legislação penal
3	documento falso	Falsificação de documentos
4	falsificação assinatura	Falsificação de documentos
5	crimes eletrônicos	Crime por computador
6	preconceito religioso	Discriminação religiosa
7	decisões penais	Decisão em matéria penal
9	drogas	Entorpecente
10	organização criminosa	Crime organizado
11	teoria finalista	Não existe um descritor
12	prescrição criminal	Prescrição extintiva (direito penal)
14	menor infrator	Delinquente juvenil

16	abuso de autoridade	Abuso de poder
17	trabalho prisional	Trabalho penitenciário
19	cena crime	Local do crime
21	deixar de prestar socorro	Omissão de socorro
23	certidão criminal	Antecedentes
26	invasão de domicílio	Violação de domicílio
29	coisa julgada penal	Coisa julgada (processo penal)
32	motorista embriagado	Embriaguez ao volante

Fonte: adaptado do APÊNDICE D.

A análise das expressões de busca que se encaixaram na categoria ‘Variantes denominativas’, conforme metodologia empregada, apresentou uma grande capacidade de fornecer subsídios para identificação de problemas de indexação, factualmente relacionados à etapa de tradução da indexação, visto que o VCB não atende a uma demanda informacional projetada e que se caracteriza pelo uso de variantes terminológicas. Lancaster (2004, p.18) diz que a tradução “[...] envolve a conversão da análise conceitual de um documento num determinado conjunto de termos de indexação”. Cabe ressaltar que a análise conceitual é a apropriação dos assuntos dos documentos pelo indexador e que, nesta pesquisa, não é passível de análise, em função do foco nas expressões de busca dos usuários. Porém, entende-se que a indexação, segundo Lancaster (2004, p. 19), objetiva “[...] representar a substância da análise conceitual mediante o emprego de termos extraídos de alguma forma de vocabulário controlado”.

Frente ao fato de que as expressões de busca que caracterizam as variantes terminológicas ocorrem na jurisprudência também, Suárez e Cabré (2002, p. 2, tradução nossa) entendem que a comunicação científica manifestada por meio de textos especializados adquire formas distintas para se referir a um mesmo conceito “[...] com a finalidade de evitar a repetição de uma mesma forma no discurso [...]”, pois “[...] uma vez introduzido o conceito [no contexto especializado], é comum referir-se ao mesmo por meio de expressões alternativas [...]”. Este tipo de fenômeno comunicacional oportuniza que as variantes terminológicas venham a compor a terminologia da área e que, quando se alça esta análise para a construção de linguagens de indexação, faz-se necessário que se observe a ocorrência das variantes e que elas sejam incluídas na estrutura da linguagem de indexação.

Cabe ainda ressaltar que há indícios de que a utilização do VCB como instrumento de linguagem de indexação controlada não esteja sendo realizada. Fato este que se caracteriza pelo uso de variantes terminológicas reconhecidas nas expressões de busca como assuntos para representar documentos no Portal LexML, em detrimento ao descritor indicado pelo

VCB. Este tipo de inconsistência de indexação pode ser contornado com base numa política de indexação. Conforme Rubi:

A política de indexação deve ser compreendida como uma decisão administrativa que reflita os objetivos da biblioteca, identificando condutas teóricas e práticas das equipes envolvidas no tratamento da informação da biblioteca para definir um padrão de cultura organizacional coerente com a demanda da comunidade acadêmica interna e externa (2009, p. 83).

Nas condutas as quais a autora cita, que devem estar incluídas na política de indexação, pode-se destacar a questão da obrigatoriedade do uso do VCB como linguagem para a representação dos documentos no Portal LexML. Essa ação faria com que representações dos assuntos presentes no documento fossem traduzidas para a linguagem de indexação de forma padronizada e conseguissem atender as exigências dos usuários, indexando de forma mais precisa e correta possível a informação (CARNEIRO, 1985).

A análise dos **itens 2, 9, 14 e 17** apresentam características de ocorrência que permitem uma análise em conjunto e levam ao mesmo resultado. Apontam para a necessidade de definição clara das políticas adotadas pelo sistema de informação, no que tange a adoção de um vocabulário controlado e de sua utilização por parte dos indexadores. Esses quatro itens acima apontados são variantes terminológicas dos descritores que os acompanham no quadro, sendo eles reconhecidos pelo VCB como termos não autorizados e, portanto, não deveriam ser utilizados para representar documentos como assunto.

Tomando por base o **item 14**, representado pela expressão de busca ‘menor infrator’, que já é reconhecida pelo VCB como uma variante terminológica e, portanto, já está incluída em sua estrutura como um termo não autorizado, percebe-se claramente o não seguimento à política de uso do VCB como vocabulário controlado pelos indexadores do Portal LexML. Ao se realizar uma pesquisa no campo assunto/indexação utilizando a expressão de busca, encontrou-se uma totalidade de 651 documentos recuperados, já na pesquisa no campo assunto/indexação pelo descritor do VCB ‘delinquente juvenil’ recuperou-se 209 documentos. A expressão de busca, que já é uma variante terminológica reconhecida e utilizada na estrutura do VCB como remissiva para o descritor autorizado, recupera mais documentos do que o próprio descritor. Ou seja, o VCB não está sendo usado como vocabulário controlado para representar os assuntos.

A indexação baseada na instituição de política que padronize o seu processo, principalmente com a definição de premissas, faz-se necessária para que o sistema de

informação consiga junto ao seu corpo técnico (indexadores) desenvolver uma indexação consistente. Nota-se que o problema reconhecido por meio das expressões de busca diz respeito a uma das etapas da indexação, a tradução, precisamente quanto à questão do uso do vocabulário controlado como ferramenta padrão para representar a informação. Strehl diz que

[...] os objetivos de uma política de indexação são a definição das variáveis que afetam o desempenho do serviço de indexação, o estabelecimento dos princípios e critérios que servirão de guia na tomada de decisões para otimização do serviço, a racionalização dos processos e a consistência das operações envolvidas (1998, p. 330).

Observa-se que uma política de indexação pode dar consistência às operações de atribuição de termos de indexação, na medida em que padroniza os critérios que envolvem o seu uso como ferramenta de indexação obrigatória.

A análise seguinte diz respeito ao segundo conjunto de expressões de busca do usuário classificadas como variantes terminológicas, que possuem afinidades na sua ocorrência devido ao fato de serem variantes denominativas dos descritores que as acompanham no quadro, porém, não são previstas como termos não autorizados. Incluem-se nessa característica os **itens 5, 6, 7, 10, 16, 19, 26 e 29**. Esse conjunto de análise aponta para um problema muito recorrente nos vocabulários controlados, também se apresentado no VCB, que é a falta de atualização e reconhecimento das diferentes formas com que um assunto é conhecido, tanto no meio especializado como na linguagem do usuário, que, por vezes, é influenciada por esse.

A não presença das variantes terminológicas na estrutura do VCB desenha uma inconsistência na indexação, ao passo que deixa de atender a uma demanda informacional representada por um conjunto de variantes terminológicas que não estão sendo absorvidas pelo mesmo. Além de prejudicar a recuperação informacional, a defasagem de um vocabulário controlado pode torná-lo obsoleto e prejudicar sua confiabilidade no meio em que atende.

Para representar esse conjunto de itens com a mesma característica de ocorrência – variante terminológica sem representação no VCB – selecionou-se a expressão de busca ‘organização criminosa’, representada pelo **item 10**, a fim de que a mesma possa, na prática, demonstrar o quanto o não reconhecimento de uma variante terminológica de um descritor pode influenciar na recuperação de documentos. A pesquisa no campo assunto/indexação utilizando descritor ‘crime organizado’, recuperou 865 documentos, já a expressão de busca e variante terminológica ‘organização criminosa’ recuperou 371 documentos. A pesquisa geral,

utilizando o descritor ‘crime organizado’, recuperou 1398 documentos e a expressão de busca e variante denominativa ‘organização criminosa’ recuperou 3387 documentos.

Destes 3387 documentos recuperados a partir da expressão de busca (variante denominativa) na pesquisa geral, que engloba a ocorrência do assunto em qualquer parte do documento, emergiram 3246 documentos que compõe a categoria ‘jurisprudência’, onde se encontra principalmente o uso da linguagem do especialista mais atualizada, que ocorre com mais frequência, ou seja, a variante terminológica representada pela expressão de busca é também utilizada pelos especialistas da área, denotando, assim, uma necessidade de inclusão da mesma no VCB como variante terminológica, a fim de aprimorar a recuperação de documentos.

Esse tipo de ocorrência também é encontrado quando se analisam os demais itens com a mesma característica (ser variante terminológica e não estar presente no VCB). A análise de expressões de busca colaborou não apenas para identificar as variantes terminológicas, mas também alertou para a possibilidade de se acompanhar as mudanças no uso da terminologia da área. Cabré (1993, p. 142, tradução nossa), ao mencionar a ocorrência das variantes terminológicas, afirma que “[...] dentro de cada linguagem especializada determinada pela temática, podemos distinguir distintos graus de abstração, que conduzem a diferentes níveis ou estilos discursivos”. Esses níveis discursivos são responsáveis pela ocorrência das variações terminológicas que indicam diferentes denominações para um mesmo conceito em função das características comunicacionais. Ao se encontrar os mesmos termos da linguagem do usuário na linguagem do especialista, identifica-se uma garantia de uso, premissa para inclusão de descritores e termos no vocabulário controlado. Dodebei (2002, p. 72) afirma que, na construção e atualização das linguagens documentárias, “[...] é necessário certificar-se de que a terminologia levantada possui duas garantias: literária e do usuário”.

A inconsistência na indexação fica deflagrada neste conjunto de dados analisados, demonstrando que as expressões de busca dos usuários podem ser uma ferramenta preciosa na condução de políticas para indexação. Percebe-se claramente, nesses itens de ocorrência relacionados acima, que se pode identificar na linguagem do usuário formas de busca que podem vir a incrementar o conjunto de termos não autorizados de um descritor, a fim de chegar a dois objetivos: o primeiro diz respeito a uma recuperação de informações mais eficaz e o segundo está relacionado à facilitação do trabalho do indexador, que, no momento em que for descrever tematicamente um documento, ao se deparar com um assunto que possua diversas formas de representação, possa ter a segurança de descrevê-lo a partir de apenas uma

forma padronizada, ao passo que as demais estarão elencadas como formas (termos) não autorizadas.

Na análise das expressões de busca referentes aos **itens 3, 4, 11, 12, 21, 23 e 32**, do Quadro 6, deparou-se com problemas de inconsistência no padrão com que o descritor representativo do assunto é formalizado, sem fornecer clareza ao conceito a que se refere, não representando assuntos da área, falta de equidade na representação dos assuntos ou falta de capacidade do VCB em compreender a linguagem do usuário, na medida em que possui uma estrutura muito rígida, que não absorve expressões de busca que não podem ser classificadas como variantes terminológicas propriamente.

O item 32, ‘motorista embriagado’, retrata muito bem este tipo de ocorrência, onde o descritor no VCB que trata do assunto é ‘Embriaguez ao volante’. A forma com que o usuário rotula a sua necessidade informacional descrita por meio da expressão é uma variação do descritor do VCB e ambas as formas trazem resultados positivos na recuperação de documentos no Portal LexML. A representação de um objeto possui variações atribuídas às situações de uso de uma linguagem por parte dos usuários, onde, segundo Freixa (2002, p. 21 tradução nossa) as variantes possuem características que se atribuem em função das situações de uso da língua, sendo elas “[...] características geográficas, sociais ou históricas dos falantes [...]”. Por isso que se destaca a necessidade do usuário em conhecer e ter acesso ao VCB, pois, segundo Strehl (p. 331, 1998), “[...] o vocabulário controlado torna-se o ponto de convergência entre as linguagens utilizadas por autores, indexadores e pesquisadores, premissa fundamental para comunicação de informações dentro de um sistema”.

A lógica básica por trás dessa perspectiva centrada no usuário é que os sistemas de informação devam ser modelados de acordo com o usuário, com a natureza de suas necessidades de informação e com seus padrões de comportamento na busca e no uso da informação, de modo a maximizar sua própria eficiência (FERREIRA, 2002, p. 6).

Entretanto, estas expressões de busca apresentam uma capacidade representativa do assunto e fazem clara referência ao descritor, decorrente de uma construção linguística que varia em função do uso de sinônimos, ou através de adjetivação ou de substantivação de suas partes, diferenciando-se assim do descritor, no entanto, tratam do mesmo assunto.

O problema apresentado acima decorre da falta de disponibilização do VCB ao usuário na página do Portal LexML, junto ao campo de pesquisa. Na medida em que o usuário desconhece a forma com que determinados assuntos estão sendo representados no VCB, dificulta-se que ocorra uma comunicação eficiente entre ele e o sistema de informação. Identificou-se que o usuário emprega na construção das expressões de busca o conceito de

determinado assunto, ou seja, ele descreve a ocorrência do assunto, rotulando-o por meio de seu conceito. Este tipo de evento ocorre pela falta de conhecimento sobre como o VCB representa o assunto em questão, por isso sugere-se que o sistema de informação, no caso o LexML, faça uma divulgação clara do uso do VCB como seu vocabulário controlado e o disponibilize em sua página de pesquisa, como uma fonte a mais para o usuário consultar, no caso de dúvidas no processo de recuperação informacional. O vocabulário controlado tem papel importante também na recuperação da informação, visto que os termos utilizados para representação do assunto de um documento, oriundos de um vocabulário controlado, podem orientar os usuários na “[...] elaboração das estratégias de busca de informações no sistema” (FUJITA E REDIGOLO, p. 128, 2009).

O objetivo da pesquisa apesar de ser a análise do uso de expressões de busca para a melhora da indexação, não pode deixar de elencar o problema supracitado como uma forma de conduzir a recuperação documental satisfatória, que também é o objetivo de uma indexação consistente. É bem provável que nunca se consiga chegar a um ponto em que se atenda a todas as necessidades informacionais dos usuários por meio da indexação, visto que o conhecimento não para e sempre produz novas formas de representação, assim como o seu uso cotidiano. Contudo, entende-se que se for ofertada ao usuário a possibilidade de pesquisar no VCB, para que possa esclarecer dúvidas sobre qual é o descritor que é utilizado para representar determinados assuntos e/ou conceitos, a probabilidade de que ele tenha sua demanda informacional atendida aumentaria consideravelmente.

As expressões de busca presentes nos **itens 3, 4, 21 e 32** caracterizam-se por uma necessidade informacional especificada pelos usuários por meio de variações linguísticas de parte do próprio descritor. Tanto o **item 3**, ‘documento falso’, como o **item 4**, ‘falsificação assinatura’, são variações de partes do assunto e fazem referência ao mesmo descritor do VCB ‘Falsificação de documentos’. O VCB possui dois descritores ‘Falsificação de documentos’, porém, um possui um qualificador que indica o seu pertencimento a área civil do Direito e, nesse descritor da área civil, encontra-se o **item 3**, sendo reconhecido como uma variante terminológica do descritor; o mesmo não ocorre com o descritor ‘Falsificação de documentos’ da área penal, o que denota uma falta de equidade no tratamento do assunto e na construção do VCB.

O papel do VCB na representação de formas controladas para atender as expectativas da indexação pode contribuir tanto ao indexador, dando-lhe aporte para a representação dos documentos, como para o usuário, esclarecendo qual descritor é usado para determinado assunto. Assim, Fujita afirma que:

Com uso de uma linguagem documentária adotada pelo sistema de recuperação da informação cada termo, então, será procurado e uma vez obtida a confirmação de que é compatível, total ou parcialmente e nesse caso adequado, será considerado um termo autorizado para a representação tanto para a indexação quanto para atendimento às questões de buscas pelo usuário (2013, p.50).

O **item 21**, ‘deixar de prestar socorro’, apresenta uma tentativa do usuário de representar o descritor ‘Omissão de socorro’ por meio do conceito do mesmo. O mesmo ocorre com o **item 32**. Como não se possui um sistema capaz de compreender este tipo de ocorrência, poderia o VCB fazer a definição do conceito do descritor, assim, o usuário teria condições de identificar nele qual é o descritor utilizado para representar o conceito empregado nas pesquisas. A definição dos conceitos pode auxiliar o usuário na busca informacional, conduzindo-o para o descritor usado pela indexação e, de certa forma, esta definição pode contribuir para a melhora da consistência, ao passo que direciona o usuário para o descritor correto, além de facilitar o trabalho do indexador e dar segurança na utilização do descritor mais indicado para representar o assunto. Fujita discorre sobre a questão dizendo que

[...] a representação por conceitos na análise de assunto para identificação e seleção de conceitos é diferente da representação realizada na tradução porque, em primeiro lugar, é realizada durante a análise de conteúdo do documento, ou seja, com o documento na íntegra e em segundo lugar, porque utiliza conceitos com os quais são representados os termos e assim identificada a tematicidade intrínseca do documento (2013, p. 50).

No que tange os **itens 11, 12 e 23**, faz-se a seguinte análise: o **item 11**, ‘teoria finalista’, é uma expressão de busca que não apresenta descritor no VCB, mas recupera documentos, tanto na pesquisa geral (192 documentos), como na pesquisa campo assunto/indexação (9 documentos). Deduz-se da análise da expressão de busca uma falha no VCB, devido à falta de representação para o mesmo já que a expressão ocorre com certa representatividade no Portal LexML frente ao número de documentos recuperados.

Já nos **itens 12 e 23**, encontrou-se um problema de indexação relacionado à falta de correspondência entre a forma padronizada do descritor de um assunto e suas variantes utilizadas no dia a dia da prática do Direito, em face de preferências de uso, ou aceitação pela comunidade especializada. Ambas as expressões não são reconhecidas como variantes terminológicas, o que acarreta problemas na recuperação tanto na pesquisa geral, como na pesquisa no campo assunto/indexação.

Tomando o **item 12** como exemplo para a análise do problema, representado pela expressão de busca ‘prescrição criminal’, que tem no VCB como descritor correspondente ‘Prescrição extintiva (direito penal)’, identifica-se que a expressão de busca é uma variante denominativa e não está inclusa no VCB. Essa expressão de busca e o descritor sofrem com a questão conceitual do assunto que tratam. Apesar de ambas referirem-se à temática da prescrição da possibilidade de se punir por um crime, o usuário desconhece ou tem preferência de uso pela forma mais utilizada na linguagem rotineira. O VCB utiliza o padrão do descritor que tem origem no Código Penal, porém o assunto tem seu uso corriqueiro mais centrado na variante encontrada na expressão de busca. Confirma-se isto através de uma pesquisa no campo assunto/indexação, em que a expressão de busca ‘prescrição criminal’ recupera 664 documentos, enquanto descritor recupera apenas trinta.

Para que a indexação seja consistente, faz-se necessária esta identificação das variantes denominativas ou expressões de busca que referenciem o descritor e as incluam na estrutura do VCB, independentemente das mesmas serem ou não figurantes da estrutura dos Códigos de Justiça Penal. O seu uso na jurisprudência e nas expressões de busca são a garantia de que tanto o usuário como o especialista fazem uso da mesma expressão para representar um assunto que fora padronizado por um descritor no VCB, mas não tem muito uso na rotina da área, sendo que se utiliza a forma variante.

Não obstante, cabe salientar que a atualização do VCB é uma realidade eminente e necessária para que, ao longo do tempo, possa ressaltar as preferências terminológicas de uso de determinados assuntos. Nem sempre a forma estabelecida pelo VCB é a forma mais utilizada e, portanto, deveria ser inclusa na sua estrutura a expressão de busca predileta para uso rotineiro, como uma forma de melhorar a consistência da indexação e que se evite que a variante terminológica ou a expressão de busca seja utilizada como assunto para indexar documentos no Portal LexML.

6.3 ANÁLISE DOS RESULTADOS DA CATEGORIA ESPECIFICIDADE

Esta categoria de ocorrência das expressões de busca dos usuários representa uma característica que remete à questão da aplicação de um tratamento mais específico para a representação dos assuntos no VCB, em oposição a uma representação mais genérica.

Quadro 7 – Expressões de busca da categoria ‘Especificidade’.

ITEM	EXPRESSÃO DE BUSCA	DESCRITOR NO VCB
14	ocultação de cadáver	Cadáver
18	constrangimento ilegal	Crime contra a liberdade individual
20	progressão regime	Regime penitenciário
22	crime de trânsito	Delito de trânsito
24	Anistia política	Anistia
27	bingos	Jogo de azar
28	bullying	Assédio moral
30	pirâmide financeira	Crime contra a economia popular
33	desacato	Desacato à autoridade

Fonte: adaptado do APÊNDICE D.

O usuário emprega, nas suas buscas, detalhamentos de tópicos ou partes de lei que são reconhecidos como assuntos devido ao seu uso na linguagem do especialista. Os **itens 14, 18, 20, 27, 28 e 30** apresentam demandas informacionais dos usuários que, presentes nas expressões de busca, demonstram a intencionalidade do mesmo em procurar assuntos específicos, que pode ocorrer dentro de um grupo maior de assuntos que estão sendo representados pelo VCB sob o mesmo descritor.

Esta forma de apresentação da linguagem documentária, deixando de representar os assuntos mais específicos em preferência ao uso de um descritor mais genérico vai em sentido contrário do que a teoria terminológica entende como adequado. Fujita e Redigolo (2006, p. 126) afirmam que a “[...] necessidade de tratar tematicamente a informação de uma forma mais específica é devido às especializações dos acervos [...]”. Este é o caso específico deste acervo em questão, que trata de informação jurídica e legislativa, visto que o usuário demanda a informação específica e, portanto, deve o sistema de informação fornecer acesso a ela de forma direta, evitando a necessidade de uma maior seleção nos resultados.

Embora possuam ou não indicação no VCB como remissivas ao descritor, identificou-se nas expressões de busca que a necessidade do usuário está ligada às especificidades do assunto e a sua representação por meio de um descritor genérico tão somente polui os resultados da pesquisa informacional, não se adequando à demanda estabelecida, fato indesejável, pois toma tempo do usuário que vai precisar filtrar manualmente os resultados obtidos, o que vai de encontro a uma das leis de Ranganathan, que orienta que se poupe o tempo do leitor.

Tomando por base o **item 18**, ‘constrangimento ilegal’, como exemplo para demonstrar esse tipo de ocorrência, o VCB representa os crimes contra as liberdades individuais no Código Penal por meio do descritor genérico ‘Crime contra a liberdade individual’, havendo o reconhecimento de quatro tipos de crimes, dentre eles, o

constrangimento ilegal. Já que a estrutura do VCB aproveita a terminologia da área que é empregada no Código Penal para instituir os descritores, poderia aprofundar-se e representar não somente o descritor genérico que institui este tipo de crime, mas, também, os assuntos específicos que são elencados dentro dele e representam os crimes específicos comportados pelo tipo criminal em questão.

Esse tipo de representação se sustenta por dois motivos: primeiro motivo entende que, ao se realizar uma pesquisa no Portal LexML no campo assunto/indexação, o descritor recupera apenas 126 documentos frente aos 11288 da expressão de busca, o que indica a preferência do indexador pelo termo mais específico e, ao mesmo tempo, apresenta um problema, já que ele não segue o VCB; o segundo motivo diz respeito ao fato de que a expressão de busca é utilizada pelo especialista. Lancaster (1993, p. 77) destaca que não se deve “[...] deixar de utilizar o termo mais específico disponível para representar um assunto”, deste modo, seria importante que tanto o VCB, quanto o indexador se utilizassem destas premissas teóricas para tornar o processo de indexação mais consistente e confiável.

Os **itens 28 e 30** apresentam assuntos novos, ligados aos avanços da área, onde novas condutas sociais ilegais são tipificadas como crimes e recebem tratamento diferenciado no VCB. Enquanto o **item 28**, ‘bullying’, está reconhecido no VCB como um termo não autorizado, o **item 30**, ‘pirâmide financeira’, que, no seu contexto de ocorrência pode ser atrelado ao descritor ‘Crime contra a economia popular’, não possui o mesmo tratamento. Seria pertinente, para uma indexação consistente, que todas as formas possíveis que um assunto possam ser identificadas e fossem reconhecidas pela linguagem de indexação utilizada, tanto com o intuito de assegurar que se utilize um descritor padrão para o assunto, quanto para que o usuário consiga recuperar todos os documentos ligados ao assunto.

A possibilidade de análise da consistência da indexação por meio das expressões de busca é alcançada também nessa categoria, visto que identifica um problema na representação dos assuntos da área. Na medida em que o VCB faz uso de termos genéricos e não utiliza os termos mais específicos, tomando por base a estrutura de ocorrência destes assuntos nos Códigos de Justiça Penal, significa que ele deixa de atender uma demanda confirmada nas expressões de busca, que primam pelo específico e não pelo genérico. Carneiro (1985, p. 229) afirma que é necessária a identificação dos assuntos cobertos pelo acervo, “[...] com a identificação de áreas que necessitem um tratamento mais específico ou superficial da informação e a identificação dos assuntos centrais e periféricos das áreas de interesse”.

Já os **itens 22, 24 e 33** apresentam uma ocorrência específica que merece um cuidado particular devido as suas características. O **item 22**, ‘crime de trânsito’, não pode ser uma

variante terminológica do descritor ‘delito de trânsito’, pois ocorrem em áreas distintas do Direito. Apesar da semelhança aparentar que seja uma sinonímia, a expressão de busca ‘crime de trânsito’ é relacionada ao Direito de Trânsito e o descritor ‘delito de trânsito’, ao Direito Penal. Faz-se necessária, aqui, tanto a definição do conceito do descritor como o uso de um qualificador para identificar o pertencimento de cada qual a sua área, facilitando, assim, o trabalho do indexador e promovendo ao usuário a possibilidade de melhorar sua pesquisa.

O **item 24** poderia ser parte de uma decisão política do VCB, visto que, na linguagem do usuário, encontrou-se tanto a expressão de busca ‘anistia política’, como ‘anistia militar’, embora ambas sejam descendentes do assunto em questão: anistia. Caso o VCB e o Portal LexML primassem por atender às especificidades das pesquisas dos usuários, tanto quanto o assunto permitisse, seria uma excelente forma de reduzir o número de documentos que seriam irrelevantes na recuperação, quando se usa o descritor.

O **item 33** segue o passo das diferenças de tratamento dos assuntos dados pelo VCB. O Código Penal trata o assunto por meio do vocábulo ‘Desacato’, já o VCB o representa como ‘Desacato à autoridade’. Realizando-se uma pesquisa no Portal LexML, por meio do campo assunto/indexação, recuperam-se apenas noventa documentos utilizando-se o descritor. Já com o emprego da expressão de busca, que não é reconhecida como variante terminológica, recuperam-se 778 documentos. Assim, a consistência da indexação exaure-se devido a não utilização do VCB pelo indexador do Portal LexML; contudo, o VCB foge ao padrão de representação que vem aplicando, de usar o descritor mais genérico, e utiliza um mais específico. Ainda assim, não atende à demanda, visto que, nos resultados de busca, pode-se identificar outros tipos de ocorrência de desacato, como na área militar e do funcionalismo público. Esse fato demonstra uma inconsistência: ou representa o assunto pelo descritor mais genérico ou representa, também, as demais ocorrências do assunto com características específicas.

6.4 IDENTIFICAÇÃO DAS INCONSISTÊNCIAS NO PROCESSO DE INDEXAÇÃO

Com o objetivo de visualizar os resultados obtidos em separado do seu fato gerador, sintetizamos os mesmos abaixo, para que seja possível destacar os problemas de indexação encontrados e, assim, assegurar que a análise de expressões de busca forneçam subsídios para a melhora da consistência da indexação. Os problemas encontrados foram:

- a) falta de descritores no VCB para representar assuntos ligados à legislação, tanto para leis como para partes de leis;
- b) indicação de uso de descritor mais genérico em detrimento da falta de descritor mais específico;
- c) não utilização do VCB como instrumento padrão para a indicação de descritores;
- d) falta de equidade na representação de assuntos no VCB de áreas temáticas idênticas presentes nas estruturas do Códigos de Justiça Penal;
- e) não utilização de qualificador para diferenciar assuntos com ocorrência em mais de uma área temática;
- f) não reconhecimento de variantes terminológicas de descritores;
- g) falta da definição dos conceitos junto aos descritores;
- h) desatualização do VCB;
- i) não utilização de remissivas para absorver o uso de formas ou nomes populares, pelos quais determinados assuntos ficaram conhecidos.

7 CONCLUSÃO

A análise de expressões de busca proporcionou a identificação de problemas na consistência da indexação, ligados diretamente à segunda etapa do processo de indexação, que diz respeito à adoção de uma linguagem controlada para descrição dos assuntos. Conforme Lancaster (1993, p. 8), “[...] a indexação de assunto possui duas etapas que são a análise conceitual e a tradução”. Assim, o uso de uma linguagem controlada é premissa para o uso de expressões de busca presentes em *logs* de pesquisa como forma de se avaliar a consistência da indexação, que ora serve como base de análise e de identificação dos problemas, pois é construída seguindo critérios ou políticas.

Os *logs* de pesquisa, por meio das expressões de busca e orientando-se pelas etapas metodológicas apresentadas, foram suficientes para a identificação de problemas de consistência da indexação. Esses problemas podem ser relacionados à falta de descritores para representar leis e partes das mesmas, diferenças na representação temática de assuntos de mesmo nível de ocorrência, inobservância de variantes terminológicas de descritores do VCB e a não utilização plena do VCB como ferramenta de descrição temática da informação no Portal LexML, pois a expressão de busca é encontrada como assunto descrevendo documentos que deveriam ser representados por descritores já estabelecidos.

A participação do usuário no processo de indexação, mesmo que indiretamente, quando se utiliza as expressões de busca para análise, pode contribuir para tornar o usuário um agente ativo do sistema de informação. Ao passo que suas demandas informacionais são reconhecidas ou possibilitam discutir formas para a melhora da indexação, eleva-se o usuário à categoria de protagonista dentro do sistema e reconhece-se o valor da integração e da junção de esforços como uma forma de ampliar o acesso à informação. A análise de expressões de busca por meio de *logs* de pesquisa demonstrou ter a capacidade para firmar-se como uma ferramenta de auxílio à melhora dos serviços de informação, uma vez que pode levar a uma aproximação com o usuário da informação, como uma forma de acompanhar as possíveis demandas de informação expressas em sua linguagem, incluindo, aqui, as mudanças que podem vir a ocorrer na atualização e avanço da terminologia utilizada nesta área do conhecimento.

Embora os problemas identificados na área do Direito Penal, pode-se deduzir que os mesmos ocorrerem também com as demais áreas, pelo fato de que esses problemas se encontram na estruturação do VCB, nos critérios de especificidade ou generalidade utilizados para representar os assuntos, na falta de descritores para leis e artigos, no não reconhecimento

de variantes terminológicas, na falta de definição de conceitos como forma de identificar quais as características do descritor e o que ele representa, e por fim, na questão da atualização do mesmo.

A baixa consistência da indexação para representar as expressões de busca da categoria ‘Legislação’ se deflagra com a falta de descritores no VCB para representar assuntos ligados tanto às leis como a partes de leis. O que se ressalta nesta categoria é o fato de que tanto os artigos dos códigos como as próprias leis são atreladas a assuntos que estão representados na estrutura da legislação em que ocorrem. Por exemplo, a expressão de busca representada pelo **item 34**, ‘Código penal art. 26’, está dentro da área temática do Código Penal, que trata da imputabilidade penal e, mais especificamente, define que são os agentes quem não serão punidos pela lei em função de crimes cometidos por falta de capacidade de discernimento de seus atos, em função de condição de mazela, desenvolvimento mental incompleto ou, ao tempo da ação, incapacidade de reconhecer a ilicitude do ato. Assim, a expressão de busca é uma variante terminológica do assunto que ela representa e que possui um conceito formalizado por meio de sua descrição no Código Penal. Embora não possua uma rotulagem que a represente por meio de um descritor (termo), ela é utilizada tanto pelo usuário como pelo especialista.

Esse tipo de ocorrência demonstra a necessidade do VCB incluir, em sua estrutura, descritores que representem leis, artigos de leis e códigos, em sua forma numérica, antecedida pelo nome do tipo de documento jurídico. Recordando, Bocatto diz que as linguagens documentárias

“[...]são linguagens estruturadas e controladas, construídas a partir de princípios e de significados advindos de termos constituintes da linguagem de especialidade e da linguagem natural (linguagem do discurso comum), com a proposta de representar para recuperar a informação documentária (2009, p. 119)”.

Assim, nada mais justo que se inclua na estrutura do VCB, podendo ser como termo não autorizado, as variantes terminológicas que advém da própria estrutura que embasa a construção desta linguagem controlada: a legislação. Esse tipo de acréscimo traria resultados positivos à consistência da indexação, que se estaria reconhecendo todas as possíveis formas com que um assunto pode ser conhecido.

No que concerne à categoria ‘Variantes denominativas’, tem-se como principal problema o não reconhecimento das variantes terminológicas dos descritores. Cabe destacar que esse tipo de problema afeta diretamente a consistência da indexação realizada, ao passo que ela possui a garantia do especialista, pois é utilizada pelo mesmo e pelo usuário, contudo

não é absorvida pela indexação, ou seja, o VCB não consegue traduzir todas as diferentes formas com que o assunto se manifesta. Destaca-se também, dentro desta categoria, o fato de que o VCB não é o utilizado como vocabulário padrão para a etapa da tradução dentro da indexação, visto que se encontraram variantes terminológicas reconhecidas pelo VCB, indexando documentos e recuperando mais documentos que o próprio descritor. O **item 14** do APÊNDICE D, representado pela expressão de busca ‘menor infrator’ ilustra claramente isso.

Dentre os demais problemas de consistência de indexação evidenciados por meio da análise de expressões de busca estão o uso de descritor mais genérico em detrimento da falta de descritor mais específico, que destaca, assim, uma falha de tradução e a falta da definição dos conceitos junto aos descritores. Esse último problema afeta tanto ao usuário, que desconhece qual a representatividade do descritor, e, principalmente, ao indexador, na medida em que desconhece qual seria o melhor descritor que representaria os assuntos oriundos da análise conceitual. Kostina (2000, p. 19, tradução nossa) diz que [...] “o termo é uma unidade léxica específica da linguagem especializada que denomina os conceitos especializados” [...]. Portanto, se o termo denomina um conceito, nada mais coerente que este seja incluindo na estrutura do VCB, dentro do registro do descritor. Isso facilitará a atuação do indexador, que terá uma garantia de estar traduzindo o conceito mais adequado da análise conceitual para o VCB e, assim, para o documento.

Já na categoria ‘Especificidade’, destacou-se o uso de descritor mais genérico em detrimento da falta de descritor mais específico no VCB, que também ocorre na categoria ‘Variantes denominativas’. Fujita e Redigolo (2006, p. 126) afirmam que a “[...] necessidade de tratar tematicamente a informação de uma forma mais específica é devido às especializações dos acervos [...]”. O Portal LexML já apresenta uma especialização em seu acervo. Considerando a área estudada, a do Direito Penal, foi identificado que não se está tratando de forma específica a informação, embora o usuário manifeste sua necessidade informacional de forma específica.

Além dos problemas acima mencionados, emergiram as seguintes falhas de indexação: a desatualização do VCB, a não utilização de qualificador para diferenciar assuntos com ocorrência em mais de uma área temática, a falta de equidade na representação de assuntos no VCB de áreas temáticas idênticas presentes nas estruturas dos Códigos de Justiça Penal e a não utilização de remissivas para absorver o uso de formas ou nomes populares pelos quais determinados assuntos ficaram conhecidos.

A análise das expressões de busca dos usuários deflagrou inconsistências de indexação que basicamente centram-se em dois eixos de ocorrência. O primeiro é representado pelo

VCB e os problemas de falta de descritores, representação dos assuntos por meio de descritores genéricos, falta de reconhecimento das variantes terminológicas e falta de equidade na representação de assuntos de áreas temáticas idênticas presentes nas estruturas dos Códigos de Justiça Penal. O segundo eixo de ocorrência está relacionado à política de indexação do Portal LexML, que é composto por catorze bibliotecas participantes, as quais possuem diferentes bibliotecários realizando as atividades de tratamento e organização da informação. Essa condição, de múltiplos indivíduos inserindo documentos no sistema de informação, indica a necessidade de controle das atividades envolvidas, que pode ser realizado por meio de uma política de indexação. Esse controle visa conduzir a utilização do VCB para a tradução dos assuntos dos documentos e evitar que termos não autorizados sejam utilizados na indexação.

Além do mais, a política de indexação atua também na identificação de áreas que necessitem um tratamento mais específico ou mais superficial da informação, no nível de exaustividade, que diz respeito à quantidade de termos que serão utilizados na representação de um documento e na especificidade do tratamento da informação, ou seja, na extensão com que a informação será descrita, influenciando diretamente na constituição do vocabulário controlado e, principalmente, sobre a avaliação do sistema, satisfação do usuário, atendimento de demandas (CARNEIRO, 1985).

Identificou-se que o Portal LexML necessita desenvolver ou aprimorar a política de indexação para que não ocorra problemas como a não utilização do VCB para a indexação de documentos. A política de indexação deve refletir os objetivos do sistema de informação e deve servir de guia para tomada de decisão no processo de indexação (RUBI, 2012). Ela também deve definir critérios sobre o nível de especificidade que irá tratar os assuntos e, principalmente, sobre a atualização do VCB, prever a inclusão da definição dos conceitos dos descritores, refletir os princípios para atualização, constituição de rede de remissivas, inclusão de variantes terminológicas e a equidade na representação de assuntos de áreas afins. Sugere-se que se dê preferência à especificidade na representação dos assuntos. Isso se justifica frente à ocorrência na linguagem do usuário de expressões de busca direcionadas a temáticas específicas do Direito Penal.

Confirma-se que a análise de expressões de busca pode colaborar com a melhora das etapas de indexação, trazendo consistência à indexação realizada no Portal LexML. Ressalta-se que o usuário poderia contribuir no processo de representação temática e na constituição de um sistema de informação eficiente.

A lógica básica por trás dessa perspectiva centrada no usuário é que os sistemas de informação devam ser modelados de acordo com o usuário, com a natureza de suas necessidades de informação e com seus padrões de comportamento na busca e no uso da informação, de modo a maximizar sua própria eficiência (FERREIRA, 2002, p. 6).

Por fim, como a consistência da indexação passa obrigatoriamente pela utilização de um vocabulário controlado e o Portal LexML faz uso do VCB, sugere-se a realização de estudos específicos para a descoberta de possíveis falhas na sua estrutura, para que, assim, possa atender às expectativas dos usuários nas suas buscas informacionais, que caminham em direção à especificidade na indicação de suas necessidades informacionais e que podem vir a conduzir a melhorias em todo o sistema.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVORCEM, Rochelle Martins; VAN DER LAAN, Regina Helena. A terminologia expressa da área da ciência da informação no discurso do especialista um estudo de caso. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v.3, n.2, p.107-115, jul-dez. 2007.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

ARAÚJO JÚNIOR, Rogério Henrique de. **Precisão no processo de recuperação da informação**. Brasília: Thesaurus, 2007.

BALBY, Claudia Negrão. **Estudos de uso de catálogos on-line (OPACs)**: revisão metodológica e aplicação da técnica de análise de log de transações a um OPAC de biblioteca universitária brasileira. 2002. 137 f. Tese (Doutorado em Ciências da Comunicação) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2002.

BARITÉ, M. La garantía literaria: vigencia y proyección teórico-metodológica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, VIII, 2007, Salvador. **Anais...** Salvador: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--068.pdf>>. Acesso em: 20 out. 2013.

BARROS, Lidia de Almeida. **Curso básico de terminologia**. São Paulo: Editora da Universidade de São Paulo, 2004.

BARROS, Lidia de Almeida. **Aspectos epistemológicos e perspectivas científicas da Terminologia**. *Ciência & Cultura*, vol. 58, n.2, São Paulo, abr./jun. 2006.

BEZERRA, F. M. P. A representação temática nos sistemas de informação e o reflexo na qualidade de comunicação com os usuários. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, n.15, 2008, São Paulo. **Anais...** São Paulo: CRUESP, 2008. Disponível em: <<http://www.repositoriobib.ufc.br/000003/000003EA.pdf>>. Acesso em: 30 set. 2013.

BOCCATO, Vera Regina Casari. A linguagem documentária vista pelo conteúdo, forma e uso na perspectiva de catalogadores e usuários. In: FUJITA, M. S. L., org., et al. **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009. p. 119-135. Disponível em: <<http://books.scielo.org/id/wcvbc/pdf/boccatto-9788579830150-06.pdf>>. Acesso em: 19 out. 2013.

BOCCATO, Vera Regina Casari; FUJITA, Mariângela Spotti Lopes. O uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal. **Perspect. ciênc. inf.**, 2010, vol.15, no.3, p.23-51.

Disponível em: <<http://dx.doi.org/10.1590/S1413-99362010000300003>>. Acesso em: 17 out. 2013.

BORBA, Diego dos Santos; VAN DER LAAN, Regina Helena; Chini, Bernadete Ros. Palavras-chave: convergências e diferenciações entre a linguagem natural e a terminologia. **Perspect. ciênc. inf.** vol.17 n. 2, Belo Horizonte abr./jun. 2012. Disponível em: <<http://dx.doi.org/10.1590/S1413-99362012000200003>>. Acesso em: 17 out. 2013.

CABRÉ, Maria Teresa . **La terminología**: teoría, metodología, aplicaciones. Barcelona: Editorial Empúries, 1993.

CABRÉ. M. Teresa. **La terminología**: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos. Barcelona: Institut Universitari de Lingüística Aplicada, 1999.

CABRÉ. M. Teresa. La terminología entre la lexicología y la documentación: aspectos históricos e importância social. In: AGUADO, G.; DURÁN, P. **La investigación em lenguas aplicadas: enfoque multidisciplinar**. Madrid: Fundación Gómez-Pardo/Universidad Politécnica de Madrid. 2001a. p. 65-78.

CABRÉ, M.T. “Sumario de principios que configuran la nueva propuesta teórica y consecuencias metodológicas”. In: CABRÉ, M. T.; FELIU, J. (edt.). **La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica**. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, p. 17-25, 2001b.

CABRÉ, Maria Teresa. La Terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro. **Revista Debate Terminológico**, Paris, n. 1, abr. 2005. Disponível em: <http://seer.ufrgs.br/riterm/article/view/21286/12263>. Acesso em: 18 out. 2013.

CABRÉ. M. Teresa. El principio de poliedricidad: La articulación de lo discurso, lo cognitivo y lo lingüístico em terminología. **Ibérica**, v.1, n.16. Castelló: AELFE, 2008.

CAMPOS, Maria Luiza de Almeida. **Linguagem documentária**: teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001.

CARNEIRO, Marília Vidigal. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v.14, n.2, p.221-241, set. 1985.

CARVALHO, Doris de Queiroz. **Classificação decimal de direito**. 4. ed. rev. e atual. Brasília: Presidência da República, 2002. 174 p. Disponível em:<<http://www4.planalto.gov.br/centrodeestudos/publicacoes/classificacao-decimal-de-direito-1> >. Acesso em: 08 nov. 2013.

CASARI, Vera Regina Boccato; FUJITA, Mariângela Spotti. Avaliação da linguagem documentária decs na área de fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação com protocolo verba. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n.21, 1º sem. 2006.

CIANCONI, Regina de Barros. Sistemas de recuperação em linha: educação x atuação profissional. **Ciência da Informação**, v.19, n.1, p.131-136, jul./dez. 1990.

CINTRA, A.M.M. Elementos de linguística para estudos de indexação. **Ciência da Informação**, v.12, n.1, p.5-22, jan./abr. 1983. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1526/1144>>. Acesso em: 16 out. 2013.

CINTRA, A.M.M. et al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002.

CÓDIGO de Catalogação Anglo-Americano (CCAA2). 2. ed. rev. São Paulo: FEBAB, 2004.

COSTA, Luciana Ferreira da; SILVA, Alan Curcino Pedreira da; RAMALHO, Francisca Arruda. (Re)visitando os estudos de usuário: entre a “tradição” e o “alternativo”. **DataGramaZero** - Revista de Ciência da Informação - v.10 n.4 ago/09. Disponível em: <http://www.dgz.org.br/ago09/Art_03.htm#R1>. Acesso em: 19 out. 2013.

CURRÁS, Emilia. **Tesauros**: linguagens terminológicas. Tradução de Antônio Felipe Corrêa da Costa. Brasília: IBICT, 1995.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de Assunto**: teoria e prática. Brasília: Thesaurus, 2007.

DINIZ, Maria Helena. Dicionário Jurídico. São Paulo: Saraiva, 1998. v. 3.

DODEBEI, Vera Lúcia Doyle. **Tesauro**: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

DUMONT, Lígia Maria Moreira. Os sentidos da leitura e a subjetividade. In: NAVES, Madalena Martins Lopes; KURAMOTO, Hélio (Org.). **Organização da informação**: princípios e tendências. Brasília: Briquet de Lemos, 2006.

FERREIRA, Sueli Mara Soares Pinto. Novos paradigmas e novos usuários de informação. **Ciência da informação**. Brasília, v. 25, n. 2, 1996.

FERREIRA, Sueli Mara S. P. **Estudos de necessidades de informação: dos paradigmas tradicionais à abordagem Sense-Making**. Disponível em: <www.eca.usp.br/nucleos/sense/index.htm>. Acesso em: 16 out. 2013.

FIGUEIREDO, Nice Menezes de. **Estudos de Uso e Usuários da Informação**. Brasília: IBICT, 1994.

FREIXA AYMERICH, Judit. **La variació terminològica. anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient**. Tese (Doutorado) - Programa de doctorat: Variació en el Llenguatge 1991-93, Universitat de Barcelona, Barcelona, 2002.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para a indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em:

<<http://www.sbu.unicamp.br/seer/ojs/index.php/rbci/article/view/287/167>>. Acesso em: 23 set. 2013.

FUJITA, Mariângela Spotti Lopes. A representação documentária no processo de indexação com o modelo de leitura documentária para textos científicos e livros: uma abordagem cognitiva com protocolo verbal. **Ponto de Acesso**, Salvador, v. 7, n. 1, p. 42-66, abr. 2013.

Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/8135>>.

Acesso em: 25 set. 2013.

FUJITA, Mariângela Spotti Lopes. Prólogo. In: GIL LEIVA, I. **Manual de indización: teoría y práctica**. Gijón: Ediciones TREA, 2008. p. 9-12.

FUJITA, Mariângela Spotti Lopes; REDIGOLO, Franciele Marques. O uso de linguagens documentárias por indexadores em contexto de bibliotecas universitárias: uma abordagem sociocognitiva com protocolo verbal. **Ibersid**, Zaragoza, v.3, p.125-132, 2009. Disponível em: <http://www.iversid.eu/ojs/index.php/iversid/article/view/3732/3493>. Acesso em: 24 set. 2013.

FUJITA, M. S. L., RUBI, M. P., BOCCATO, V. R. C. O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária. **Datagramazero**, v.10, p.1-24, 2009.

FUJITA, Mariângela Spotti Lopes. O contexto profissional do indexador no ensino de indexação. R. Eletr. Bibliotecon. **Ci. Inf.**, Florianópolis, v. 15, n. 30, p.91-104, 2010.

FUJITA, M. S. L. **O contexto da leitura documentária de indexadores de bibliotecas universitárias em perspectiva sociocognitiva para a investigação de estratégias de ensino**. Marília: Unesp - Faculdade de Filosofia e Ciências, 2006 (Projeto Integrado de Pesquisa- CNPq).

GIBBS, Graham. **Análise de dados qualitativos**. Tradução de Roberto Cataldo Costa. Porto Alegre: Artmed, 2009.

GIGANTE, Maristela Cid. Os sistemas de classificação bibliográfica como interface biblioteca/usuário. **Ciência da Informação**, Brasília, v. 25, n. 2, 1995. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/436/394>>. Acesso em 06 out. 2013.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.

GIL LEIVA, I. **Manual de indización: teoría y práctica**. Gijón: Ediciones TREA, 2008.

GIL LEIVA, Isidoro. Aspectos conceituais da indexação. In: _____. **Política de indexação**. São Paulo: Cultura Acadêmica, 2012.

GUINCHAT, Claire; MENOU, Michel. **Introdução geral às ciências e técnicas da informação e documentação**. Brasília: IBICT, 1994.

GONÇALVES, Maria Carolina. A percepção de usuários sobre a indexação na análise de assuntos para a catalogação. In: FUJITA, M. S. L., org., et al. **A indexação de livros: a**

percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009. p. 96-117. Disponível em: <<http://books.scielo.org/id/wcvbc/pdf/boccatto-9788579830150-06.pdf>>. Acesso em: 19 out. 2013.

IZQUIERDO ALONSO, Mónica. Nuevos retos en el análisis documental de contenido: la gestión de la forma documental del contenido. **Scire**. n. 10:1, jun. 2004. p. 31-50. Disponível em: <<http://www.iberid.eu/ojs/index.php/scire/article/view/1477/1455>>. Acesso em: 13 out. 2013.

KOSTINA, Irina Y. **Dinamicidad de los conceptos especializados en los textos de diferente nivel de especialización**: trabajo de investigación. 2000. Tese (Doctorado em Teoría de La Traducción) – Programa Del Doctorado em Teoría de La Traducción, Universidad Autónoma de Barcelona, Bella Terra, 2000.

KRIEGER, Maria da Graça. Terminologia revisitada. In: KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker (org.) **Temas de Terminologia**. Porto Alegre/São Paulo: UFRGS/ USP, 2001.

KRIEGER, Maria da Graça; FINATTO, Maria José Bocorny. **Introdução à Terminologia: teoria e prática**. São Paulo: Contexto, 2004.

LANCASTER, F.W. **Indexação e Resumos**: teoria e prática. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 1993.

LANCASTER, F.W. **Indexação e Resumos**: teoria e prática. Tradução de Antonio Agenor Briquet de Lemos. 2. ed. Brasília: Briquet de Lemos, 2004.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ci. Inf.**, Jan 2002, vol.31, n.1, p.41-52. Disponível em: <<http://www.scielo.br/pdf/ci/v31n1/a05v31n1.pdf>>. Acesso em: 25 out. 2013.

MARTÍNEZ TAMAIO, Ana María; VALDEZ, Julia C. **Indización y clasificación em bibliotecas**. Buenos Aires: Alfagrama, 2008. 235 p.

MIRANDA, Antonio. **Ciência da Informação**. Brasília: Thesaurus, 2003.

MOREIRA, M. P.; MOURA, M. A.. Construindo tesouros a partir de tesouros existentes: a experiência do TCI – Tesouro em Ciência da Informação. **Data Grama Zero**: Revista em Ciência da Informação, v. 7, n. 4, ago. 2006. Disponível em: <http://www.dgz.org.br/ago06/F_I_art.htm> Acesso em: 25 set. 2013.

MOREIRO, José Antonio. Criterios e indicadores para evaluar la calidad del análisis documental de contenido. **Ci. Inf.**, Brasília, v. 31, n. 1, p. 53-60, jan./abr. 2002. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/174/153>>. Acesso em: 23 set. 2013.

MOURA, Maria Aparecida. Leitor-bibliotecário: interpretação, memória e as contradições da subjetividade em processos de representação informacional. In: NAVES, Madalena Martins

NASCIMENTO, Lúcia Maria Barbosa; GUIMARÃES, José Augusto Chaves. Documento jurídico digital: a ótica da diplomática. In: Passos, Edilenice (Org). Informação jurídica: teoria e prática. Brasília: Thesaurus, 2004. p.33-77.

NAVES, Madalena Martins Lopes; KURAMOTO, Hélio (Org.). **Organização da informação**: princípios e tendências. Brasília: Briquet de Lemos, 2006.

NEVES, Dulce Amélia de Brito; DIAS, Eduardo Wense; PINHEIRO, Ângela Maria Vieira. Uso de estratégias metacognitivas na leitura do indexador. **Ci. Inf.**, Brasília, v. 35, n. 3, p. 141-152, set./dez. 2006. Disponível em: <<http://www.scielo.br/pdf/ci/v35n3/v35n3a14.pdf>>. Acesso em: 23 set. 2013.

OLIVEIRA, Dalgiza Andrade; ARAÚJO, Ronaldo Ferreira de. Construção de linguagens documentárias em sistemas de recuperação da informação: a importância da garantia do usuário. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 17, n. 34, p.17-30, maio./ago., 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p17/22620>>. Acesso em: 30 set. 2013.

PASSOS, Edilenice. O Futuro da Biblioteca Jurídica. Brasília, DF:Infolegis, 2005. Disponível em: <<http://www.infolegis.com.br>>. Acesso em: 16 maio 2014.

PINTO Molina, María; GÁLVEZ, Carmen. **Análisis documental de contenido** : procesamiento de información. Madrid : Sintesis, 1996.

PONTES, Adriana Moura de. **OPAC como recurso para a gestão da informação no contexto da biblioteca central da UFPB**. João Pessoa: UFPB, 2006. 85 p. Monografia (Especialização) – Curso de Especialização em Gestão de Unidades de Informação, da Universidade Federal da Paraíba, João Pessoa, 2006.

ROBREDO, J.; CUNHA, M. B. da. **Documentação de hoje e amanhã**. Brasília: ed. Autor, 1986.

ROSENTAL, Claude; FRÉMONTIER-MURPHY, Camille. Introdução aos métodos quantitativos em ciências humanas e sociais. Lisboa: Instituto Piaget, 2002.

ROWLEY, J. **A biblioteca eletrônica**. Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 2002.

RUBI, Milena Polsinelli. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: FUJITA, M. S. L., org., et al. **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias**. Um estudo de observação do contexto sociocognitivo com protocolos verbais [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009. p. 81-93. Disponível em: <<http://books.scielo.org/id/wcvbc/pdf/boccatto-9788579830150-06.pdf>>. Acesso em: 16 out. 2013.

SANZ CASADO, Elías. **Manual de estudios de usuarios**. Madrid: Pirâmide, 1994.

STREHL, Leticia. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. **Ci. Inf.**, Brasília, v. 27, n. 3, p. 329-335, set./dez. 1998.

UNISIST. Princípios de indexação. Revista da Escola de Biblioteconomia da UFMG, Belo Horizonte, v. 10, n. 1, p. 83-94, mar. 1981.

STUMPF, Ida Regina Chitto. A comunicação da ciência na universidade: o caso da UFRGS. In: MUELLER, Suzana Pinheiro Machado; PASSOS, Edilenice Jovelina Lima. **Comunicação científica**, Brasília : UnB, 2000.

SUÁREZ, M.; CABRÉ, M. Teresa. La variación denominativa em los textos de especialidad: indícios lingüísticos para su recuperación automática. In: VIII Simposio Iberoamericano de Terminologia, 2002, Cartagena de Indias. **Actas...** Cartagena de Indias: IULATERM, 28 a 31 oct. 2002.

VAN DER LAAN, Regina Helena. **Tesouro e terminologia**: uma inter-relação lógica. 2002. 185 f. Tese [Doutorado em Letras] – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2002.

VARGAS, Dóris Fraga; VAN DER LAAN, Regina Helena. A contribuição da terminologia na construção de linguagens documentárias como os tesauros. **Biblos**: Revista do Instituto de Ciências Humanas e da Informação, v. 25, n.1, p.21-34, jan./jun. 2011.

VIEIRA, Simone Bastos. Indexação automática e manual: revisão de literatura. **Ciência da Informação**, Brasília, v. 17, n.1, p. 43-57, jan./jun. 1988.

WEITZEL. Simone da Rocha. Fluxo de informação científica. In: POBLACION, Dinah Aguiar; WITTER, Geraldina Porto; SILVA, José Fernando Modesto da (Org.). **Comunicação & produção científica** : contexto, indicadores e avaliação. São Paulo: Angellara, 2006.

APÊNDICE A – Seleção inicial de expressões de busca.

COLETA	EXPRESSÃO DE BUSCA	LINHA
9	"LEI MARIA DA PENHA"	1931
1	"prisão devedor alimentar"	4336
3	"garantia aplicação da lei penal tráfico de drogas"	444
3	"ato doloso de improbidade administrativa"	3427
3	"abeas corpus"	5541
3	"anecéfalo"	9626
2	"uso documento falso"	10867
2	"crimes eletrônicos"	7014
31	"preconceito religioso"	9339
29	""a motivação das decisões penais""	1461
30	"falsificação documento atestado"	3711
30	"falsificação assinatura"	5077
30	"Novo Código PENal"	5795
24	"Contravensões Penais"	8273
27	"drogas"	3028
27	"organização criminosa"	3287
27	"Os crimes hediondos"	6456
26	"organizações criminosas"	1503
26	"coisa julgada penal"	2407
26	"motorista profissional embriagado "	3345
26	"motorista embriagado "	3346
26	motorista profissional embriagado ao volante	3372
26	"Código penal art. 26"	5578
26	"execução de prestação de alimento"	6262
26	"execução alimentícia"	6238
26	"desacato"	6561
25	"anteprojeto do código penal"	4550
25	art 20 3º código penal"	5899
25	art 20 §3º homicídio culposo"	6081
22	"crimes internet"	45
22	"Anistia Pol?tica"	406
21	"abuso de autoridade"	6168
21	"anistia politica"	7858
20	"certidao criminal"	1771
20	"pirâmide financeira"	7456
20	"deixar de prestar socorro"	8472
19	"crimes cibernéticos"	4264
19	"legalizaçãõ dos bingos"	6689
18	"alteracao cena crime pericia"	563
18	"bullying"	3320
17	"constrangimento ilegal insignificância"	86
17	"anistia de Policiais Militares"	3103

16	“progressão de regime falta de estabelecimento”	1
16	“invasão de domicílio”	367
15	"crime de trânsito perícia local do acidente"	671
15	"ocultação de cadáver"	2172
14	"LEI 4.898/65"	5702
13	"organizações criminosas"	2480
13	"lei anticorrupção"	2957
13	"prisão alimentos pagamento 30% "	5674
6	"LEI DE ORGANIZACOES CRIMINOSAS"	1179
5	“Teoria finalista do delito”	1134
6	"anistia de policiais militares"	2148
6	"menor infrator"	3353
7	"Repressão ao crime organizado"	10264
8	"Trabalho prisional"	320
8	"trabalho apenado"	325
8	"181 código penal"	1962
8	“prescrição criminal	2383
12	"crimes cibernéticos"	221
12	"crimes digitais"	222

APÊNDICE B – Expressões de busca qualificadas para a análise.

ITEM	EXPRESSÃO DE BUSCA	PARTE DA EXPRESSÃO E/OU EXPRESSÕES USADAS PARA ANÁLISE
1	LEI MARIA DA PENHA	Lei Maria da Penha
2	lei penal tráfico de drogas	lei penal
3	uso documento falso	documento falso
4	falsificação assinatura	falsificação assinatura
5	crimes eletrônicos; crimes cibernéticos; crimes digitais; crimes internet	crimes eletrônicos
6	preconceito religioso	preconceito religioso
7	a motivação das decisões penais	decisões penais
8	Novo Código Penal; anteprojeto do código penal	Novo Código Penal
9	drogas	drogas
10	organização criminosa; organizações criminosas; Repressão ao crime organizado; LEI DE ORGANIZACOES CRIMINOSAS;	crime organizado
11	Teoria finalista do delito	Teoria finalista
12	prescrição criminal	prescrição criminal
13	181 código penal	181 código penal
14	menor infrator	menor infrator
15	ocultação de cadáver	ocultação de cadáver
16	LEI 4.898/65; abuso de autoridade	abuso de autoridade
17	Trabalho prisional; trabalho apenado	Trabalho prisional
18	constrangimento ilegal insignificância	constrangimento ilegal
19	alteração cena crime perícia	cena crime
20	progressão de regime falta de estabelecimento	progressão de regime
21	deixar de prestar socorro	deixar de prestar socorro
22	crime de trânsito perícia local do acidente	crime de trânsito
23	certidão criminal	certidão criminal
24	Anistia Política; anistia de Policiais Militares	Anistia Política; anistia de Policiais Militares
25	lei anticorrupção	lei anticorrupção
26	invasão de domicílio	invasão de domicílio
27	legalização dos bingos	bingos
28	bullying	bullying
29	coisa julgada penal	coisa julgada penal
30	pirâmide financeira	pirâmide financeira
31	art 20 3º código penal; art 20 §3º homicídio culposo	art 20 3º código penal
32	motorista embriagado; motorista profissional embriagado ao volante;	motorista embriagado

	motorista profissiona embriagado	
33	desacato	desacato
34	Código penal art. 26	Código penal art. 26

APÊNDICE C – Comparação entre expressão de busca e descritor representativo do mesmo no VCB.

ITENS	EXPRESSÃO DE BUSCA	VCB DO SENADO FEDERAL
1	LEI MARIA DA PENHA	Não existe um descritor
2	lei penal	Legislação penal
3	documento falso	Falsificação de documentos
4	falsificação assinatura	Falsificação de documentos
5	crimes eletrônicos	Crime por computador
6	preconceito religioso	Discriminação religiosa
7	a motivação das decisões penais	Decisão em matéria penal
8	Novo Código Penal	Não existe um descritor
9	drogas	Entorpecente
10	organização criminosa	Crime organizado
11	Teoria finalista	Não existe um descritor
12	prescrição criminal	Prescrição extintiva (direito penal)
13	181 código penal	Não existe um descritor
14	menor infrator	Delinquente juvenil
15	ocultação de cadáver	Cadáver
16	LEI 4.898/65; abuso de autoridade	Abuso de poder
17	Trabalho prisional	Trabalho penitenciário
18	constrangimento ilegal	Crime contra a liberdade individual
19	cena crime	Local do crime
20	progressão de regime	Regime penitenciário
21	deixar de prestar socorro	Omissão de socorro
22	crime de trânsito	Delito de trânsito
23	certidão criminal	Antecedentes
24	Anistia Política; anistia de Policiais Militares	Anistia
25	lei anticorrupção	Corrupção administrativa
26	invasão de domicílio	Violação de domicilio
27	bingos	Jogo de azar
28	bullying	Assédio moral
29	coisa julgada penal	Coisajulgada (processo penal)
30	pirâmide financeira	Crime contra a economia popular
31	art 20 3º código penal	Não existe um descritor para os artigos. Homicídio culposo
32	motorista embriagado	Embriaguez ao volante
33	desacato	Desacato à autoridade
34	Código penal art. 26	Não existe um descritor

APÊNDICE D - Categorias das expressões de busca e descritores respectivos.

Categoria: Legislação.

ITEM	EXPRESSÃO DE BUSCA¹	DESCRITOR NO VCB¹
1	Lei Maria da Penha	Não existe um descritor
8	novo código penal	Não existe um descritor
13	181 código penal	Não existe um descritor
25	lei anticorrupção	Corrupção administrativa
31	art 20 3º código penal	Não existe um descritor
34	Código penal art. 26	Não existe um descritor

Categoria: Variantes denominativas.

ITEM	EXPRESSÃO DE BUSCA¹	DESCRITOR NO VCB¹
2	Lei penal	Legislação penal
3	documento falso	Falsificação de documentos
4	falsificação assinatura	Falsificação de documentos
5	crimes eletrônicos	Crime por computador
6	preconceito religioso	Discriminação religiosa
7	decisões penais	Decisão em matéria penal
9	drogas	Entorpecente
10	organização criminosa	Crime organizado
11	teoria finalista	Não existe um descritor
12	prescrição criminal	Prescrição extintiva (direito penal)
14	menor infrator	Delinquente juvenil
16	abuso de autoridade	Abuso de poder
17	trabalho prisional	Trabalho penitenciário
19	cena crime	Local do crime
21	deixar de prestar socorro	Omissão de socorro
23	certidão criminal	Antecedentes
26	invasão de domicílio	Violação de domicílio
29	coisa julgada penal	Coisa julgada (processo penal)
32	motorista embriagado	Embriaguez ao volante

Categoria: Especificidade.

ITEM	EXPRESSÃO DE BUSCA¹	DESCRITOR NO VCB¹
14	ocultação de cadáver	Cadáver
18	constrangimento ilegal	Crime contra a liberdade individual
20	progressão regime	Regime penitenciário
22	crime de trânsito	Delito de trânsito
24	Anistia política	Anistia
27	bingos	Jogo de azar
28	bullying	Assédio moral
30	pirâmide financeira	Crime contra a economia popular
33	desacato	Desacato à autoridade

¹ Conforme APÊNDICE C.