



**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO**

GUSTAVO DINIZ DO NASCIMENTO

**DOS SINTAGMAS NOMINAIS AOS DESCRITORES DOCUMENTAIS: ESTUDO DE
CASO NA INDEXAÇÃO DE TESES E DISSERTAÇÕES DA ÁREA DE DIREITO**

**Recife - PE
2015**

GUSTAVO DINIZ DO NASCIMENTO

**DOS SINTAGMAS NOMINAIS AOS DESCRITORES DOCUMENTAIS: ESTUDO DE
CASO NA INDEXAÇÃO DE TESES E DISSERTAÇÕES DA ÁREA DE DIREITO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de mestre em Ciência da Informação.

Área de concentração: Informação, Memória e Tecnologia.

Linha de pesquisa: Comunicação e Visualização da Memória.

Orientador: Prof. Dr. Renato Fernandes Corrêa

**Recife- PE
2015**

Catálogo na fonte
Bibliotecária Maria Valéria Baltar de Abreu Vasconcelos, CRB4-439

N244d Nascimento, Gustavo Diniz
Dos sintagmas nominais aos descritores documentais: estudo de caso na indexação de teses e dissertações da área de Direito / Gustavo Diniz Nascimento. – Recife: O Autor, 2015.
198 f.: il.

Orientador: Renato Fernandes Corrêa.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro de Artes e Comunicação. Ciência da Informação, 2015.
Inclui referências, apêndices e anexos.

1. Ciência da informação. 2. Indexação automática. 3. Recuperação da informação. 4. Língua portuguesa - Sintagma nominal. 5. Sistemas de recuperação da informação - Direito. 6. Representação do conhecimento (Teoria da informação). I. Corrêa, Renato Fernandes (Orientador). II. Título.

020 CDD (22.ed.) UFPE (CAC 2015-215)

GUSTAVO DINIZ DO NASCIMENTO

*Dos sintagmas nominais aos descritores documentais: estudo de caso na
indexação de teses e dissertações da área de Direito*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de mestre em Ciência da Informação.

Aprovada em: 20/11/2015

BANCA EXAMINADORA

Prof. Dr. Renato Fernandes Corrêa (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Fabio Assis Pinho (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Bruno Tenório Ávila (Examinador Externo)
Universidade Federal de Pernambuco

Dedico este trabalho a minha mãe Irisma Diniz Nascimento, porque nunca mediu forças para que eu estudasse. Sempre me incentivou a prosseguir e olhar para o futuro. Foi uma verdadeira guerreira, amiga e uma mãe MARAVILHOSA.

AGRADECIMENTOS

Agradeço ao Programa de Pós-graduação em Ciência da Informação da UFPE pela oportunidade dada para desenvolvimento desta pesquisa.

Agradeço a todo o corpo docente do programa, em especial ao meu orientador Prof. Dr. Renato Fernandes Corrêa pela contribuição para o meu crescimento profissional, pelas discussões e pela paciência que teve comigo perante as dúvidas que surgiam no decorrer do curso e da pesquisa. Sou grato, também, ao professor Fábio Pinho, pela paciência sempre e ajuda na compreensão e entendimento mais amplos acerca da organização da informação.

Sou agradecido a Sheila Cavalcante, com a qual construí uma grande amizade durante o desenvolvimento desta pesquisa.

Sou grato a Edson de Souza, um amigo e irmão que me ajudou durante os aperreios com a pesquisa.

Por fim, à minha Mãe, a qual sempre me apoiou em tudo, especialmente, no desenvolver desta pesquisa. A ela, dedico este trabalho, pois sem o apoio da mesma nada teria alcançado. Sou grato a ela por tudo que até hoje fez por mim

RESUMO

O uso dos Sintagmas Nominais como instrumentos de organização e recuperação da informação digital vem se mostrando uma alternativa promissora para sistemas de informação. Nesse contexto, a indexação automática por meio de Sintagmas Nominais se mostra como uma alternativa que minimiza alguns problemas encontrados na indexação baseada em palavras isoladas, uma vez que os Sintagmas Nominais se constituem em unidades sintáticas que possuem semântica/sentido específico. No entanto, é notório que nem todos os Sintagmas Nominais que se encontram em um documento digital são representativos do mesmo, o que demonstra por sua vez a necessidade de uma seleção dos Sintagmas Nominais que realmente possam funcionar como descritores documentais. Nesse contexto, o presente trabalho tem como objetivo geral investigar a seleção de sintagmas nominais com valor de descritor no contexto do processo de indexação automática por meio de sintagmas nominais de resumos de teses e dissertações em português da área jurídica. Pretende-se: 1. Investigar o processo de indexação automática por meio de sintagmas nominais; 2. Verificar quais são as características de um Sintagma Nominal como valor de descritor documental; 3. Identificar na literatura científica nacional metodologias para seleção de sintagmas nominais em textos em português, bem como os critérios de seleção de cada metodologia; 4. Planejar experimento, onde os Sintagmas Nominais extraídos são categorizados quanto ao atendimento ou não a critérios de seleção propostos na literatura e quanto ao valor como Descritores, quando semelhantes aos descritores documentais resultantes da indexação manual; 5. Avaliar os critérios de seleção na indexação automática por meio de Sintagmas Nominais para teses e dissertações da área jurídica. Para o alcance dos objetivos propostos, fez-se uso de uma pesquisa bibliográfica e de um experimento. A pesquisa bibliográfica permitiu a identificação de pesquisas voltadas para a indexação automática por meio de Sintagmas Nominais, principalmente no que se refere à seleção de Sintagmas que funcionem como descritores documentais. Com base nas leituras dessas pesquisas, puderam-se identificar vários critérios utilizados para a seleção de Sintagmas. O experimento constituiu-se na aplicação dos critérios identificados na literatura aos sintagmas nominais extraídos de um conjunto de resumos de teses e dissertações da área jurídica, com o intuito de mensurar a utilidade ou não desses critérios no que diz respeito à seleção de sintagmas nominais descritores. Assim, realizou-se a indexação manual dos documentos, a extração automática dos Sintagmas Nominais constituintes dos mesmos, a categorização dos sintagmas nominais como descritores com base na semelhança aos descritores documentais advindos da indexação manual e a aplicação dos critérios de seleção aos sintagmas nominais extraídos. Por meio do experimento, foram percebidos comportamentos distintos entre cada critério, onde a maioria foi considerada útil para a seleção de sintagmas nominais.

Palavras-chave: Sintagmas Nominais. Indexação Automática. Recuperação da Informação. Representação da Informação. Seleção de Sintagmas Nominais. Informação Jurídica.

ABSTRACT

The use of noun phrases as tools for information organization and retrieval has proven to be a promising alternative in information systems. In this context, automatic indexing through the noun phrases shows a new alternative that minimizes some problems encountered in indexing based on single words, given that noun phrases are syntactic units with specific semantic/meaning. However, it is evident that not all the noun phrases in a digital document are representative of the content, which in turn demonstrates the need for a selection of noun phrases that may serve as documentary descriptors. In this context, this work aims to investigate the selection of noun phrases with descriptor value in the context of the automatic indexing process by noun phrases from abstracts of theses and dissertations in the area of law in Portuguese. The specific objectives are: 1. Investigate the automatic indexing process by noun phrases; 2. Identify what are the characteristics of a noun phrase with descriptor value; 3. Identify the methodologies in national scientific literature for selection of noun phrases in texts in Portuguese, as well as the selection criteria of each methodology; 4. Organization of an experiment where each extracted noun phrase is categorized as descriptor and if it meets or not the proposed selection criteria in the literature; 5. Evaluate the selection criteria in the automatic indexing by noun phrases for theses and dissertations in the legal field. To achieve the objectives, it was made use of a bibliographic research and experiment. The literature review allowed the identification of research on automatic indexing by noun phrases, especially criteria used to choose noun phrases that act as documentary descriptors. Based on the readings of this research, it was possible to identify several criteria used for the selection of noun phrases. The experiment consisted in applying the criteria identified in the literature to noun phrases extracted from a set of abstracts of theses and dissertations in the legal field, in order to measure the usefulness or not of these criteria for the selection of noun phrases descriptors. Thus it was made manual indexing of documents, automatic extraction of noun phrases thereof, the categorization of noun phrases as descriptors based on their resemblance to the descriptors of manual indexing and the application of criteria in the extracted noun phrases. Through the experiment, we notice a different behavior for each criterion, where most of the criteria was considered useful for noun phrase selection.

Keywords: Noun phrases. Automatic indexing. Information retrieval. Information organization. Noun phrase selection. Legal information.

LISTA DE FIGURAS

Figura 1 – Página de busca no Tesouro jurídico do site do Superior Tribunal de Justiça / Brasil.....	81
Figura 2 – Página de busca no Tesouro jurídico do site do Superior Tribunal de Justiça / Brasil com a expressão de busca: “legislação brasileira”.....	82
Figura 3 – Página de busca no Tesouro jurídico do site do Superior Tribunal de Justiça / Brasil já com os relacionamentos dos termos do vocabulário com o termo de busca inicial.....	83
Figura 4 – Lista de Palavras-chave.....	90
Figura 5 – Análise feita pelo PALAVRAS.....	96
Figura 6 – Falha na identificação de SNs pelo PALAVRAS.....	98
Figura 7 – Falha do PALAVRAS ao omitir termos.....	99
Figura 8 – Quantidade de SNs únicos e expressões que compuseram a lista de stop words.....	116

LISTA DE QUADROS

Quadro 1 – Exemplo de SNs identificados por um software.....	20
Quadro 2 – Subsistemas que compõem uma Biblioteca / Sistema de Recuperação de Informação.....	27
Quadro 3 – Atividades de Representação Temática da Informação.....	28
Quadro 4 – Equações de índices de consistência.....	38
Quadro 5 – Elementos que compõem o SN.....	45
Quadro 6 – Ferramentas necessárias para a indexação automática por meio de Sintagmas Nominais.....	48
Quadro 7 – Etapas do processo de indexação automática por meio de SNs.....	50
Quadro 8 – Valor atribuído ao SN de acordo com sua relevância.....	54
Quadro 9 – Valor atribuído ao SN de acordo com sua estrutura e nível.....	56
Quadro 10 – Heurísticas utilizadas para extração de conceitos em textos em língua portuguesa.....	61
Quadro 11 – Valores ideais das Categorias do Sintagma Nominal – CSN.....	67
Quadro 12 – Síntese de critérios utilizados por pesquisadores para se selecionar SNs.....	70
Quadro 13 – Critérios de seleção, escopo de aplicação dos critérios, trabalhos que utilizaram um ou outro critério de seleção.....	72
Quadro 14 – Termos Jurídicos na linguagem especializada e na linguagem comum.....	79
Quadro 15 – Palavras-chave semelhantes resultantes da indexação manual.....	93
Quadro 16 – Médias de revocação e precisão sem aplicação de critério.....	103
Quadro 17 – Quantitativo de SNs, descritores e não descritores eliminados pelo critério: descarte de SNs com numerais.....	105
Quadro 18 – Exemplos de SNs descritores eliminados pelo critério de descarte de SNs com numerais.....	106
Quadro 19 – Médias de revocação e precisão / Critério: descarte de SNs com numerais.....	107
Quadro 20 – Quantitativo de SNs descritores e não descritores, eliminados pelo critério: descarte de SNs com pronome no núcleo.....	110
Quadro 21 - Médias de relocação e precisão / Critério: descarte de SNs com pronome no núcleo.....	111
Quadro 22 – Quantitativos de SNs descritores e não descrtores eliminados pelo critério: descarte de SNs que iniciam com advérbio.....	113
Quadro 23 – SNs descritores eliminados pelo critério de descarte de SNs com advérbio no início.....	113
Quadro 24 - Médias de relocação e precisão / Critério: descarte de SNs que iniciam com advérbio.....	114
Quadro 25 – Quantitativo de SNs descritores e não descritores, eliminados pelo critério: descarte de SNs categorizados como <i>stopwords</i>	117
Quadro 26 – Médias de revocação e precisão / Critério: uso de <i>stopwords</i>	118
Quadro 27 – Quantidade de SNs que seriam gerados com a aplicação do critério de adição de SNs com	

múltiplos adjetivos.....	120
Quadro 28 - Quantitativo de SNs descritores e não descritores selecionados pelo critério: detecção de SNs por meio da remoção sucessiva de adjetivos múltiplos.....	120
Quadro 29 - Quadro 33 - Médias de revocação e precisão / Critério: detecção de SNs com múltiplos adjetivos.....	121
Quadro 30 – Quantidade de SNs que seriam gerados com aplicação do critério de adição de SNs com múltiplos adjetivos ligados por conjunção.....	123
Quadro 31 – Quantitativo de SNs descritores e não descritores selecionados pelo critério: detecção de SNs por meio de múltiplos adjetivos ligados por conjunção.....	123
Quadro 32 - Médias de revocação e precisão / Critério: detecção de SNs com múltiplos adjetivos ligados por conjunção.....	124
Quadro 33 - Quantitativo de SNs descritores e não descritores, que atenderam ao critério de nível, sendo categorizados em nível 1a, 1b, 2, 3, 4 ou 5 ou mais.....	126
Quadro 34 – Taxas de revocação e precisão alcançadas com a análise de cada nível de SN.....	128
Quadro 35 – Médias de revocação e precisão / Critério: nível.....	130
Quadro 36 – Quantitativo de SNs descritores e não descritores que se encontraram em uma das quatro partes dos documentos.....	132
Quadro 37 – Taxas de revocação e precisão alcançadas com a análise de cada uma das quatro partes dos documentos.....	133
Quadro 38 - Médias de revocação e precisão / Critério: posição do SN.....	134
Quadro 39 – Ocorrências de SNs percebidas para várias frequências.....	136
Quadro 40 – Quantitativos de SNs descritores e não descritores de acordo com a frequência de ocorrência no texto/documento.....	137
Quadro 41 – Taxas de revocação e precisão para cada frequência de ocorrência dos SNs nos textos/documentos.....	138
Quadro 42 – Médias de revocação e precisão / Critério: freq. absoluta e normalizada.....	139
Quadro 43 – Exemplos de SNs com frequências altas em todo o <i>corpus</i>	142
Quadro 44 – Quantitativo de SNs descritores e não descritores que apareceram em 1,2,3,4 ou 5 ou em mais de cinco documentos da coleção.....	143
Quadro 45 – Taxas de revocação e precisão alcançadas com as diferentes frequências de ocorrências dos SNs na coleção.....	144
Quadro 46 – SNs ordenados de acordo com o índice <i>idf</i>	144
Quadro 47 – Médias de revocação e precisão / Critério: freq. <i>idf</i>	146
Quadro 48 – Resumo das médias de revocação e precisão de todos os critérios.....	148

LISTA DE TABELAS

Tabela 1 – Regras de formação de SNs.....	46
Tabela 2 – Exemplo de SNs e níveis.....	127

LISTA DE ABREVIATURAS E SIGLAS

AFP News Briefes	Agence France Presse
BDTD	Biblioteca Digital de Teses e Dissertações
BDTD – IBICT	Biblioteca Digital de Teses e Dissertações do Instituto Brasileiro de Informação em Ciência e Tecnologia
BDTD-UFPE	Biblioteca Digital de Teses e Dissertações da Universidade Federal de Pernambuco
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CI	Ciência da Informação
CSN	Categoria do Sintagma Nominal
IDF	Inverse Document Frequency
ExATOP_{PL}	Extrator Automático de Termos para Ontologias em língua Portuguesa
KWIC	Key word in Context
KWOC	Key word out of Context
LA	Linguagem Artificial
LC	Linguagem Controlada
LDs	Linguagens Documentárias
LI	Linguagem de Indexação
LN	Linguagem Natural
PLN	Processamento da Linguagem Natural
PPGCI – UFPE	Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco
PRECIS	PREserved Context Indexing System
RI	Recuperação de Informação
SciELO	Scientific Eletronic Library Online
SDU	University of Southern Denmark
SIDSN	Sistema Identificador de Sintagmas Nominais
SISNOP	Sistema Identificador de Sintagmas Nominais do Português
SLIC	Selective Listing Combination
SN	Sintagma Nominal
SNs	Sintagmas Nominais
SRI_s	Sistema de Recuperação de Informações
TF	Term Frequency
TF-DCF	Term Frequency, Disjoint Corpora Frequency

TF-IDF	Term Frequency - Inverse Document Frequency
UFPE	Universidade Federal de Pernambuco
UNISIST	Sistema Mundial de Informação Científica
VISL	Visual Interactive Syntax Learning
WEKA	Waikato Environment for Knowledge Analysis

SUMÁRIO

1 INTRODUÇÃO.....	15
1.1 Objetivo Geral	22
1.2 Objetivos Específicos	22
2 QUADRO TEÓRICO CONCEITUAL	26
2.1 Representação Temática da Informação.....	26
2.1.1 Indexação.....	30
2.1.1.1 Análise de Assunto	33
2.1.1.2 Tradução	35
2.1.1.3 Qualidade e Avaliação da Indexação.....	37
2.1.2 Indexação Automática	39
2.1.2.1 Indexação Automática por meio de Sintagmas Nominais.....	43
2.1.2.1.1 Seleção de Sintagmas Nominais com valor de descritores	51
2.1.2.1.2 Avaliação da Indexação Automática e da extração e seleção de SNs	74
2.2 Informação Jurídica	77
3 METODOLOGIA DA PESQUISA.....	84
4 ANÁLISE DOS RESULTADOS EXPERIMENTAIS	103
4.1 Descarte dos SNs que Contêm Numerais	104
4.2 Descarte de SNs que possuem Pronome como Núcleo	109
4.3 Descarte de SNs que Iniciam com Advérbios	112
4.4 Eliminação de SNs Não Descritores categorizados como <i>Stop Words</i>	115
4.5 Detecção de SNs contidos em SNs maiores por meio da remoção sucessiva de Adjetivos.....	119
4.6 Identificação de SNs por meio de Conjunção entre Adjetivos.....	123
4.7 Estrutura e Nível do SN.....	126
4.8 Posição do Sintagma Nominal no Documento	131
4.9 Frequência de ocorrência dos SNs no Texto/Documento e Frequência Normalizada ..	136
4.10 Frequência de ocorrência dos SNs no Conjunto de Documentos – <i>Frequência Inversa de Documento – (Idf)</i>	141
5 CONSIDERAÇÕES FINAIS	152
REFERÊNCIAS	159
APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	165
APÊNDICE B – CARTA DE ANUÊNCIA DAS DUAS INSTITUIÇÕES AS QUAIS OS BIBLIOTECÁRIOS PARTICIPANTES ESTÃO LIGADOS	168
APÊNDICE C – LISTA DE SINTAGMAS NOMINAIS QUE COMPUSERAM AS <i>STOPWORDS</i>	169

APÊNDICE D – PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS (PRODUTO DA INDEXAÇÃO MANUAL).....	170
APÊNDICE E – PALAVRAS-CHAVE E SNS DESCRITORES DE CADA DOCUMENTO	180
APÊNDICE F – TÍTULOS E RESUMOS CONSTITUINTES QUE COMPUSERAM O CORPUS DESTA PESQUISA.....	192
ANEXO A – COMPROVANTE DE ENVIO DE PROJETO AO COMITÊ DE ÉTICA – UFPE	198

1 INTRODUÇÃO

No século XX, com o desenvolvimento da ciência e da tecnologia, identificou-se um momento denominado de “explosão informacional”, caracterizado pelo surgimento de uma grande massa documental. Logo, surgia a preocupação com a recuperação dessa gama de informações disponibilizadas para toda a sociedade. E é nesse momento que nasce uma ciência com o propósito de resolver esse problema informacional, onde se tinha, de um lado, a crescente produção de informação de maneira desordenada, e, do outro, indivíduos necessitados de informação, porém, sem mecanismos e recursos que facilitassem o acesso a essas informações. Esse novo campo de estudo denominado de Ciência da Informação (CI) volta seus olhares para a “informação” desenvolvendo estudos que vão desde a construção da mesma até o seu fluxo.

Segundo Oliveira (2011, p. 9) “[...] a ciência da informação nasceu no bojo da revolução científica e técnica que se seguiu à Segunda Guerra Mundial”. Vale salientar aqui o relacionamento dessa nova área com as novas tecnologias que foram surgindo à época.

Toda a conjuntura vivenciada após a segunda guerra mundial levou ao surgimento da CI, a qual pode ser compreendida como um campo que se dedica ao estudo das propriedades e do comportamento da informação, bem como as forças que governam seu fluxo e os meios de processamento para viabilizar de forma eficaz a sua acessibilidade e utilização. (BORKO, 1968, apud OLIVEIRA, 2011)

A preocupação com a organização, a guarda e a recuperação de informações já existia bem antes do surgimento da CI, no entanto as disciplinas que estavam voltadas para essas atividades se resumiam às técnicas e métodos de guarda e organização. Mas é com a CI que a preocupação vai mais além das técnicas, uma vez que a mesma constitui-se de um conjunto de teorias e práticas, que segundo Saracevic (1996, p.47) é um campo:

[...] dedicado a questões científicas e a prática profissional, voltadas para os problemas da efetiva comunicação do conhecimento e de registros de conhecimentos entre seres-humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. (SARACEVIC, 1996, p.47)

Apesar dos estudos desempenhados pela CI, a recuperação de informação na atual conjuntura ainda não é uma tarefa livre de problemas e de inconvenientes. Recuperar informação relevante na atualidade se tornou uma atividade difícil, devido a gama de informações que é disponibilizada nos vários suportes informacionais que por sua vez se

encontram nas múltiplas instituições que guardam, organizam e disponibilizam informações tanto em meio físico como em meio virtual.

Na atual cenário, onde “ter” informação a disposição não é mais o problema, e sim “tê-la de forma organizada, que se torne achável”, uma área nuclear da Ciência da Informação que se torna elemento-chave nessa situação de produção e uso da informação é a *representação temática da informação*, a qual se encontra, de modo mais amplo, dentro da *Representação da Informação*¹. (GUIMARÃES; SALES; GRÁCIO, 2012).

A representação temática da informação compreende as atividades de *catalogação de assunto*, *classificação* e *indexação*, as quais se ocupam de representar os conteúdos informacionais contidos nos documentos. Tais processos fazem parte de um conjunto de atividades que são executadas a fim de tornar disponíveis recursos informacionais com base em seu(s) assunto(s) em Unidades de Informação, tanto em meio físico quando em meio virtual. Essas atividades, em especial a “indexação” (objeto da presente pesquisa), são essenciais para que as unidades de informação alcancem seus propósitos.

Não basta que as instituições forneçam informação, faz-se necessário que a informação esteja organizada para que assim seja encontrada pelo indivíduo que dela necessita e busca. Nesse cenário, uma atividade desempenhada pelo profissional bibliotecário voltada para a representação do conteúdo dos documentos é a “indexação”. São três as abordagens que norteiam a Representação Temática da Informação, são elas: *Abordagem Norte-Americana*, *Abordagem inglesa* e *Abordagem Francesa*. A primeira voltada para catálogos, portanto, denominada de catalogação de assunto, a segunda voltada para a informação especializada, denominada de indexação e a terceira e última voltada não mais para os produtos e sim para o processo (as etapas) que são desenvolvidas no processo de representação temática de um documento, denominada de análise documentária. Esta pesquisa usará a terminologia da corrente inglesa, ou seja, utilizar-se-á o termo “indexação”, fazendo referência sempre a um processo que tem como intuito a representação de conteúdos de documentos para fins de organização e recuperação desses documentos.

Maia (2008, p. 27) de forma sucinta resume o objetivo primordial da atividade de indexação, considerando-a como:

¹ Neste trabalho, utiliza-se a expressão “Representação da Informação” para designar as atividades que representam o documento, tanto fisicamente quanto tematicamente. No entanto, são encontradas na literatura outras expressões que possuem o mesmo sentido. Em alguns momentos desse trabalho utiliza-se expressões diferentes de “Representação da informação”, como, por exemplo, Organização da Informação, em respeito ao uso de terminologia diferente por alguns autores citados nesse trabalho.

O processo de indexação produzindo uma lista de descritores visa à representação dos conteúdos dos documentos. Ou seja, esse processo tem como objetivo extrair as informações contidas nos documentos, organizando-as para permitir a recuperação destes últimos. Contudo, na maioria dos sistemas convencionais de recuperação de informação, os descritores não passam de uma simples lista de palavras extraídas dos documentos, que constituem a coleção. (MAIA, 2008, p. 27).

A indexação, em particular, é considerada como uma das principais atividades que são desempenhadas em bibliotecas e unidades de informação, principalmente, na atual conjuntura, onde coleções nas unidades de informação crescem rapidamente, acarretando, por sua vez, em uma grande demanda de materiais para serem indexados.

O ambiente virtual, por suas características peculiares, em detrimento ao ambiente tradicional, disponibiliza uma maior quantidade de informações, o que acarreta em uma necessidade maior de procedimentos de organização dessas informações. Com o grande volume de informação disponibilizado nos meios físicos e “mais” ainda nos meios virtuais, surge a necessidade emergente da prática da indexação, bem como o desenvolvimento de mecanismos que auxiliem a indexação realizada manualmente. A variedade de materiais informacionais que são disponibilizados atualmente requer um processamento que acompanhe a velocidade com que as informações são produzidas.

No ambiente virtual, as coleções estão crescendo em ritmo acelerado, como é o caso, por exemplo, das Bibliotecas Digitais de Teses e Dissertações - BDTDs. Essas Bibliotecas têm como propósito organizar e tornar disponíveis teses e dissertações produzidas no âmbito da instituição a qual pertencem, ou seja, trata-se de uma coleção que cresce rapidamente, pois ao final de cada turma de mestrado e doutorado da instituição, são incorporados novos materiais à coleção.

Como resposta ao fomento governamental, as instituições de ensino superior estão aumentando vertiginosamente os seus cursos de pós-graduação, proporcionando novos cursos e mais vagas. Como resultado desse aumento de cursos nas instituições de ensino tem-se um considerável volume de dissertações e teses produzidas. Assim, as BDTDs são ricas em conteúdo informacional científico e tecnológico. Esse conteúdo necessita ser bem indexado para que assim contribua para a geração de novas pesquisas, bem como para o desenvolvimento tecnológico e científico de diversas áreas do conhecimento científico em tempo hábil.

A produção vertiginosa de informações requer, por sua vez, um grande número de profissionais que processem esses materiais, no entanto essa quantidade necessária de

recursos humanos nem sempre é existente nas unidades de informação. Essa inexistência de profissionais e a grande velocidade com que novos materiais informacionais são produzidos conduzem para o surgimento de estudos e alternativas voltadas para novos métodos e técnicas de indexação que abarquem toda essa massa informacional.

Nesse contexto, a indexação tradicional que até então era uma atividade desempenhada essencialmente por um ser humano, passa a ser exercida com o auxílio do computador, chegando a ser executada completamente por ele. Na literatura de Ciência de Informação, encontram-se várias expressões relacionadas com a indexação automática. Assim, tem-se: indexação auxiliada por computador, indexação automatizada, indexação semiautomática e indexação automática, sendo esta última, realizada completamente pelo computador, e aquelas pelo ser humano com o auxílio da máquina. Essas definições são detalhadas no referencial desta dissertação.

Pode-se dizer, então, que a indexação automática surge e se desenvolve com o propósito de dar conta dessa grande quantidade de documentos digitais que precisam ter seus conteúdos representados para que assim sejam recuperados em outro momento por um indivíduo que deles necessita, além de contribuir para a solução de problemas inerentes à indexação manual, como, por exemplo, a morosidade dessa prática (BORGES; MACULAN; LIMA, 2008). Nesse contexto, a indexação automática contribui para o não esquecimento dos registros que se encontram nas várias instituições de memória, como, por exemplo, nas bibliotecas digitais, representando os registros do conhecimento que compõem essa memória para que os mesmos sejam recuperados em um dado momento, pois não se pode falar em recuperação sem se falar em guarda, em memória, uma vez que o acesso à memória exige instrumentos de organização e recuperação da mesma.

Além da capacidade de abarcar grandes volumes de coleções digitais, a indexação automática vem minimizando o problema da inconsistência que é evidenciado na indexação manual. A inconsistência que se encontra na prática manual é resultante da subjetividade que essa atividade carrega, pois como é o ser humano que a realiza, ele pode sofrer influências externas, cognitivas, pessoais etc. Vieira (1988, p. 44) diz que “essa inconsistência é o ponto de estrangulamento da indexação manual”.

A *indexação automática* tem como característica principal a *objetividade*, em contrapartida à subjetividade da manual, pois ela utiliza a mesma forma, a mesma programação para processar vários documentos em diferentes momentos. No entanto, isso não ocorre na indexação manual, pois um mesmo indexador pode representar um mesmo

documento de forma diferente em momentos distintos, ou um grupo de indexadores pode representar um mesmo documento de forma totalmente diferente.

Apesar de ter sido com base nas “palavras isoladas” que a indexação automática se desenvolveu, mais especificamente com os estudos de H. P Luhn em 1957 e Baxendale (1958), verifica-se que em muitos casos as “palavras sozinhas” não conseguem dizer muita coisa. Isso se dá pela existência da sinonímia e da polissemia na linguagem. Nesse contexto, as pesquisas sobre indexação automática buscam cada vez mais desenvolver metodologias que se preocupam com a sintaxe e com a semântica imbuídas nos textos (documentos).

Dentre as metodologias de *indexação automática* que vem se aprimorando, identificam-se algumas que levam em consideração a semântica intrínseca aos documentos, como, por exemplo, a utilização de *sintagmas nominais* (SNs) como descritores representativos de conteúdos de documentos ao invés da utilização das “palavras isoladas”.

Segundo Kuramoto (2002. p. 6) “O sintagma nominal é a menor parte do discurso portadora de informação”, em outras palavras, é o conjunto de palavras que possuem significado, possuem um sentido.

Segundo Borges, Maculan e Lima (2008) “sintagmas” são expressões que definem uma relação de dependência, onde são estabelecidos elos de subordinação entre outros elementos, que por sua vez também são sintagmas.

Martins (2014, p. 42) afirma que “O SN é um conjunto de elementos que constituem uma unidade significativa dentro da oração e que mantêm a dependência e a ordem entre seus constituintes.” Os elementos a que se refere esse autor são palavras que giram em torno de outra fundamental, definida como Núcleo do Sintagma.

Algumas pesquisas vêm sendo desenvolvidas com o intuito de demonstrar a viabilidade do uso dos SNs como recursos potenciais para a recuperação de informações como, por exemplo, Kuramoto (1995; 2002), Souza (2005; 2006), Maia (2008), Corrêa et al. (2011), Lopes (2012), Silva (2014), Souza e Raghavan (2014) entre outros.

Apesar do grande potencial que os SNs têm no que se refere à representatividade informacional, é preciso que se tenha em mente que nem todo SN que se encontra em um texto tem potencial para representar o conteúdo temático desse texto. Ou seja, a extração automática de SNs de um texto não resultará, necessariamente, em descritores documentais², pois, apesar de saber que os SNs possuem uma grande carga de semântica intrínseca em suas

² Descritores documentais são considerados no escopo desta dissertação, como termos de indexação relativos a um documento, convencionalmente palavras ou conjunto de palavras que representem conceitos relacionados ao(s) assunto (s) principal (s) desse documento.

estruturas, nem todos os SNs que se encontram em um determinado documento serão representativos do conteúdo informacional desse documento, ou seja, nem todos os SNs podem ser considerados descritores com alto potencial discriminatório para a representação de assunto. Por exemplo, do mesmo modo que existem palavras isoladas que são mais discriminantes de conteúdos de documentos para fins de indexação e recuperação de informação, alguns SNs são melhores descritores do que outros pelo fato de serem mais imprevisíveis e raros numa coleção, uma vez que SNs muito comuns têm pouco potencial discriminatório, visto que, são generalistas e comuns a diversos domínios, por exemplo.

No Quadro 1, a seguir, pode-se ver um exemplo de estruturas que apesar de serem SNs, não contribuem muito para descrição de conteúdos documentais, não servindo assim como descritores. Os SNs que se encontram nesse Quadro foram extraídos de um documento, cujo título é "Direitos Fundamentais: constituição e tratados internacionais (análise do ordenamento jurídico brasileiro)" e que trata de uma reflexão e análise dos tratados internacionais sobre Direitos Humanos, enfocando essa análise principalmente no que se refere ao art.5º, § 2º da Constituição Federal. Desse documento, por meio dos softwares, são extraídos vários SNs, dentre os quais alguns são demonstrados:

Quadro 1 – Exemplo de SNs identificados por um Software

Sintagmas Nominais extraídos do DOC.2
Direitos fundamentais
Constituição
Análise do Ordenamento Jurídico Brasileiro
o ordenamento jurídico brasileiro
a presente tese
nível hierárquico
Os tratados internacionais sobre Direitos Humanos
os Direitos Humanos
o foco deste trabalho

Fonte: desenvolvido pelo autor.

Como pode ser visto no quadro anterior, a maioria dos SNs possuem potencial discriminatório para o documento a que se referem, todavia, alguns Sintagmas pouco contribuem para a representação de documento em questão, como é o caso dos SNs “a presente tese” e “o foco deste trabalho”. É nesse sentido que esta dissertação tenta trabalhar, ou seja, propor uma reflexão acerca de critérios de seleção dos SNs que realmente funcionem

como descritores documentais, uma vez que nem todos os SNs presentes num texto representam o conteúdo desse texto.

Nesse contexto, Corrêa et al. (2011) dizem que a extração de SNs

[...]não garante por si só a seleção de bons descritores, sendo necessário que a ferramenta de extração de sintagmas nominais possa fazer a análise dos textos e pontuar os sintagmas nominais com a potencialidade de serem bons descritores [...] (CORRÊA et al., 2011, p. 11).

Nesse contexto, surge o problema da presente pesquisa, o qual consiste na avaliação das metodologias de seleção de SNs com valor de descritor, com vistas a contribuir para uma seleção mais coerente de SNs que funcionem como descritores na representação temática de documentos digitais, mais especificamente de dissertações e teses jurídicas no âmbito da BDTD/UFPE. Nesse contexto, o presente estudo volta-se especificamente para a seleção de SNs que realmente possam funcionar como descritores documentais, uma vez que, como foi mencionado, nem todos os SNs possuem poder discriminante suficiente para representar o conteúdo informacional de um documento.

A problemática do trabalho está direcionada para uma reflexão e uma análise da seleção de SNs. Sendo assim, a problemática desta pesquisa gira em torno das seguintes indagações: Em que consiste a indexação automática por meio de sintagmas nominais? O que caracteriza um SN como descritor Documental e como identificá-lo? Como selecionar sintagmas nominais com valor de descritor? Como avaliar a seleção de Sintagmas Nominais na indexação automática de teses e dissertações? Quais são os melhores critérios para a seleção de sintagmas nominais com valor de descritor?

Diante dessa problemática, a presente pesquisa tem como propósito investigar os critérios de seleção que contribuem para seleção de SNs que funcionem como descritores documentais no contexto da indexação automática de resumos de teses e dissertações do Programa de Pós-Graduação em Direito/UFPE. Escolheu-se o Programa de Direito devido ao fato de o mesmo se constituir em uma área temática de interesse do pesquisador. Além disso, a escolha dessa área se deu, também, pelo fato de a mesma possuir um vocabulário abrangente incluindo não só termos técnicos como também muitos termos da linguagem comum, como podem ser vistos na classificação de Maciel (2009) baseada em Cornu (1990), a qual será discutida em subseção do referencial teórico deste trabalho.

1.1 Objetivo Geral

No que se refere aos objetivos da presente pesquisa, como objetivo geral tem-se:

- Investigar a seleção de sintagmas nominais com valor de descritor no contexto do processo de indexação automática por meio de sintagmas nominais de resumos de teses e dissertações em português da área jurídica.

1.2 Objetivos Específicos

No tocante aos objetivos específicos, o trabalho em questão possui os seguintes objetivos:

1. Investigar o processo de indexação automática por meio de sintagmas nominais;
2. Verificar quais são as características de um SN como um bom descritor documental;
3. Identificar na literatura científica nacional metodologias para seleção de sintagmas nominais em textos em português, bem como os critérios de seleção de cada metodologia;
4. Planejar experimento, onde os Sintagmas Nominais extraídos são categorizados quanto ao atendimento ou não à critérios de seleção propostos na literatura e quanto ao valor como Descritores, quando semelhantes aos Descritores documentais resultantes da indexação manual;
5. Avaliar os critérios de seleção na indexação automática por meio de Sintagmas Nominais para teses e dissertações da área jurídica.

O interesse de se estudar essa temática, *a indexação automática por meio de sintagmas nominais*, se dá, de modo geral, pela aproximação do pesquisador com a área de representação temática da informação de forma automática, bem como com as metodologias desenvolvidas para uma representação automatizada mais fidedigna de conteúdos informacionais. A vivência profissional com a indexação manual de documentos contribuiu ainda mais para a motivação que levou a esse estudo. Esse interesse se manifestou mais intensamente com os contatos com trabalhos da área, como, por exemplo, Corrêa et al.(2011), Lapa (2014), Maia (2008), Souza (2005), Silva (2014), entre outros, por meio dos quais foi

possível perceber os avanços e as contribuições efetivas que a indexação automática de modo geral, e mais especificamente, a que faz uso de SNs vem alcançando na área da Organização e Recuperação da Informação.

Além do que foi exposto anteriormente, pode-se ressaltar como ponto motivador a identificação dos problemas suscitados pela “indexação manual” não só na literatura da área, como também na vivência da prática dessa atividade por parte do pesquisador.

Todo esse contexto influenciou diretamente a motivação em estudar essa temática, a qual se relaciona diretamente com a linha de pesquisa denominada “Comunicação e visualização da memória” do Programa de Pós-Graduação em Ciência da Informação da UFPE. Dentro dessa linha de pesquisa, o presente trabalho reside, mais especificamente, no tema: Métodos, técnicas e instrumentos de recuperação da informação, uma vez que possui como propósito a investigação das metodologias de seleção de SNs na indexação automática de teses e dissertações, com vistas a alcançar uma recuperação mais eficiente desses recursos informacionais que constituem uma parte da memória científica e tecnológica da instituição registrada na coleção digital da BDTD/UFPE.

Esta pesquisa se torna importante, pois contribui para as pesquisas que envolvem a utilização de SNs como recursos de acesso a informação, principalmente no atual contexto, onde somente parte da informação produzida é devidamente organizada. Os SNs, quando bem selecionados, possuem potencial suficiente para servirem como descritores de conteúdos de documentos em sistemas de recuperação de informação, facilitando assim o processo de busca e recuperação de documentos. Nesse contexto, a relevância deste trabalho se torna evidente, ao passo que o mesmo volta-se especificamente para os procedimentos de seleção dos SNs mais relevantes dentro de um determinado documento, contribuindo assim para as Bibliotecas Digitais, bem como os Sistemas de Recuperação de Informação que fazem uso ou estão aptas a fazerem uso desses novos recursos de acesso à informação.

Acredita-se que a presente pesquisa poderá trazer benefícios, tanto no plano teórico, uma vez que, pretende-se fazer um levantamento da literatura acerca do processo de indexação automática por meio de sintagmas nominais e das peculiaridades imbuídas em tal processo, como também no plano prático, com a avaliação das metodologias de seleção de sintagmas nominais na indexação automática de Teses e Dissertações.

Este trabalho se constitui de cinco capítulos.

Este primeiro capítulo, sendo a **Introdução**, onde foram apresentados o contexto social da pesquisa, o problema, a problemática, os objetivos propostos e a justificativa de se desenvolver este estudo.

O segundo capítulo trás o **Quadro Teórico Conceitual**, como forma de fundamentar e embasar teoricamente os assuntos tratados nesse trabalho, bem como os conceitos pertinentes para uma melhor compreensão do estudo. O referido capítulo inicia-se versando sobre a Representação Temática da Informação como uma área nuclear da CI. Em seguida passa-se para o estudo da Indexação, esta sendo subdividida em duas etapas gerais: Análise de Assunto e Tradução. Prosseguindo, coloca-se em questão a Indexação Automática, explicitando algumas características, bem como funcionalidades da mesma. Mais adiante, passa-se para a Indexação Automática por meio de SNs, onde são colocadas em relevo as contribuições desta indexação para a representação da informação digital. Finaliza-se este capítulo com um breve estudo acerca da informação jurídica.

O terceiro capítulo, **Metodologia da Pesquisa**, descreve todo o percurso metodológico utilizado para se realizar a pesquisa, desde a coleta dos materiais informacionais para a elaboração do Quadro Teórico Conceitual até a execução da parte empírica deste trabalho, bem como a coleta e análise dos resultados.

O quarto capítulo, **Análise Dos Resultados Experimentais**, faz uma análise e reflexão dos resultados alcançados com a realização da parte empírica deste trabalho, a qual se constitui de um experimento envolvendo a aplicação de um conjunto de critérios de seleção a SNs que foram extraídos de determinadas partes de dissertações e teses da área jurídica, tendo em vista verificar a viabilidade de cada critério ao selecionar ou não os SNs descritores e evitar os SNs não descritores. A rotulação dos SNs em descritores ou não foi feita com base na comparação dos SNs extraídos dos documentos com as palavras-chave atribuídas a estes mesmos documentos por um conjunto de cinco indexadores, sendo quatro bibliotecários e o próprio autor de cada documento. Em síntese, este capítulo explicita os resultados com a aplicação de cada critério, demonstrando, assim, a utilidade ou não dos critérios no contexto desse experimento.

O quinto capítulo, **Considerações Finais**, traz uma reflexão e algumas considerações acerca de todo o estudo desenvolvido, retomando os objetivos propostos no início deste trabalho. Este capítulo constitui-se de uma síntese dos resultados alcançados, retomando alguns pontos principais desta dissertação, bem como algumas sugestões de continuidade deste trabalho e trabalhos futuros.

Logo em seguida, têm-se as referências utilizadas para a construção desta pesquisa, assim como os apêndices construídos durante este trabalho. Constituem os apêndices deste trabalho: APÊNDICE A, que se constitui do termo de consentimento e livre esclarecimento que foram assinados pelos sujeitos (indexadores) do experimento deste trabalho; APÊNDICE B, que se constitui da carta de anuência que foi solicitada aos responsáveis pelas instituições, as quais os sujeitos estavam ligados; APÊNDICE C, o qual contém os SNs que compuseram a lista de stopwords, a qual foi utilizada no experimento deste trabalho; APÊNDICE D, contendo as palavras-chave atribuídas pelos indexadores aos documentos constituintes do corpus deste trabalho; APÊNDICE E, o qual trás as palavras-chave que foram escolhidas para representarem cada documento; e por fim, o APÊNDICE F, o qual trás os títulos e resumos que compuseram o *corpus* deste trabalho. Após estes apêndices, encontra-se como anexo o comprovante de submissão desta pesquisa ao comitê de ética da instituição a qual a mesma se encontra vinculada.

2 QUADRO TEÓRICO CONCEITUAL

No referencial teórico do presente trabalho, optou-se por se fazer um levantamento acerca da **Representação Temática da Informação**, abrangendo as atividades de representação de conteúdos documentais para fins de recuperação.

Logo em seguida, faz-se uma revisão da **Indexação Manual**, identificando as peculiaridades que envolvem essa atividade realizada pelo ser humano.

Depois de revisada a atividade da indexação manual, parte-se para a **Indexação Automática**, onde são analisadas as características dessa atividade e estudados os processos de indexação por meio da máquina.

Em seguida, o referencial desta pesquisa revisa a **Indexação automática por meio de sintagmas nominais (SNs)**, identificando as contribuições que essa atividade tem a dar para grandes coleções de recursos informacionais digitais. Depois de abordada a temática da **indexação automática por meio de SNs**, passa-se para um esclarecimento acerca das metodologias de **seleção de SNs com valor de descritor**, na qual são abordadas as formas pelas quais alguns pesquisadores selecionaram SNs com valor de descritores.

Conclui-se o referencial desta pesquisa com uma breve explanação acerca da **Informação Jurídica**, devido ao *corpus* da parte empírica do presente trabalho.

2.1 Representação Temática da Informação

A organização da informação é fundamental no processo de recuperação da informação, principalmente para facilitar o acesso aos documentos. São elementos constituintes da Representação Temática da Informação: a catalogação, a classificação e a indexação. Esses processos são realizados por etapas, de acordo com regras, normas e políticas adotadas pelas unidades de informação.

A expressão utilizada neste trabalho será “Representação da Informação”, mas, comumente, são usadas como sinônimos as seguintes expressões: “Tratamento da Informação” e “Organização da Informação”.

A preocupação com a organização da informação não é recente, mas adquiriu ênfase na Ciência da Informação, com o aumento de informação em circulação e pela diversidade de suportes disponíveis para registro e em especial, devido ao desenvolvimento das tecnologias da informação (NAVES; KURAMOTO, 2006).

Dentre os vários processos realizados dentro de uma Biblioteca, encontra-se o “Tratamento da Informação”, este por sua vez se subdivide em dois tipos de tratamento, o tratamento temático e descritivo. Dentre as disciplinas que constituem o “Tratamento da informação”, as que se dedicam ao tratamento temático da informação, ou seja, a catalogação de assunto, a classificação e a indexação, merecem um olhar cuidadoso, tendo em vista que as mesmas podem ser consideradas as bases dessa disciplina maior que é o tratamento da informação.

Uma Biblioteca ou um Sistema de Recuperação de Informação (SRI) são considerados sistemas que englobam vários subsistemas. De acordo com Dias e Naves (2007, p.15) esses sistemas (bibliotecas e SRIs) possuem três subsistemas que por sua vez se subdividem. São eles:

Quadro 2 – Subsistemas que compõem uma Biblioteca / Sistema de Recuperação de Informação.

<ul style="list-style-type: none"> • SUBSISTEMA DE ENTRADA <ul style="list-style-type: none"> ➤ Desenvolvimento da Coleção ➤ Tratamento da Informação ➤ Armazenagem • SUBSISTEMA DE SAÍDA <ul style="list-style-type: none"> ➤ Análise/Negociação de Questões ➤ Estratégia de Busca ➤ Busca ➤ Disseminação • SUBSISTEMA DE ADMINISTRAÇÃO
--

Fonte: Dias e Naves (2007, p. 15)

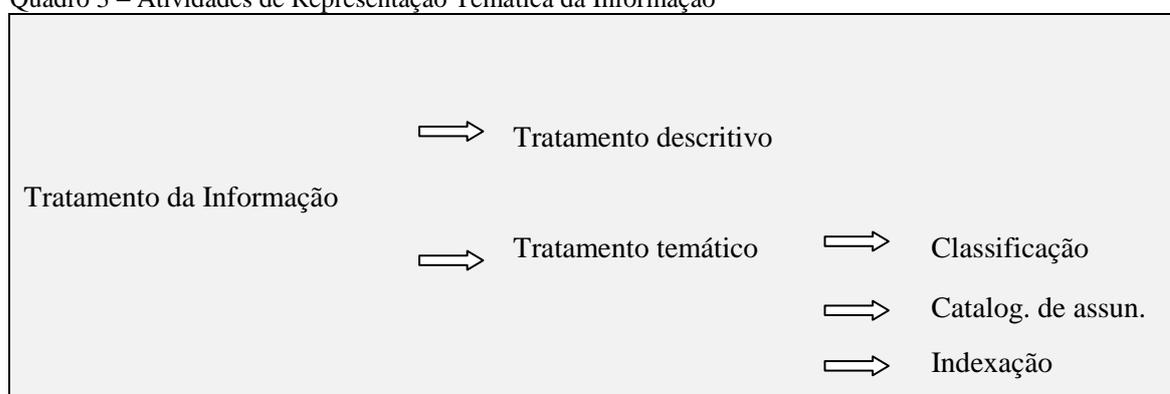
O “Tratamento da Informação” se subdivide em dois tipos de tratamento da informação, a representação descritiva (descrição física) e a representação temática (descrição temática).

A representação descritiva envolve todas as atividades que descrevem fisicamente, materialmente um determinado recurso informacional, preocupando-se com os aspectos físicos do recurso. Já a representação temática se preocupa com a representação do conteúdo, do tema, do assunto tratado no recurso informacional. A descrição do conteúdo visa proporcionar aos usuários uma representação do assunto tratado em determinada obra com vistas a mostrar-se relevante ou não para o usuário no momento da busca por informações.

Dentro do tratamento temático encontram-se as seguintes atividades: a indexação, a catalogação por assunto e a classificação (DIAS e NAVES, 2007; GUIMARÃES, 2009; REDIGOLO, 2010). Apesar de diferirem na nomenclatura e também possuírem algumas "particularidades" que as diferenciam, essas atividades possuem a mesma essência que é a identificação do assunto do documento e a tradução dessa identificação para uma linguagem de indexação.

Dentro do subsistema de entrada "Tratamento da Informação", mencionado anteriormente, as atividades são distribuídas da seguinte forma:

Quadro 3 – Atividades de Representação Temática da Informação



Fonte: Dias e Naves (2007, p. 15)

Dentro da Representação temática, um ponto que merece ser mencionado diz respeito à variação terminológica que muitas vezes acaba por gerar uma confusão conceitual e nomenclatural dentro dessa área, utilizando-se sinônimos quando não deveriam ser utilizados e confundindo atividades que são semelhantes, mas não iguais. Um bom exemplo disso é a "Análise de Assunto", a qual será estudada mais adiante em outra subseção, mas que vem ao caso mencioná-la aqui. São encontrados na literatura termos como: "Análise de Assunto", "Análise de Informações", "Análise Documentária" e "Análise Conceitual", todos voltados para a representação do conteúdo de documentos para fins de organização e recuperação.

Apesar de serem atividades distintas, a representação descritiva e temática estão relacionadas uma com a outra, formando, em um sentido mais amplo, um conjunto de atividades que juntas possibilitam a organização e disseminação de informação. Ruiz Perez (1992) rotula esses dois grupos de atividades, mencionadas anteriormente, de análise documental de forma e análise documental de conteúdo.

As atividades voltadas para a representação do conteúdo dos documentos, catalogação por assunto, classificação e indexação se apresentam, teoricamente, sob três vertentes, as

quais são explicitadas e diferenciadas logo abaixo. São elas: *Abordagem Norte-Americana-subject cataloging (Catalogação de Assunto)*; *Abordagem Inglesa- Indexing (Indexação)*; e *Abordagem Francesa – Analyse Documentaire (Análise Documentária)* (GUMARÃES, 2009).

A seguir, analisam-se um pouco essas três correntes teóricas, identificando as peculiaridades que as tornam diferentes. São elas:

a) Catalogação de Assunto³

Essa corrente anterior à abordagem inglesa e francesa, com foco pragmático é direcionada para o desenvolvimento de produtos pelas atividades de Representação Temática da Informação. Sanchez Luna (2004, p.83) entende a catalogação como uma “operação pela qual se identifica o documento em função de suas características formais e de seu conteúdo, tais como o autor, o título, o local de publicação, o editor, o ano de publicação, assim como o tema da obra”. Por esse concebimento de Sanchez Luna, percebe-se que o entendimento de “catalogação” é bem mais amplo, abrangendo operações que envolvem a descrição da forma do documento e o conteúdo do documento, esta última (descrição do conteúdo do documento) definida como *Catalogação de Assunto*, e é nessa atividade que reside o tratamento temático da informação.

b) Indexação

Esta segunda abordagem da Representação Temática da Informação se aproxima da função de pesquisa da Biblioteca, não mais voltada para o repositório documental, como na *catalogação de assunto*. Nessa abordagem a Representação Temática da Informação ocorre em centros de informação especializados, com uma ênfase maior para a pesquisa. A Representação aqui é mais voltada para o tratamento de informações que estão contidas em documentos, abordando não só o documento, como também a informação (contida nele). O direcionamento aqui é para o documento, lembrando também das necessidades do usuário.

c) Análise Documental

³ Encontra-se na literatura de Ciência da informação a utilização, em alguns casos, do termo “indexação alfabética de assunto” como sinônimo de “catalogação de assunto”, uma vez que a indexação alfabética de assunto está relacionada com a determinação de cabeçalhos de assuntos. (SILVA; FUJITA, 2004)

Verifica-se a partir do final da década de 60 e início da de 70, na França, uma preocupação, em relação ao tratamento da informação, mais voltada para a explicitação e o entendimento do processo de tratamento da informação em contrapartida às abordagens mencionadas anteriormente, as quais estavam preocupadas mais com os produtos, a Norte-Americana voltada para os catálogos e a segunda para a geração de índices.

Conforme essa abordagem francesa, introduzida por Jean-Claude Gardin (1966), a Análise Documental é entendida como um processo mais amplo, onde nele se insere a indexação. Essa abordagem francesa possui nítidos reflexos no contexto brasileiro e espanhol. No contexto brasileiro os reflexos dessa abordagem são percebidos a partir dos trabalhos de Smit (1974, 1978) e no contexto espanhol, a partir dos trabalhos de Izquierdo Arroyo e Sagredo Fernández (1983) e de García Gutiérrez (1984). (GUIMARÃES, 2009).

Em relação a essas três correntes mencionadas anteriormente, Silva e Fujita (2004, p. 137) dizem que “A existência de diferentes correntes teóricas explica o uso de termos como análise de assuntos, análise de conteúdos documentários e análise documentária.”.

Sempre que necessário, utiliza-se nesta pesquisa os autores da corrente inglesa, os quais entendem a *indexação* como um processo. A abordagem inglesa é representada por autores como Cavalcante, Foskett, Campos, Fujita, VanSlype, Lancaster entre outros.

Mais adiante, na subseção 2.1.1 Indexação, analisa-se o processo de “Indexação” de forma mais detalhada.

2.1.1 Indexação

Segundo Dias e Naves (2007), no contexto do tratamento da informação, a indexação possui dois sentidos. O primeiro sendo um sentido mais amplo quando é entendida e realizada como uma atividade de criar índices (de autor, de título, de assunto etc.). O segundo sentido é mais restrito e se refere somente à indexação de conteúdos dos documentos. Esses dois sentidos são explicitados por Farrow (1995 apud DIAS; NAVES, 2007, p. 27) ao afirmar que a:

Indexação back – of – book [mais ampla] permite ao leitor localizar informação sobre um tópico dentro do livro; a tarefa do indexador é ler o texto, distinguir entre informação relevante e periférica e empregar juntos o processamento top-down (conceitual) e bottom- up, presentes na leitura fluente. Por sua vez a indexação acadêmica [mais restrita] fornece um termo útil estabelecido pela indexação

praticada em bases de dados de resumos e em catálogos de bibliotecas usando a abordagem top-down. (FARROW, 1995 apud DIAS; NAVES, 2007, p.27)

A indexação é definida como a forma de caracterizar os conceitos que estão presentes na escrita de certo tipo de documento, de maneira a permitir a sua recuperação. O propósito principal da indexação é o de representar tematicamente esses documentos, através de termos de indexação, organizados em índices, que podem também ser denominados descritores, palavras-chave⁴ ou metadados.

É pertinente ressaltar o conceito de indexação elaborado pelo Sistema Mundial de Informação Científica (UNISIST). Esse sistema é responsável por elaborar um documento contendo “os princípios da indexação”, sendo esse documento a primeira tentativa internacional de se normalizar o processo de indexação. De acordo com o UNISIST (1977 apud VIEIRA, 1988, p. 43) a indexação é “a atividade que descreve e identifica o conteúdo de um documento, através de termos. Os conceitos dos documentos podem ser representados por termos selecionados através da linguagem natural ou por símbolos”.

Vieira (1988, p. 43) conceitua a indexação dizendo que é “uma técnica de análise de conteúdo que condensa a informação significativa de um documento através de atribuição de termos, criando uma linguagem intermediária entre o usuário e o documento”.

Para Araújo e Oliveira (2011), a “Indexação” consiste na descrição dos conteúdos dos documentos e possui como maior propósito a recuperação da informação necessitada por parte do usuário. Essas autoras assinalam ainda que a indexação “é uma das principais atividades desenvolvidas numa Biblioteca ou Unidade de Informação” (2011, p. 41).

Segundo Robredo (2005, p. 165) a indexação “[...] consiste em indicar o conteúdo temático de uma unidade de informação, mediante a atribuição de um ou mais termos (ou códigos) ao documento, de forma a caracterizá-lo de forma unívoca.”.

A indexação pressupõe um processo que, de acordo com a visão de cada autor, vai de duas até oito etapas. Há autores que a veem como sendo um processo constituído por duas etapas (LANCASTER, 2004; CHAUMIER, 1986), outros se referem a quatro etapas (ROBREDO, 2005; VAN SLYPE, 1991), encontrando na literatura autores que veem a indexação como um processo composto por oito etapas (GUINCHAT e MENO, 1983). (RUBI, 2008).

⁴ Denomina-se “palavras-chave” uma palavra ou conjunto de palavras utilizadas como termos de indexação, ou seja, termos que representem conteúdos informacionais de documentos para fins de representação e recuperação desses documentos em Bibliotecas e Sistemas de Recuperação de Informações.

A Associação Brasileira de Normas Técnicas - ABNT, em 1992, publicou uma norma, sob o número 12676, dedicada à metodologia de análise e determinação de assuntos de documentos. Essa norma intitulada: “Métodos para análise de documentos - determinação de seus assuntos e seleção de termos de indexação” define a indexação como sendo o “Ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p.2)

Para a Associação Brasileira de Normas Técnicas (1992, p.2) a indexação se constitui de três estágios, a saber:

- a) Exame do documento e estabelecimento do assunto de seu documento;
- b) Identificação dos conceitos presentes no assunto; e
- c) Tradução desses conceitos nos termos de uma linguagem de indexação.

Chaumier (1986 apud RUBI, 2008, p. 26) vê a indexação como sendo constituída por duas etapas:

- a) o reconhecimento e extração de conceitos informativos composto pela apreensão do conteúdo do documento por meio da leitura; pela identificação dos conceitos tendo em vista os objetivos do sistema de informação as necessidades dos usuários e pela seleção de conceitos segundo a exaustividade e especificidade do sistema de informação;
- b) a tradução desses conceitos.

Segundo Lancaster (2004) o processo de indexação também é constituído por duas etapas, são elas:

- a) A análise conceitual.
- b) A tradução.

Já Chu & O'Brien (1993 apud DIAS; NAVES, 2007, p. 28) consideram o processo de indexação como sendo constituído por quatro etapas, a saber :

- Análise de Assunto do texto;
- Expressão do conteúdo do assunto nas palavras dos indexadores (Linguagem Natural);
- Tradução para um vocabulário de indexação;
- Expressão do assunto em termos do índice.

Adota-se no decorrer deste trabalho a visão de Lancaster (2004) no que se refere à etapas da indexação, compreendendo esta atividade como sendo constituída pela “análise conceitual” e “tradução”. Como sinônimo de “análise conceitual” utiliza-se neste trabalho o termo “análise de assunto”, pelo fato de alguns autores utilizarem comumente esta última expressão, autores esses que são enfaticamente utilizados no decorrer deste trabalho.

Mais adiante, verifica-se um entendimento mais detalhado das duas etapas da indexação, com ênfase na primeira etapa que é a “análise de assunto”.

2.1.1.1 Análise de Assunto

Antes de qualquer coisa, faz-se necessário que se tenha em mente que a Análise de Assunto não se constitui no simples fato de ler um documento superficialmente e decidir arbitrariamente o assunto do mesmo. Essa primeira etapa será crucial para que todo o processo de representação (indexação) ocorra de forma eficiente.

A análise de assunto, de modo geral, se constitui na subetapa mais importante da indexação de assuntos, pois é a análise, como o próprio nome indica, do assunto que é tratado em um documento. Silva e Fujita (2004, p. 149) a esse respeito ressaltam que “A determinação do tema ocorre, na leitura documentária, mediante análise conceitual para identificação dos conceitos presentes no conteúdo textual.”.

O termo *assunto* foi traduzido do inglês *aboutness*, por esse motivo encontra-se, na literatura da área de representação temática da informação (indexação), um relacionamento entre os termos: assunto, *aboutness*, tematicidade e *atinência*⁵. Este último pode ser encontrado comumente no livro intitulado “Análise de assunto” de Dias e Naves (2007).

Encontra-se na literatura nacional autores que usam o termo *tematicidade* e outros que utilizam do termo *atinência* para se referirem ao *aboutness* do documento, ao assunto do documento. Ainda sobre o uso desses termos, bem como os seus significados Begthol (1986 apud SILVA e FUJITA, 2004, p. 149) resalta que “o documento tem uma tematicidade/atinência que lhe é relativamente permanente, porém um número variado de

⁵ Termo utilizado para designar aquilo de que trata o conteúdo substantivo de uma obra, ao invés de se preocupar tanto, a princípio, com a forma ou o suporte em que a informação esta registrada. Dias e Naves (2007) ressaltam que o texto possui uma *atinência* relativamente estável, permanente, no entanto esse mesmo texto possui um número variado de significados. Esse texto pode ter um significado para um usuário em um determinado tempo de sua vida e possuir outro totalmente diferente em outra época. Enquanto que a *atinência* de um determinado documento permanece estável no tempo. O que vai variar é o significado que o usuário irá fazer daquela *atinência*.

mensagens ou significados que podem ser medidos, conforme o uso exato do documento para o usuário.”. Encontram-se alguns autores que falam de *tematicidade intrínseca* e *tematicidade extrínseca*. Assim, a tematicidade intrínseca refere-se ao conteúdo permanente do documento e a tematicidade extrínseca refere-se às diferentes interpretações que podem ser dadas ao conteúdo permanente do documento, referindo-se a questões como, por exemplo, como o documento poderá ser utilizado, para quais necessidades informacionais ele será útil, para qual clientela ele servirá, etc. (SILVA, FUJITA, 2004).

De acordo com Dias e Naves (2007) a *análise de assunto* constitui-se de três momentos. São eles: **Leitura técnica do documento; Extração dos conceitos e Determinação da atenção**. Somente depois dessas etapas é que se passa para a *tradução*. Essa análise de assunto é proposta por Vieira (1988) com a denominação de “análise intelectual”.

Para Vieira (1988) a “análise intelectual” (análise de assunto) é dividida em três fases, sendo a primeira, a **compreensão do conteúdo do documento através de leitura do documento**, a segunda, a **identificação dos conceitos principais**, e a terceira, a **seleção dos conceitos**, levando em consideração a exaustividade, especificidade e a consistência. Verifica-se que, apesar da existência de algumas diferenças em termos de nomenclatura, essas três atividades mencionadas por Vieira são semelhantes às três colocadas por Dias e Naves (2007). Apesar de semelhantes, serão expostos nessa subseção o entendimento e os conceitos colocados por esses dois autores, Dias e Naves (2007), com a complementação de outros autores também.

Com base em sua capacidade e em seus conhecimentos prévios o indexador irá iniciar o processo de análise de assunto pela **Leitura Técnica do Documento (Primeiro momento)**. Essa leitura técnica constitui-se de uma leitura do conteúdo do documento. Segundo Lancaster (1993), a leitura técnica é um mister de ler e passar os olhos pelo texto. O propósito da leitura técnica feita pelo indexador é o mesmo da que é feita por um usuário final que é o de assimilar o conteúdo do documento, com uma característica adicional, que é o fato de que o indexador tem que assimilar esse conteúdo de forma a conseguir representá-lo posteriormente nas etapas sucessivas desse processo.

A análise das partes mais informativas de um documento pode fornecer contribuições relevantes para a identificação de seu conteúdo Depois da Leitura técnica do documento passa-se para a **Extração de conceitos (Segundo momento)**.

Depois de ter analisado as partes mais “importantes” do documento e com isso tenha se chegado a um entendimento do que trata o documento, o indexador irá extrair os conceitos do documento resultando em um determinado “assunto” ou “assuntos”.

Finalizando a **Extração de conceitos**, passa-se para o terceiro momento dentro da análise de assunto que é a **Determinação da atinência**. Esse terceiro momento constitui-se de selecionar os conceitos que realmente sintetizam o conteúdo do material que esteja sendo indexado. Nesse ponto inicia-se um processo linguístico, e é aí que recai o problema de como a linguagem é utilizada na descrição para posterior recuperação.

Nesse terceiro momento da Análise de Assunto é preciso distinguir e ao mesmo tempo relacionar os conceitos de atinência e de significado, os quais já foram mencionados neste trabalho.

Uma característica peculiar da *Análise de Assunto* que merece ser mencionada aqui é o seu caráter interdisciplinar. O profissional indexador ao realizar a análise de assunto está pensando e interpretando, levando em consideração vários fatores, como, a necessidade de seus usuários, a política da instituição a qual pertence, e o vocabulário utilizado para mediação entre o documento e o usuário. Ao desempenhar essa complexa atividade, o indexador sofre influência de fatores de diversos campos do conhecimento, como, por exemplo, influências advindas da linguística, da ciência cognitiva e da lógica. (DIAS; NAVES, 2007).

A respeito das contribuições que a linguística, a lógica e a ciência cognitiva vem trazendo para a análise de assunto, encontra-se na literatura autores com estudos mais aprofundados, como, por exemplo, Pinto Molina (1994), Kobashi (1996), Naves (2000), Silva e Fujita (2004), entre outros.

Ao final dessa terceira etapa, o indexador terá em mente, com base em todo esse processo intelectual, o entendimento do que trata o documento, e poderá passar para a segunda etapa da indexação, vista mais adiante.

2.1.1.2 Tradução

Com o resultado da etapa de análise de assunto, o indexador irá converter esse resultado para uma linguagem de indexação⁶ (tesauros; listas de cabeçalhos de assuntos e

⁶ Utiliza-se, no decorrer deste trabalho, a expressão “Linguagem de indexação”, todavia encontram-se na literatura, vários termos sinônimos, como “ linguagens artificiais”, “linguagens controladas” e ” linguagens documentárias”. Em alguns momentos utiliza-se um ou outro termo distinto, tendo em vista há autores que

esquemas de classificação.), ou poderá utilizar a própria linguagem natural (a que ocorre no próprio documento, dos próprios autores). Ou ainda, pode-se utilizar uma linguagem mista, ou seja, fazendo uso tanto da linguagem natural como também da linguagem de indexação.

Em bases de dados textuais os campos do *título* e do *resumo* registram os termos da *Linguagem Natural (LN)*, enquanto que os campos de descritores termos de indexação ou identificadores registram os termos da *Linguagem Controlada (LC)*, sendo esta última também denominada de vocabulário controlado, linguagem artificial linguagem documentária ou linguagem de indexação.

A Linguagem Natural é a linguagem que ocorre no próprio documento, ou seja, nos títulos, nos resumos e no corpo do trabalho. Lancaster (2004, p.250) diz que esta linguagem é “utilizada habitualmente na escrita e na fala, e que é o contrário de vocabulário controlado”.

As linguagens de indexação (controladas, artificiais ou documentárias) são instrumentos essenciais para a representação do conteúdo dos documentos. Do mesmo modo que a catalogação precisa de seus manuais de catalogação, a indexação também necessita de suas linguagens de indexação, seus manuais e suas políticas.

Van Slype (1983 apud FUJITA e GIL LEIVA, 2010,) compreende a linguagem de indexação ou linguagem documentária como sendo um sistema de representação do conteúdo dos documentos e das perguntas, tendo como finalidade a recuperação dos documentos. Para isso é dotada de estrutura própria, controlada, padronizada e hierarquizada.

Um bom exemplo de linguagem de indexação que é bem discutido na literatura da área de organização e recuperação da informação é o “Tesouro”. De acordo com Curras (1995), citado por Moreira (2003, p. 24), pode-se considerar o Tesouro como

uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos linguísticos que a compõem – termos, simples ou compostos – encontram-se relacionados entre si sintática e semanticamente”. (CURRAS,1995 apud MOREIRA 2003, p. 24)

Em relação à escolha de uma ou de outra linguagem para a indexação, Lancaster (2004) sintetiza alguns pontos que devem ser levados em consideração na escolha da Linguagem Natural ou Linguagem de Indexação. Dentre os fatores estudados pelo teórico, tem-se o da “RECENTIDADE” dos termos, que na LN é bem mais evidente devido à ocorrência dos termos empregados pelos autores, enquanto que nas Linguagens de Indexação

utilizam a expressão “linguagem documentária”, outros a expressão “linguagem controlada”, etc. Essa variação se deve pelo fato das correntes teóricas, as quais já mencionadas neste trabalho.

esses termos irão demorar um pouco para serem adicionados. Outro ponto é que, em uma busca exaustiva se recomendará o texto completo (LN), enquanto que em uma busca seletiva, provavelmente se recomendará a indexação com termos controlados. (LANCASTER, 2004).

Depois de vistas algumas peculiaridades acerca do processo de indexação, passa-se mais adiante para uma reflexão acerca dos conceitos que estão envolvidos na avaliação do processo de indexação.

2.1.1.3 Qualidade e Avaliação da Indexação

No âmbito da atividade de indexação são utilizados alguns conceitos referentes à qualidade do processo, bem como referentes à avaliação dessa atividade, como, por exemplo, os conceitos de *consistência* e *relevância*. Quando se fala em buscar uma indexação consistente, significa indexar de forma coerente. E o que seria uma indexação coerente? Na indexação, a coerência é a concordância na atribuição de termos por diferentes indexadores para representar o conteúdo de um documento, ou seja, o grau de semelhança com que indexadores representam um mesmo documento, ou um mesmo indexador em diferentes momentos. Já a relevância é um conceito na indexação que diz respeito ao nível de satisfação do usuário final com a representação do documento que ele procura, ou seja, refere-se ao grau de satisfação por parte do usuário ao buscar e encontrar um determinado documento em uma base. Outros conceitos presentes na literatura da área de indexação são: Pertinência, Exaustividade e Especificidade.

Em relação à especificidade, Lancaster (1993, p. 27) afirma que

O princípio que, isoladamente, é o mais importante da indexação de assuntos, e que remota a Cutter (1876), é aquele segundo o qual um tópico deve ser indexado sob o termo mais específico que o abranja completamente. Assim, um artigo que trate do cultivo de laranjas será indexado sob LARANJAS e não sob FRUTAS CÍTRICAS ou FRUTAS. (LANCASTER, 1993, p. 27)

No tocante à consistência, Vieira (1988) comenta que a “*Consistência*” é o ponto de estrangulamento na prática da indexação manual.

Uma medida bem comum para avaliação da indexação, de acordo com Lancaster (1994, p. 61), é a simples relação “ $AB (A+B)$ ”, onde A representa os termos atribuídos pelo indexador a , B representa os termos atribuídos pelo indexador b , e AB representa os termos com os quais a e b concordam.”

Segundo Gil Leiva (2008) a avaliação da indexação pode ser intrínseca ou extrínseca. A avaliação intrínseca refere-se, como o próprio nome indica, interno, ou seja, voltado para os resultados da indexação, descritores, cabeçalhos, etc., com vistas a identificar a qualidade desses descritores. Narukawa, Gil Leiva e Fujita (2009, p. 108) ressaltam que “A avaliação intrínseca da indexação pode ser qualitativa, isto é, por meio de valoração e consensos entre especialistas, ou quantitativa, através de fórmulas”.

Logo abaixo, encontra-se um quadro onde Gil Leiva, Rubi e Fujita, (2008, p.236) reúnem outras fórmulas utilizadas para calcular a consistência entre duas indexações, inclusive uma variação da fórmula mencionada mais acima:

Quadro 4 – Equações de índices de consistência.

Hooper (1965)	Rolling (1981)
$\frac{C}{A+B-C}$ <p>Uma variante dessa equação é:</p> $\frac{100C}{C+A+B}$ <p>onde,</p> <p>C= Termos comuns nas duas indexações A= Termos usados na indexação A mas não na B B= Termos usados na indexação B mas não na A</p>	$\frac{2C}{A+B}$ <p>onde,</p> <p>C= Termos comuns nas duas indexações A= Termos usados na indexação A B= Termos usados na indexação B</p>

Fonte: (GIL LEIVA, 2008, p, 386 apud GIL LEIVA; RUBI; FUJITA, 2008, p.236)

No que se refere à avaliação extrínseca, Gil leiva (2008, apud NARUKAWA; GIL LEIVA; FUJITA, 2009, p. 109) diz que nessa avaliação ”o resultado da indexação é utilizado para compará-lo com a indexação de outra unidade de informação que também indexou os mesmos documentos (interconsistência) ou para testar a função da indexação na recuperação.”. (exaustividade e precisão na recuperação).

Através da incorporação da indexação em uma base de dados, a mesma poderá ser avaliada mediante o cálculo de dois índices: índices de revocação e índices de precisão. Essas medidas, segundo Lopes (1985) podem ser representadas pelas seguintes formulas:

$$\text{Revocação} = \frac{\text{quantidade de referências relevantes recuperadas} \times 100}{\text{quantidade total de referências relevantes existentes na base}}$$

$$\text{Precisão} = \frac{\text{quantidade de referências relevantes recuperadas} \times 100}{\text{quantidade total de referências recuperadas}}$$

Na prática esses dois parâmetros se mostram inversos, tendo em vista que ao passo que se melhora em uma medida, piora-se em outra. (GIL LEIVA, 2012).

Ainda em relação à consistência na indexação é pertinente considerar que, segundo Gil Leiva (2008, p. 76 apud GIL LEIVA; RUBI; FUJITA, 2008, p. 236)

[...] da revisão dos resultados obtidos nos experimentos levados a cabo nos últimos anos, depreende-se que a média dos índices oscila entre os 25 % e os 60 % de coincidências. (GIL LEIVA, 2008, p.76 apud GIL LEIVA; RUBI; FUJITA, 2008, p. 236).

Outro conceito que se encontra na atividade de indexação é o de “pertinência”, o qual está intimamente relacionado com a importância de um determinado documento para o indivíduo que dele necessita. Levando em consideração o conceito de pertinência nesse contexto, deve-se entender que a indexação deve ser feita sempre em função do indivíduo que utilizará o documento a ser indexado. Lancaster (1993) se utiliza do termo “pertinente” e “relevante” como sinônimos para se referirem a itens úteis.

Para Lancaster (2004), na representação de conteúdos, um dos pontos mais importantes no momento da representação é sua **extensão**, pois, quanto maior a representação de um documento, mais pontos de acesso ele terá na base de dados da unidade de informação, e conseqüentemente com mais pontos de acesso, maior a probabilidade de esse documento ser recuperado de forma eficaz. Essa extensão é quem determinará se a indexação vai ser **seletiva** ou **exaustiva**. Quanto mais extensa a indexação maior será a **revocação**⁷ e quanto maior a especificidade maior será a **precisão**⁸, o ideal é que se consiga um equilíbrio entre a revocação e a precisão. (LANCASTER, 2004).

Passa-se mais adiante para uma análise dessa indexação realizada não mais de forma manual e sim automática.

2.1.2 Indexação Automática

⁷ A **revocação** é a razão entre o número de documentos pertinentes recuperados e o número total de documentos pertinentes disponíveis na base de dados. Alguns autores da área de Ciência da Informação utilizam o termo “abrangência” para se referirem à revocação.

⁸ A **precisão** é a razão entre o número de documentos pertinentes recuperados e o número total de documentos recuperados. Ou seja, como o próprio nome indica, capacidade de ser Preciso, evitando o que não é relevante.

Com base na literatura da área de CI, bem como na área de Representação Temática da Informação em conjunto com outras áreas, como, por exemplo, a Computação, percebe-se que vários são os estudos acerca das alternativas de se resolverem os problemas suscitados pela indexação manual, bem como também o desenvolvimento de sistemas de indexação automática como forma de acompanhar o crescimento da massa documental no ambiente virtual.

A indexação manual, desenvolvida pelo ser humano, vem com o decorrer do tempo se mostrando inadequada para suprir os problemas relacionados à subjetividade dessa atividade, uma vez que, ao ser desenvolvida pelo ser humano, sofre influências externas o que acaba afetando o resultado de todo o processo.

O ser humano por natureza é subjetivo e como tal, acaba passando essa subjetividade para as suas atividades cotidianas. Nesse sentido, foram identificados problemas relacionados à consistência no processo de indexação manual. O indexador pode indexar um documento um dia e pode também indexar esse mesmo documento outro dia de forma totalmente diferente, o que pode ocasionar a falta de consistência no processo de representação do conteúdo. Fatores externos ao indexador, como, por exemplo, ambiente de trabalho, podem influenciar esse profissional e contribuir para representações não tão fidedignas, não tão consistentes. Além disso, a crescente produção de informação, principalmente em meio digital, motivou pesquisadores e estudiosos a estudarem formas alternativas de indexação, surgindo assim a indexação automática.

Gil Leiva (2012) ressalta a existência de três âmbitos que interferem na indexação, mais precisamente, na consistência da indexação manual, a saber: o âmbito do indexador (formação, conhecimentos profissionais e motivações); o âmbito do objeto a ser indexado; e o âmbito do contexto onde a indexação é realizada. Em contrapartida, esse mesmo autor salienta que na indexação automática não ocorre fatores externos que podem afetar o próprio sistema, pois um programa de computador (software) indexará da mesma maneira vários documentos em diferentes momentos a menos que suas regras sejam alteradas.

O surgimento e o desenvolvimento da indexação automática relacionam-se diretamente com a tentativa de solucionar, ou talvez minimizar, alguns problemas suscitados pela indexação manual e também como uma forma de indexação que dê conta de todas as informações digitais que estão sendo produzidas no ambiente virtual.

No que se refere à conceituação da indexação automática, Vieira (1988, p. 48), de forma simples e sucinta, define-a como “uma operação que identifica, através de programas

de computador, palavras ou expressões significativas dos documentos para descrever de forma condensada o seu conteúdo.”.

A indexação automática, muitas vezes, é rotulada como *indexação assistida por computador*, *indexação auxiliada por computador etc.*, no entanto, alguns autores distinguiram bem as várias expressões utilizadas para designar a indexação que é totalmente realizada pela máquina (computador), ou a que é realizada com o auxílio de recursos tecnológicos, etc. Gil Leiva (1997 apud LAPA, 2014) é um desses autores que define bem essa variedade de expressões, uma vez que para este autor a *indexação auxiliada por computador* refere-se a programas que auxiliam na representação temática de documentos com base na indexação intelectual (feita pelo homem), a *indexação semiautomática* refere-se àquela que ocorre em sistemas que indexam os documentos, mas que no final do processo os termos são validados pelo ser humano, e a *indexação automática* é a que é totalmente desenvolvida pelo computador, o qual é preparado e estruturado para desempenhar tal atividade.

No que se refere às pesquisas voltadas para a associação da informática com o processamento de textos em linguagem natural, Coulon e Kaiser (1992 apud DODEBEI, 2002, p. 45) ressaltam que as pesquisas nessa vertente se dividem em quatro fases, são elas: “fase estatístico-morfológica (1945-1955); fase sintática (1955-1970); fase lógica (1963-); e fase cognitiva (1974-)”. O surgimento e o desenvolvimento da *indexação automática* podem ser relacionados com o uso de programas computacionais para geração de índices pré-coordenados. Nesse contexto identificam-se, por exemplo, o desenvolvimento do Selective Listing Combination – SLIC, o PREserved Context Indexing System –PRECIS e o Keyword in context – KWIC (BORGES; MACULAN; LIMA, 2008)

Na década de 50 surgiram os primeiros trabalhos voltados para a indexação automática baseada em frequência de ocorrência de palavras. Luhn (1957) e Baxendale (1958) foram os precursores nesses estudos. (BORGES; MACULAN; LIMA, 2008)

É a partir do KWIC que a indexação se baseia nas palavras significativas dos títulos dos documentos. Desenvolvido por H. P. Luhn em (1957), o KWIC corresponde a um índice rotativo em que cada palavra-chave que aparece nos títulos dos documentos torna-se uma entrada do índice. (BORGES; MACULAN; LIMA, 2008). A partir do KWIC desenvolveram-se outros sistemas, no entanto, levando em consideração a “palavra isolada” na maioria dos casos.

Ao indexar um documento, o indexador, como foi colocado anteriormente neste trabalho, se utiliza de sua capacidade cognitiva para interpretar e relacionar os signos (palavras) com o intuito de entender e apreender o conteúdo informacional de um determinado texto, documento etc., ou seja, é com a junção das palavras isoladas e com o relacionamento entre elas que o indexador compreende o assunto tratado em um determinado documento e assim indexa-o. Assim, a palavra solta, fora do contexto, pode muitas vezes pouco contribuir para a representação do conteúdo informacional de um documento, demonstrando assim a sua incapacidade em representar conteúdos documentais.

De acordo com Borges, Maculan e Lima (2008) “signo” é uma palavra que, isoladamente, pode ter sentido para um determinado indivíduo, mas não possui significado. Para que esse signo possua significado é preciso que ele esteja inserido em um determinado contexto, um discurso, uma situação. Essa autora faz essa distinção de forma simples e objetiva, informando que o signo isolado possui um sentido que é uma abstração pessoal, ou seja, é a forma como cada pessoa entende um determinado signo.

Complementando o entendimento acerca do uso da palavra “isolada”, Kuramoto (1995) diz que sistemas que se utilizam da “palavra isolada” como forma de se representar e recuperar informação podem não ser eficazes devido a fatores como a sinonímia, a polissemia e a combinação de palavras em ordem diferentes, o que pode acarretar em alteração de sentido.

Sendo assim, surge a indexação baseada não nas palavras “soltas”, “isoladas”, e sim em grupos de palavras que possuam significado (sentido), os chamados *Sintagmas Nominais* – SNs.

Antes de adentrar numa reflexão acerca dos SNs, vale ressaltar aqui a importância de uma área que vem contribuindo em muito para que a indexação automática alcance seus propósitos, principalmente no que diz respeito à indexação que faz uso dos SNs. Essa área é a *Linguística Computacional*⁹.

Silva (2014, p.32) sintetiza de forma objetiva e clara o propósito do PLN dizendo que:

O PLN tenta reproduzir, compreender, interpretar e apresentar informações na linguagem que os seres humanos usam no cotidiano, usando leis gramaticais e observando as relações das palavras com suas categorias gramaticais e seus significados. (SILVA, 2014, P. 32)

⁹ Encontram-se, na literatura, autores que se utilizam da expressão Processamento de Língua Natural, ou Processamento de Linguagem Natural como sinônimos de Linguística Computacional, como também são evidenciados autores que entendem a Linguística Computacional como sendo dividida em duas subáreas distintas: a Linguística de *corpus* e o Processamento da Língua Natural.

Mais adiante, são apresentadas as formas de avaliação da indexação automática, uma vez que é por meio da avaliação que se julga a eficiência ou não de determinado procedimento, ou seja, é a avaliação de determinado procedimento que dirá se o mesmo alcança seus objetivos eficientemente ou não.

2.1.2.1 Indexação Automática por meio de Sintagmas Nominais

Como foi mencionado anteriormente, o uso da “palavra”, apesar de ter sido a primeira unidade base para a indexação automática, foi se mostrando aos poucos ineficiente para fins de representação e recuperação de informações.

Algumas tentativas foram sendo desenvolvidas com o intuito de se utilizar outras unidades ou conjuntos de unidades em SRIs com o propósito de tornar a recuperação de informação mais eficiente. Exemplos dessas tentativas são as técnicas de grupamentos de substantivos, do inglês, *noun groups*. Uma metodologia que segue essa mesma linha, no entanto extrapola a proposta de identificação de grupamentos de substantivos é a que propõe a identificação de Sintagmas Nominais. Esta nova proposta vem se mostrando promissora, uma vez que faz uso dos SNs presentes nos documentos para servirem como recursos de acesso a esses documentos em outro momento.

Kuramoto (2002, p.3) em artigo voltado para o estudo dos sintagmas nominais como unidade base para a recuperação de informações lança a seguinte pergunta acerca do uso da “palavra isolada” como fonte para acesso a informação: “a utilização dessa unidade seria um fator limitante na melhoria da precisão dos resultados nos procedimentos de recuperação de informação?”. Com base nesse questionamento colocado por este autor, várias indagações surgem a respeito da validade dessa unidade como fonte informativa para a indexação e recuperação de informação.

O autor da ideia de se utilizar os Sintagmas Nominais como descritores ao invés das palavras isoladas foi Michel Le Guern (1991). Esse autor é responsável pelo desenvolvimento conceitual acerca desse recurso como unidade portadora de significado para a indexação e recuperação de informação. O referido autor faz uma distinção relevante entre **descriptor** e **palavra**, uma vez que o descriptor utilizado para a recuperação de informação deveria ser uma unidade do discurso e não uma unidade da língua (signo isolado sem significado).

Os **descritores** deveriam fazer referência à realidade extralinguística do autor. A palavra como uma unidade da língua constitui um conjunto de propriedades, no entanto sem referência a realidade extralinguística. Complementando esse entendimento, Kuramoto (1995), acerca do potencial das palavras tidas de forma isolada e das palavras relacionadas com outras palavras, diz que as **palavras** passam a ter valor referencial a partir do momento que as mesmas se encontram dentro de um universo do discurso.

Kuramoto (1995), em artigo dedicado ao estudo do tratamento temático e da recuperação de documentos por meio de sintagmas nominais, exemplifica a distinção entre as palavras isoladas e os SNs, tomando como exemplo a palavra **livro** e o termo **livro de bolso**. Enquanto que “livro” isoladamente possui várias definições, ou seja, vários predicados, o termo “livro de bolso” já não possui vários predicados livres, fazendo referência a uma classe específica de objetos do universo do discurso.

Outro estudo que complementa o que foi mencionado anteriormente acerca dos SNs, e que é direcionado para a indexação automática é o de Brito (1992). O estudo desse autor pode ser considerado um dos primeiros estudos nessa vertente no Brasil. Esse autor apresentou uma visão diferente sobre a análise e descrição linguística, fundada sobre uma descrição mais rica dos fenômenos linguísticos e que está na origem das reflexões sobre o tratamento automático da informação. Assim, pode-se considerar Le Guern (1991) como um dos precursores no estudo do uso dos SNs como descritores, e Kuramoto (1995) e Brito (1992) como precursores no uso dos SNs na indexação automática. Sendo Kuramoto (1999) orientado em seu doutorado pelo próprio Michel Le Guern.

De acordo com a literatura, os Sintagmas Nominais podem ser utilizados, também, por exemplo, para a criação de ontologias em determinado domínio (LOPES, 2012). Como o propósito deste trabalho é a indexação e recuperação de informações dentro do âmbito de Bibliotecas Digitais, os Sintagmas Nominais aqui são entendidos e usados como recursos de acesso à informação digital, uma vez que os mesmos são potenciais para a indexação e recuperação de informação.

Segundo Perini (2010, p. 251) o Sintagma Nominal é “um constituinte composto de uma ou mais palavras, que apresenta certas propriedades”. As propriedades a que se refere esse autor são: a) o SN pode ocorrer nas funções de **sujeito**, **objeto** ou **complemento de preposição**, e b) semanticamente, o SN pode se referir a uma entidade do mundo (real ou imaginário); essa entidade pode ser entendida como um objeto específico (por exemplo, *minha caneta*), uma classe geral (*os seres humanos*) ou uma abstração (*a inteligência*).

Kuramoto (2002, p. 6), no tocante a conceituação dos SNs, diz que “um sintagma nominal é a menor unidade do discurso portadora de informação”. Um SN pode ser tanto uma palavra isolada como também um conjunto de palavras que possui semântica e sintaxe.

De acordo com Perini (1998 apud CORRÊA et al. 2011) os Sintagmas Nominais possuem duas estruturas, uma estrutura à esquerda do núcleo e outra à direita do núcleo. O núcleo é o elemento essencial à existência do SN, esse núcleo pode ser um substantivo, um pronome substantivo, um numeral ou uma palavra substantivada. A estrutura à esquerda do núcleo é composta por determinantes, como, artigos, possessivos etc., e na estrutura à direita se encontram modificadores ou até outros sintagmas nominais. Assim, um SN pode ser constituído por apenas uma estrutura, o núcleo, (um nome), como também por três estruturas, como, por exemplo, determinantes + núcleo + modificadores (ou outros SNs).

Os elementos que se posicionam ao lado do núcleo (nome), denominados de limitadores, limitam o sentido expresso pelo núcleo, ou seja, limita a referência do nome. Esses limitadores são constituídos por elementos pré-nucleares (predeterminantes, determinantes, quantificadores, possessivos sintéticos¹⁰ e numeral) e pós-nucleares (modificadores). Os modificadores podem ser compostos de uma ou mais palavras e podem ocorrer antes ou depois do núcleo. (PERINI, 2010)

Segue abaixo um quadro elaborado com base em Perini (2010) com o intuito de esclarecer os possíveis elementos que compõem o SN:

Quadro 5 - Elementos que compõem o SN

Elementos que compõem os SNs		
Elementos pré-nucleares	Núcleo	Elementos pós-nucleares
Predeterminantes, determinantes, quantificadores, possessivos sintéticos, numeral.	Nome (substantivo, pronome substantivo, numeral ou palavra substantivada)	Modificadores (palavra ou conjunto de palavras que qualificam o núcleo, restringem o sentido do núcleo, inclusive outros nomes que podem ser núcleos também)

Fonte: Perini (2010, p. 259).

É importante salientar que os elementos colocados no quadro anterior podem ser utilizados na construção de um SN, não sendo obrigado que para que seja um SN, se tenha

¹⁰Os possessivos sintéticos são opostos dos possessivos analíticos, tendo em vista que os possessivos analíticos ocorrem sempre após o núcleo. Os possessivos analíticos são: *dele, deles e de vocês*. (PERINI, 2010)

todos os três tipos de elementos, uma vez que apenas um nome sozinho já se constitui em um SN.

Logo abaixo segue uma tabela com regras de formação de SNs elaborada por Silva (2014, p. 50), baseado em Miorelli (2001) e Santos (2005):

Tabela 1 – Regras de formação de SNs

Regras	Exemplos
Regra geral: DET + MOD + N + MOD	A interdisciplinar Ciência da Informação
Regra 1: DET + N + MOD	A Ciência da Informação
Regra 2: N + MOD	Informação estratégica
Regra 4: DET + N	A informação
Regra 5: N	Informação
Regra 6: DET + N + DET + N + MOD	A filosofia e a ciência juntas
Regra 7: DET + DET + N + MOD	A minha recuperação da informação
Regra 8: MOD + N + MOD	Grande área da informação
Regra 9: DET + DET + N	Uma certa área

Fonte: Silva, (2014, p. 50), baseado em Miorelli (2001) e Santos (2005).

Como se pôde perceber, os SNs possuem várias possibilidades de formação e, apesar de serem regidos por algumas regras na ordenação, eles possuem algumas exceções, tornando assim o sistema flexível e maleável. Apesar de o próprio sistema dos SNs ainda carecer de estudos e pesquisas em alguns pontos (PERINI, 2010), são perceptíveis as contribuições que essas unidades do discurso, SNs, vêm alcançando como fontes de indexação e recuperação de informação em meio digital.

Os SNs podem ser organizados de acordo com os SNs que se encontram embutidos neles, uns dentro dos outros. Com base em Corrêa et al. (2011) tem-se o seguinte exemplo de SN: “Sistema de Recuperação de Informações”. Tomando esse exemplo pode-se perceber que se trata de um SN de terceiro nível¹¹, visto que, o mesmo contém dois outros SNs, sendo um de segundo nível e outro de primeiro nível. Sendo assim, tem-se: “Sistema de recuperação de informações” (nível 3), “Recuperação de informações” (nível 2) e “Informações” (nível 1)

Depois de vistas as características e algumas peculiaridades dos SNs como recursos potenciais para a indexação e recuperação de informações textuais em meio digital, faz-se necessário que se aborde um pouco as formas e as metodologias adotadas para a identificação e extração dos SNs pelos programas de computadores (sistemas) no âmbito da indexação automática por meio de SNs.

¹¹ Os sintagmas nominais, ao passo que vão sendo extraídos de outros SNs, vão sendo classificados em níveis. (KURAMOTO, 1995). Sendo assim, um SN mais simples é definido como de nível 1. Já o SN de nível 2 será aquele que contém o SN de nível 1, e assim sucessivamente. (SN 1 = sintagma nominal de nível 1, SN 2 = sintagma nominal de nível 2, SN 3 = sintagma nominal de nível 3 [...])

A percepção e a identificação dos SNs pelo homem parece ser uma atividade não muito complexa, desde que o mesmo tenha o discernimento necessário para se distinguir os “nomes” que fazem referência a alguma entidade do mundo real ou imaginário, ou seja, distinguir os “nomes” de sua língua que possuem potencial referencial e os que não possuem. Falantes e ouvintes intuitivamente identificam as palavras que se referem a algo ou a alguém e as que não se referem. No entanto, em alguns casos e situações a identificação de SNs não é tão simples, principalmente, quando é feita pela “máquina”. Em outras palavras, essa “simples” identificação de SNs, muitas vezes, não é tão natural, devido a recursos que a língua portuguesa possui como, por exemplo, as construções anafóricas e as elipses.

Exemplificando esses obstáculos encontrados no processo de identificação de SNs, pode-se citar a pesquisa de Kuramoto (1995), na qual este autor procedeu com a identificação e extração dos SNs manualmente devido a não existência de software que fizesse esse trabalho à época. Seguem abaixo algumas dificuldades encontradas por esse autor na identificação e extração manual de SNs:

- Sintagmas nominais escondidos em frases com fatoração (em outras palavras, são as frases que contêm uma sequência de palavras que precedem outro conjunto de palavras coordenadas pelas conjunções “e” e “ou”. Por exemplo: “seleção, organização e disseminação da informação pelos meios tradicionais”. Outro exemplo de fatoração são as construções que aparecem dentro de parênteses como um complemento da frase que as antecedem.);
- Artigo zero (Ausência de determinantes em SNs);
- Cálculo de anáforas (Elementos anafóricos já citados anteriormente), e
- Cálculo de elipses.

Como o propósito é extrair descritores, ou seja, Sintagmas Nominais Descritores Documentais, e não mais palavras isoladas, faz-se necessário que a máquina ao analisar o texto se utilize dos mesmos conhecimentos utilizados pelo ser humano, ou seja, conhecimentos morfológicos, sintáticos e semânticos para que assim seja capaz de reconhecer e identificar os SNs. Sendo assim, o processo de indexação utilizando os SNs compõe-se de algumas ferramentas capazes de analisar um texto levando em consideração esses conhecimentos supracitados.

Pode-se dizer que as ferramentas necessárias para que haja a indexação automática por meio de SNs são: *Etiquetadores*, *Identificadores de SNs*, *Extratores de SNs* e *Selecionadores de SNs*. Atualmente, tem-se a disposição softwares criados especificamente para uma única

atividade, como, por exemplo, etiquetar as palavras de um texto, como, também, são desenvolvidos sistemas que desempenham mais de uma dessas atividades acima citadas.

Vários autores se debruçaram em desenvolver métodos e instrumentos de extração e seleção de SNs de forma automática. Kuramoto (1995) pode ser considerado um dos precursores nesses estudos. Este autor desenvolveu um protótipo de sistema de recuperação de informação baseado na navegação nas estruturas internas dos SNs. Seguido desse autor vieram outros trabalhos, relacionados direta e indiretamente com a extração e utilização de SNs, como, por exemplo, Kuramoto (2002), Bick (2000), Vieira et. al. (2000), Miorelli (2001), Othero (2004), Souza (2005; 2006), Borges, Maculan e Lima (2008), Maia e Souza (2010), Corrêa et al. (2011), Lopes (2012), Souza e Raghavan (2014), Chaves (2013), Silva (2014) e Martins (2014).

Cada um desses trabalhos contribuiu para o desenvolvimento e o aprimoramento da indexação por meio de SNs. Uns estudaram e desenvolveram metodologias para identificação e extração de SNs, outros desenvolveram produtos (sistemas) voltados para a extração de SNs e assim por diante.

Como já mencionado, algumas ferramentas são essenciais para que a identificação e extração dos SNs se desenvolvam de forma automática. Procede-se agora com uma discussão sobre essas ferramentas na seguinte sequência: Ferramentas de etiquetagem (Tagger); Ferramentas de identificação de SNs; Ferramentas de extração de SNs; e, por fim, Ferramentas de seleção de SNs, salientado que, apesar de essas ferramentas estarem sendo analisadas individualmente aqui, elas, muitas vezes, são desenvolvidas e desempenhadas por um único sistema (software) como serão vistas no decorrer do trabalho.

Mais adiante, segue o Quadro 6 com uma síntese das ferramentas necessárias para a identificação, extração e seleção de SNs para fins de indexação e recuperação de informações digitais.

Quadro 6 – Ferramentas necessárias para a indexação automática por meio de SNs

<i>Ferramentas/sistemas</i>	<i>Funções desempenhadas</i>
Etiquetadores 1ª Ferramenta: Etiquetagem das palavras do texto.	Os etiquetadores (Taggers) têm como função identificar e rotular as palavras que compõem um texto em determinadas classes gramaticais. Categoriza as palavras e as rotulam através de etiquetas, para que essas palavras etiquetadas passem para a ferramenta seguinte, que é a identificação dos SNs. Como exemplo de etiquetadores, tem-se o FORMA que é um programa de código aberto. Outro software bem utilizado é o parser PALAVRAS, entre outros.
	Programas computacionais que têm como propósito analisar um determinado texto,

Identificadores de SNs 2ª Ferramenta: Identificação dos SNs.	observando as sequências de léxicos e aplicando esses léxicos às regras gramaticais internas dos SNs (gramática sintagmática) com o propósito de que sejam identificadas as sequências de palavras que constituam SNs.
Extratores de SNs 3ª Ferramenta: Extração de SNs.	Os extratores são ferramentas desenvolvidas para realizar, além da atividade de identificação dos SNs, a extração dos mesmos, mostrando os fora do texto.
Selecionadores de SNs 4ª Ferramenta: Seleção dos SNs.	A seleção se faz necessária, pois, muitas vezes, os SNs extraídos pela máquina não são representativos do conteúdo de determinado documento. Em suma, essa função escolhe, com base em determinados critérios, SNs com valor de descritores dentre os vários SNs extraídos pela máquina, ou seja, os melhores SNs.

Fonte: desenvolvido pelo autor.

As aplicações desses sistemas (softwares), mencionados no quadro anterior, para a Indexação automática se dão pelo fato de que esses sistemas criam termos, “descritores” de indexação e recuperação.

Atualmente têm-se vários sistemas que desempenham as funções mencionadas mais acima, como, por exemplo, tem-se o Sistema Identificador de Sintagmas Nominais do Português – SISNOP, de Morellato (2010), que identifica e extrai SNs, o OGMA, de Maia (2008), entre outros.

Miorelli (2001) propôs um método de extração de Sintagmas Nominais de textos em português aplicados a SRIs. A referida autora desenvolveu um método denominado ED-CER, o qual se constitui de dois módulos que juntos realizam a extração dos SNs para indexação e recuperação de informações.

Outra ferramenta que identifica e extrai SNs, e que se encontra disponível online ao público, é o OGMA de Maia (2008). Esta, além de outras funções, etiqueta, identifica, extrai e seleciona os SNs. Essa ferramenta foi utilizada por alguns autores em pesquisas, como, por exemplo, Corrêa et al.(2011).

O PALAVRAS, de autoria de Bick (2000), trabalha com a gramática e o léxico da Língua Portuguesa, fazendo análise do texto morfológicamente (lexemas), de grupos de palavras e da composição das orações. Uma das contribuições inovadoras desse autor está relacionada com uma nova abordagem visual das sentenças, permitindo a visualização em árvores sintáticas, marcações em cores, entre outras opções proporcionadas pelo conjunto de programas que constituem o VISL (Virtual Interactive Syntax Learning). O referido sistema pode ser acessado e utilizado livremente, com algumas exceções via internet. O PALAVRAS foi desenvolvido pelo Instituto de Linguagem e Comunicação (ISK), University of Southern Denmark (SDU).

O Parser PALAVRAS foi utilizado por Vieira et al.(2000), Miorelli (2001), Souza (2005), Santos (2005), Arcoverde (2007), Maia (2008), Lopes (2012), Silva (2014), Martins (2014), entre outros. No entanto, existem outros parsers, como, por exemplo, o Grammar Play, OLX - Parser e o Tropes, alguns disponíveis ao público, outros não.

Com base no que foi visto nessa subseção, pode-se estruturar o processo de indexação automática por meio de Sintagmas Nominais em três etapas gerais, as quais são realizadas pelas ferramentas/sistemas já explicitados anteriormente. Essas três etapas constituintes da indexação automática por meio de SNs são estruturadas no quadro abaixo, salientando que essas etapas foram elaboradas como forma de proporcionar uma melhor visualização de todo o processo de indexação automática que faça uso de SNs. Essas etapas entendidas pelo autor deste trabalho como sendo fases necessárias para uma indexação baseada em SNs foi constituída tomando como referência trabalhos que abordaram direta e indiretamente o uso dos SNs para fins de indexação, como, por exemplo, os trabalhos de Kuramoto (1995: 2002), Vieira et al.(2000), Miorelli (2001), Souza (2005:2006), Santos (2005), Souza e Raghavan (2006), Arcoverde (2007), Borges, Maculan e Lima (2008), Maia (2008), Corrêa et. al. (2011) Lopes (2012), Souza e Raghavan (2014), Silva (2014), Martins (2014).

Quadro 7 – Etapas do processo de indexação automática por meio de SNs.

Processo de indexação automática por meio de SNs	
1ª Etapa	<i>Identificação dos SNs</i> através das subetapas de “etiquetagem” e de “cotejamento dos léxicos etiquetados com as regras dos SNs”
2ª Etapa	<i>Extração dos SNs</i> do texto, mostrando-os em listas, por exemplo.
3ª Etapa	<i>Seleção dos SNs</i> , com base em critérios que os classifiquem como “Bons Descritores”

Fonte: desenvolvido pelo autor.

Essas etapas são fundamentais para que ocorra uma seleção eficiente de SNs tidos como **Descritores**, tendo em vista que o propósito é que se selecionem os SNs mais representativos do tema tratado no documento. Fazendo analogia com a indexação humana, do mesmo modo que o indexador humano seleciona os termos mais representativos de um determinado documento, a máquina também deve ser capaz de selecionar os SNs mais apropriados para a descrição do conteúdo de um documento.

O OGMA é um exemplo de software que executa todas essas três etapas mencionadas no quadro anterior. Além realizar análise de texto e calcular a similaridade entre documentos, executa a extração e a seleção de SNs. Essa extração inicia-se com a etiquetagem de cada palavra do texto, depois passa-se para a submissão dessas palavras marcadas a um conjunto de regras que são as regras necessárias à constituição dos SNs, e com base nessas regras o OGMA identifica e extrai os SNs, mostrando os SNs em ordem de importância, de acordo com critérios de seleção de SNs.

Mais adiante, analisa-se detidamente a etapa de seleção de SNs dentro do processo de indexação automática por meio de SNs, sintetizando-se algumas pesquisas que contribuem diretamente para a seleção de SNs.

2.1.2.1.1 Seleção de Sintagmas Nominais com valor de descritores

Em relação à *seleção de SNs*, Corrêa et al. (2011) dizem que alguns dos SNs extraídos não apresentam relevância para o usuário no momento de busca, ou seja, embora sejam SNs, não constituem **descritores** e não correspondem a necessidade de informação do usuário como também não são representativos dos documentos. Tal fato mostra que a extração de SNs deve ser acompanhada de estratégias de ordenação por relevância dos SNs, levando em conta critérios de frequência e posicionamento, semelhante às propostas existentes para as palavras isoladas.

Complementando o que os autores acima mencionaram, Lopes (2012, p. 33) diz que: “Um aspecto importante para a recuperação de informações textuais é o passo posterior à extração de termos, que consiste em escolher dentre os termos extraídos aqueles que são portadores de valor conceitual, e não apenas terminológico”.

Kuramoto (2002) enfatiza o que foi dito até o momento sobre *a seleção de SNs*, ressaltando que um aspecto importante e essencial para a recuperação de informações textuais é a atividade posterior à extração de termos (SNs), que consiste na escolha dentre os termos extraídos daqueles que são portadores de valor conceitual, ou seja, aqueles que funcionem como descritores de um determinado documento.

Segue-se com um breve detalhamento de algumas pesquisas que contribuem diretamente não só para os estudos voltados para o uso dos SNs como recursos para indexação e recuperação, como também contribuem para a *seleção dos* mais representativos de um documento e que, por conseguinte, podem ser considerados **Descritores Documentais**, para

que assim funcionem como Descritores Documentais. Ao final dessas sínteses, encontra-se um quadro onde são colocados os *critérios de seleção de SNs* identificados em cada pesquisa, de forma a proporcionar uma melhor visualização de como está sendo feita a seleção de SNs, isto é, a **terceira etapa da indexação automática por meio de SNs**.

Souza (2005) propôs uma metodologia para a indexação de documentos digitalizados de texto completo por meio da identificação, extração e seleção de SNs. O referido autor utilizou dois *corpora* em sua pesquisa, sendo, um primeiro corpus (composto de 15 textos) e o segundo constituído por 60 documentos textuais. O primeiro *corpus* foi utilizado para validar a extração automática de SNs, este primeiro *corpus* foi o mesmo utilizado por Kuramoto (1999). Já o segundo *corpus* foi utilizado para validar a metodologia proposta por Souza (2005). Esse segundo corpus, constituído por 60 documentos textuais, por sua vez, foi dividido em dois outros corpora, sendo um utilizado para a metodologia prospectiva, composto por 6 textos, e outro para a metodologia consolidada, composto por dois conjuntos de textos, sendo um primeiro com 30 textos e o segundo com mais 30 textos. Esse primeiro conjunto de 30 textos engloba os 6 do teste inicial. Os textos são provenientes das revistas: DataGramZero e Ciência da Informação.

Souza (2005) aplica a metodologia prospectiva, depois segue para a análise dos dados, em seguida efetua as melhoras na metodologia, passando para a metodologia consolidada e finaliza com a avaliação geral da metodologia proposta. Segue abaixo, resumidamente, as etapas de sua metodologia prospectiva, com destaque para as etapas que se constituem em critérios de seleção que podem ser automatizados:

1. Escolha do *corpus* significativo de documentos, funcionando como universo empírico da pesquisa;
2. Conversão dos formatos de arquivos para texto simples;
3. Retirar os resumos e as palavras-chave atribuídas pelos autores;
4. Extração dos SNs do corpo do texto;
5. **Ordenação dos SNs nas planilhas através da verificação da frequência de ocorrência dos SNs nos documentos;**
6. **Descarte dos SNs que apresentam frequências de ocorrência inferiores a um patamar preestabelecido;**
7. Agrupamento dos SNs remanescentes a partir dos determinantes de suas formas “canônicas”, e reordená-los;
8. **Análise manual dos SNs pré-escolhidos e decisão sobre a sua relevância como descritores, com o intuito de criação de uma *stoplist*;**
9. **Verificação da incidência dos SNs nos outros documentos do *corpus*;**
10. **Análise da estrutura e o nível do SN;**
11. **Verificação da ocorrência desses SNs, parcial ou totalmente, em tesauro específico;**

12. Avaliação da relevância dos SNs como descritores;
13. Análise comparativa entre os produtos (resumos e palavras-chave) e os SNs escolhidos como descritores, para avaliação da metodologia e
14. Análise por especialistas / Garantia literária. (SOUZA, p. 77-78)

Analisa-se, neste momento, algumas etapas específicas dessa metodologia (etapas destacadas em negrito), as quais se detêm mais à seleção dos SNs tidos como descritores. O terceiro momento da metodologia prospectiva, como pode ser visto, é a retirada dos resumos e das palavras-chave dos documentos originais, tendo em vista que os mesmos sirvam como parâmetro para comparações com os produtos da metodologia, que são os SNs extraídos e selecionados.

A extração dos SNs foi praticamente toda automática utilizando o identificador de SNs PALAVRAS e o extrator de SNs Xtractor. O resultado dessa extração foi um arquivo contendo os SNs na ordem de ocorrência nos textos. Em seguida, os SNs que apresentavam frequência inferior a um patamar pré-estabelecido eram considerados descritores insignificantes. O patamar estabelecido depende de análises que levassem em consideração a relevância dos SNs extraídos em cada faixa de frequência, além do tamanho dos textos originais. Em seguida, os SNs remanescentes passam por um agrupamento levando em consideração suas estruturas, suas formas (SOUZA, 2005). Com base nos dados empíricos analisados da metodologia prospectiva, Souza (2005) verificou que para a metodologia consolidada os dados empíricos analisados na prospectiva sugeriram a adoção de frequências de corte inferiores de 2 ou 3 ocorrências, com a possibilidade de análise concomitante de outros dos parâmetros analisados.

Depois de agrupados, os SNs passam pela análise manual com vistas a ser julgado como descritor relevante ou não. Os SNs, escolhidos por meio do julgamento humano que compuseram a lista de *stopwords* puderam ser descartados de qualquer conjunto posterior de SNs extraídos. Esta oitava etapa, segundo o autor, pode ser adotada para aperfeiçoar a metodologia automática. Depois desse “julgamento”, passa-se para a nona etapa que se dedica a verificar o aparecimento dos SNs no conjunto de documentos que compõe o *corpus*. Essa etapa se baseou no pressuposto de que quanto mais um determinado SN aparecer em um conjunto de documentos, menor é a sua relevância como descritor, ou seja, o inverso da frequência de ocorrência no *corpus*, valorizando os SNs que ocorrem em poucos documentos e penalizando os SNs extremamente frequentes em todo o conjunto de documentos. Percebe-se nesse momento a presença do conceito de especificidade nesse procedimento, uma vez que

busca SNs específicos, onde SNs generalistas podem não contribuir muito para a representação informacional em indexação (SOUZA, 2005)

Depois de se verificar a ocorrência dos SNs em outros documentos, passa-se para a análise da estrutura e o nível dos SNs como recurso complementar para a definição da relevância ou não de um determinado SN. Passando para a décima primeira etapa, verifica-se a ocorrência destes SNs em um tesouro específico. Na metodologia consolidada, essa etapa de “conferencia no tesouro” foi removida, considerada opcional, pois pouco contribuiu para a seleção dos SNs como descritores ou não. A décima segunda etapa tem como propósito avaliar a relevância dos SNs como descritores. Os dados empíricos da pesquisa contribuíram para a lógica desenvolvida nessa etapa. Essa avaliação levou em consideração os seguintes fatores: A frequência da ocorrência dos SNs no conjunto de documentos; Os níveis dos SNs; Suas estruturas sintáticas; e Sua ocorrência no tesouro da área (SOUZA, 2005).

Como produto dessa avaliação mencionada mais acima, tem-se a ordenação dos SNs segundo critérios de relevância estabelecidos, em planilhas como candidatos a descritores.

Souza (2005) se utiliza de dois conceitos para avaliar a relevância semântica dos SNs. O conceito de “Pontuação” e o de “Taxa de relevância”. Esse autor associa alguns valores (criados por ele mesmo) aos SNs de acordo com a relevância semântica percebida, conforme o Quadro 8, logo abaixo:

Quadro 8 – Valor atribuído ao SN de acordo com sua relevância

Relevância descritiva do SN	SÍMBOLO	VALOR ASSOCIADO
SN extremamente relevante como descritor	SN***	1,0
SN razoavelmente relevante como descritor	SN**	0,5
SN moderadamente relevante como descritor	SN*	0,25
SN não relevante como descritor	--	0,0

Fonte: Souza (2005, p. 93)

O autor se utilizou da seguinte fórmula para se chegar aos “valores ponderados” da pontuação dos SNs:

$$Pontuação (desc) = 1,0x(Núm. SN ***) + 0,5x(Núm. SN **) + 0,25x(Núm. SN *)$$

$$TxRelev = \left(\frac{Pontuação (desc)}{Soma das ocorrências} \right)$$

A pontuação foi mensurada atribuindo-se valor numérico arbitrário aos SNs de acordo com sua relevância percebida como descritores, e a taxa de relevância apresentou esse valor normalizado. Sendo assim, quanto maior a taxa de relevância, melhor seria a representação do assunto pelos descritores, sendo que o valor máximo é 1 – valor este que seria alcançado se a totalidade dos descritores fosse extremamente relevante, caso bastante incomum mesmo para processos de indexação manual. (SOUZA, 2005)

Em suma, a metodologia de seleção dos SNs com valor de “bons descritores” levou em consideração quatro fatores, os quais são: *As frequências dos SNs no documento; As relevâncias semânticas dos SNs ou descarte de Sns irrelevantes presentes em stoplist; A quantidade de ocorrência dos SNs na totalidade do corpus de documentos; Os níveis e as estruturas sintáticas dos SNs relevantes como descritores; e A ocorrência no Tesouro da área da CI dos SNs frequentes e relevantes*, sendo este último fator considerado como opcional, na metodologia prospectiva. (SOUZA, 2005)

Os critérios adotados para avaliação da relevância semântica dos SNs escolhidos baseou-se, como já foi mencionado, nas considerações do referido autor da pesquisa, comparando os SNs com as palavras-chave e resumos dos documentos originais.

Depois de toda a aplicação da metodologia prospectiva e de todas as mudanças necessárias para a metodologia consolidada, o autor em questão ao desenhar a “Metodologia Consolidada” leva em consideração os seguintes fatores: a frequência, a análise de relevância e o nível (estrutura) do SN. Nessa fase, Souza (2005) utilizou uma fórmula para atribuir a pontuação, para efeitos de ranking, a qual pode ser vista logo abaixo:

$$Pontuação(SN) = [(k1 * frequência(Xar)) - (k2 * ocorrência(Ytot)) + (k3 * CSN)]$$

Assim, a pontuação (SN) é o valor atribuído ao SN de acordo com os critérios apresentados. Frequência (Xar) é a frequência do SN no artigo. Ocorrência (Ytot) corresponde ao número de artigos em que o SN ocorre com frequência maior que Y. Já X, Y, k1, k2 e k3 são as constantes ajustadas de acordo com os testes. A CSN diz respeito à categoria do SN, que assume um valor de acordo com o Quadro 9 (SOUZA, 2005, p. 123):

Quadro 9 – Valor atribuído ao SN de acordo com sua estrutura sintática e nível

CSN	ESTRUTURA E NÍVEL DO SN	VALOR ASSOCIADO
1 ^a	Nível 1, estrutura (D+N)	0,25
1b	Nível 1, qualquer estrutura, exceto (D+N)	0,75
2	Nível 2, qualquer estrutura	1,0
3	Nível 3 qualquer estrutura	0,75
4	Nível 4, qualquer estrutura	0,5
5	Nível 5, ou superior, qualquer estrutura	0,25

Fonte: Souza (2005, p. 123)

Pode-se considerar essa metodologia, como um todo, como sendo realizada de forma manual e automática, uma vez que houve etapas desempenhadas manualmente e, também, automaticamente, complementando-se uma com a outra.

Com essa síntese da pesquisa de Souza (2005), percebe-se que este estudo possui contribuição impar para *Seleção de SNs com valor de descritores*. No decorrer de sua pesquisa, verifica-se a utilização de vários critérios para se chegar a possíveis SNs (Descritores). Nessa pesquisa, a atenção recai não somente sob a extração de SNs, mas também sob a seleção. Retomando a metodologia prospectiva, verifica-se que a partir da quinta etapa o autor já faz uso de critérios, pois o mesmo já ordena os SNs levando em conta a frequência de cada um. Verifica-se, também, que o autor faz uso de exclusões de SNs com baixas frequências, ou seja, é um critério de exclusão que contribuirá para que fiquem somente os melhores SNs. Outro ponto da metodologia do autor que evidencia critérios de seleção diz respeito ao fato de na oitava etapa, o autor criar uma *stoplist* com SNs que não possuem natureza de descritor, ação essa semelhante aos processos utilizados na indexação baseada em palavras “isoladas”. A verificação da incidência dos SNs em outros documentos também é um critério, ao passo que não se recomenda a utilização do SN que aparece comumente em outros documentos do mesmo conjunto, levando em consideração que um termo muito genérico tem poucas chances de ser representativo de um determinado documento. Por fim, a seleção dos SNs levando em consideração as suas estruturas e seus níveis, também, se constituem critérios, tendo em vista que a complexidade da estrutura do SN e seu nível são proporcionais a sua densidade informacional.

Souza, Alvarenga Neto e Mendes (2007) realizaram uma pesquisa que teve como foco uma proposta analítica quádrupla de análise e compreensão de fenômenos sociais, apoiando-se no desenho de um mapa terminológico-conceitual. Na referida pesquisa, esses autores complementaram uma pesquisa já realizada por Souza e Alvarenga Neto (2003), na qual foi proposto um método tríplice para delimitar o escopo conceitual da Gestão do Conhecimento.

Esse método proposto por Souza e Alvarenga Neto (2003) baseava-se em três estratégias, a saber, a análise da prática organizacional; análise intelectual da ocorrência conceitual em publicações específicas sobre o assunto; e análise de eventos relacionados ao tema. Assim a pesquisa de Souza, Alvarenga Neto e Mendes (2007) pode ser considerada como uma extensão daquela pesquisa, uma vez que adicionam a essas três estratégias uma quarta estratégia que é justamente a análise de ocorrências de descritores sobre gestão do conhecimento em um conjunto de teses e dissertações, contribuindo assim para um mapeamento semântico de determinados domínios.

Assim, Souza, Alvarenga Neto e Mendes (2007) aplicam a metodologia de seleção de descritores de Souza (2005) a um conjunto de cinco dissertações e teses, das quais foram extraídos um total de 20 conceitos mais relevantes de cada documento. Na seleção dos referidos conceitos, foram expurgados os descritores que mencionavam diretamente o termo “Gestão do Conhecimento” (SOUZA; ALVARENGA NETO; MENDES, 2007). Esses autores salientam que não foi utilizado *stoplist*, que melhoraria o desempenho da metodologia de seleção.

Como foi mencionada, a metodologia de seleção de descritores utilizada por Souza, Alvarenga Neto e Mendes (2007) é a metodologia proposta por Souza (2005), sendo assim os critérios utilizados para a seleção de descritores foram os mesmos utilizados por Souza (2005), a saber, a frequência de ocorrência do SN no texto analisado; a frequência de ocorrência no *corpus*; o nível e a estrutura do SN.

A pesquisa de Souza, Alvarenga Neto e Mendes (2007, p. 25) em suas conclusões considerou que essa proposta “em suas quatro vertentes, confirma a percepção de que é mister o afastamento da estéril dissensão puramente terminológica acerca de diversas temáticas e campos do conhecimento”.

Maia (2008) em sua tese investigou a utilização de Sintagmas nominais “pontuados” como recursos para classificação por similaridade e aglomerados de documentos eletrônicos. Esse autor propõe no arrolar de sua pesquisa o desenvolvimento de uma ferramenta denominada OGMA, a qual tem como objetivo proporcionar a automação da extração de SNs e, também, o cálculo da pontuação de cada SN na indexação de documentos, além de trabalhar com classificação automática de documentos.

Para a classificação e o agrupamento de documentos, o autor utilizou o OGMA e a ferramenta WEKA¹² (*Waikato Environment for Knowledge Analysis*). O OGMA foi utilizado como ferramenta de análise de texto.

O primeiro desafio para o desenvolvimento do OGMA diz respeito à constituição do léxico da língua portuguesa com vistas à futura etiquetagem, atividade essencial para a identificação e a extração de SNs. Maia (2008) se utilizou do vocabulário utilizado pelo BR/ISPELL, depois utilizou a ferramenta Conjugue para a elaboração da base dos verbos, diferenciando também os verbos no particípio. Baseando-se na gramática de Tufano (1990) reuniu-se 475 palavras de diversas classes gramaticais. Essas mesmas palavras foram utilizadas para compor a lista de *stopword*. Depois de elaborado o dicionário, passou-se para a resolução das “ambiguidades” recorrentes no léxico, como, por exemplo, os termos: mato (verbo) e mato (substantivo). Para a resolução desse problema, a ferramenta em questão (o OGMA) formou uma lista com todas as combinações possíveis de etiquetas gramaticais encontradas e submeteu a frase e cada combinação às regras de extração dos SNs.

Depois de etiquetados, fazia-se necessário que os SNs fossem identificados e extraídos e, para tal, o OGMA utilizou um conjunto de regras para conseguir extrair os SNs, as quais foram baseadas no método ED-CER de Miorelli (2001).

O desenvolvimento do OGMA baseou-se na metodologia proposta por Souza (2005), o que levou essa ferramenta a automatizar as atividades colocadas por esse autor em sua proposta de seleção dos SNs de forma manual. O OGMA foi desenvolvido por Maia (2008) com orientação de Souza (2005). Sendo assim, esse software OGMA foi projetado com o intuito de desempenhar as seguintes funcionalidades: extração dos SNs; atribuição de pesos aos SNs de acordo com sua frequência no texto e, também, sua ocorrência dentro de outros SNs; identificação da classe dos SNs de acordo com a CSN proposta por Souza (2005); cálculo da pontuação de todos os SNs extraídos; extração dos SNs e atribuição dos pesos aos mesmos; extração dos termos isolados, exceto os constantes na lista de *stopwords* e, por último, calcular a similaridade entre os documentos, através da comparação de duas listas de termos (extraídos dos documentos) utilizando o coseno¹³.

¹² Essa ferramenta se constitui de um conjunto de algoritmos de aprendizado de máquina para tarefa de mineração de dados. Agrega diversos algoritmos provenientes de diferentes abordagens na subárea da inteligência artificial. Entre esses algoritmos, têm-se os de classificação e agrupamento, entre outros. Disponível em: < <http://www.cs.waikato.ac.nz/ml/weka/>>.

¹³ No campo da estatística, existem duas medidas que são utilizadas para verificação de similaridade e que são aplicáveis a outros estudos, essas medidas são: correlação e coseno. O coseno é resultante da comparação de similaridade entre dois documentos, onde (1) equivale a um documento que é semelhante (igual) a outro, e

Maia (2008), no que se refere à extração dos SNs, faz comparações entre o sistema ED-CER, o OGMA e o VISL (Visual Interactive Syntax Learning). Por meio da comparação da extração de SNs feita pelo OGMA, o ED-CER e o VISL, percebeu-se que o OGMA alcançou um resultado eficaz em relação aos outros sistemas.

Maia (2008) em sua pesquisa utilizou dois *corpora*, um primeiro *corpus* composto por 50 artigos selecionados do Encontro Nacional de Pesquisa em Ciência da Informação, e o segundo *corpus* composto por textos menores e de conteúdo jornalístico. O primeiro *corpus* foi aplicado somente à metodologia prospectiva e o segundo foi aplicado à metodologia consolidada. Ao segundo *corpus*, além de textos jornalísticos o autor adicionou outro *corpus* de outro tema. O interesse em se utilizar dois *corpora* foi o de perceber como cada *corpus* se comportava, tendo em vista que eram de assuntos distintos.

A pesquisa de Maia (2008) pode ser dividida em três momentos: o primeiro voltado para o desenvolvimento da ferramenta OGMA, o segundo com a aplicação do experimento prospectivo e o terceiro já com a metodologia consolidada.

O resultado alcançado pelo experimento prospectivo demonstrou que os métodos de análise por termo isolado obtiveram melhor classificação do que os que utilizaram SNs, sendo que o método de SNs pontuados atingiu um maior índice de classificações corretas, demonstrando uma expressiva melhora em relação ao método que considerou apenas os SNs.

Ao final do experimento consolidado, Maia (2008) verificou que, apesar de resultados semelhantes, o uso de SNs necessitou de um processamento computacional bem mais complexo em comparação com o uso do método de termos isolados sem *stopwords*, que por sinal é bem mais simples. Sendo assim, a relação custo *versus* benefício de se utilizarem os SNs na classificação e no agrupamento de documentos eletrônicos não se mostrou tão viável. Os piores resultados de classificação ficaram para o uso de todas as palavras que compõem o documento. Já na comparação do uso de SNs pontuados ou não, verificou-se um maior grau de eficiência com o uso de SNs pontuados.

Por fim, a pesquisa de Maia (2008) demonstrou entre outras coisas, que a utilização de SNs é capaz de representar o conteúdo dos documentos, servindo como descritores ou características para o processo de classificação ou descoberta de conglomerados.

No tocante à seleção de SNs com valor de descritores, pode-se considerar que a metodologia de seleção utilizada por Maia (2008) consiste no cálculo da pontuação dos SNs levando em consideração os mesmos critérios levantados por Souza (2005), ou seja, a

(0) equivale a um documento que não é semelhante a outro, ou seja, não possui similaridade. Quanto mais próximo de (1) mais semelhantes são os documentos.

frequência dos SNs no documento; a incidência dos SN em outros documentos; os níveis dos SNs e as estruturas sintáticas dos SNs.

Corrêa et al. (2011) realizaram pesquisa com o intuito de verificar como os SNs podem ser utilizados para a indexação e recuperação de informações em meio digital, especificamente no âmbito da Biblioteca Digital de Teses e Dissertações da UFPE. Foram realizados estudos de caso com o propósito de analisar a extração de SNs e a utilização dos mesmos como fontes de acesso a informação.

O *corpus* utilizado por esses autores se constituiu de 30 resumos de três programas de pós-graduação da UFPE, sendo dez do programa de direito, dez do programa de ciência da computação e outros dez do programa de nutrição. Esses autores utilizaram o OGMA como ferramenta para a extração dos SNs dos resumos. Os resumos foram extraídos do sistema TEDE – BDTD-UFPE a partir dos metadados no formato MTD-BR das primeiras teses e dissertações inseridas no sistema da BDTD.

O processo de extração dos SNs através do OGMA foi avaliado através do cálculo e da análise dos percentuais de precisão em extrair SNs relevantes como descritores, da taxa de erro em extrair conjunto de caracteres que não constituíam SNs, mas que o software reconheceu como SN, e, também, o percentual dos SNs extraídos, mas que não são relevantes como descritores.

Como resultados, os autores identificaram que a ferramenta OGMA se comportou de maneira distinta perante os resumos das três diferentes áreas. Esse comportamento diferente pode ser compreendido devido à terminologia de cada domínio. Foi alcançada uma maior eficiência na extração de SNs nos resumos do programa de Direito. Esse dado pode ser justificado pelo fato de os termos da área jurídica serem mais comuns no léxico do OGMA.

Corrêa et al. (2011) ressaltam que à “ordenação por relevância” deve ser dada uma maior atenção. No que diz respeito à “seleção de SNs com valor de descritores”, percebe-se que a pesquisa desses autores evidencia bem a questão de que somente a “extração de SNs” não é suficiente para a indexação e recuperação de informações em ambiente digital, pois existem SNs que em nada contribuem para a representação de um determinado documento, o que por sua vez corrobora com o que Kuramoto (2002) e Lopes (2012) ressaltaram sobre a importância de uma seleção dos SNs extraídos.

Outra pesquisa que se faz pertinente para este trabalho de dissertação é a de Lopes (2012), a qual propõe um processo de extração de conceitos/termos, conceitualmente relevantes para um determinado domínio caracterizado por um *corpus* em língua portuguesa,

por meio de Sintagmas Nominais. Esse processo é implementado por uma ferramenta denominada ExATO_{LP} (Extrator Automático de Termos para Ontologias em Língua Portuguesa). Esse processo é composto basicamente por quatro etapas gerais, a saber: 1) Extração de termos e contextos; 2) Ordenação de termos de acordo com sua relevância; 3) Identificação de conceitos; e 4) Geração de recursos linguísticos (aplicações de conceitos gerados).

Foram utilizados um *corpora* constituído por cinco *corpus*. São eles: Pediatria; Modelagem Estocástica; Mineração de Dados; Processamento Paralelo e Geologia, no entanto, somente o primeiro (pediatria) foi utilizado inicialmente para avaliação de todas as etapas propostas pelo processo de Lopes (2012), devido ao fato de ser o único domínio que possuía uma lista de termos de referência para serem comparados com os termos extraídos pelo processo proposto.

Extraídos os SNs candidatos a conceitos, foram aplicados um conjunto de heurísticas, criadas pela própria autora como forma de refinar os SNs extraídos. As heurísticas propostas baseiam-se em análises linguísticas. Essas heurísticas propostas por Lopes (2012) ajustam, reestruturam, adicionam e eliminam SNs com vistas a deixarem os SNs mais próximos dos “conceitos” que se encontram na lista de referência para um determinado domínio, sendo nesse caso, o domínio: Pediatria. Logo abaixo, apresenta-se um quadro com cada uma dessas heurísticas:

Quadro 10– Heurísticas utilizadas para a extração de conceitos em textos em língua portuguesa.

Heurísticas de ajuste	Remoção de artigos que aparecem no início dos SNs. Essa heurística ajustaria, por exemplo, o SN: “o leite materno” para “leite materno”.
	Remoção de todos os artigos encontrados em um SN. Essa heurística remove todos os artigos que aparecem dentro de um SN. Por exemplo, o SN: “o leite da mãe” passaria a ser “leite de mãe”.
	Remoção dos pronomes que se encontram no início dos SNs. Essa heurística tem o propósito de manter o SN genérico o suficiente para ser considerado como conceito.
	Remoção de pronomes que se encontram em qualquer posição de um SN. Por exemplo, o SN: “o objetivo de seu movimento” passa a ser “objetivo de movimento”.
Heurísticas de descarte	Descarte de SNs que contêm numerais seja em forma escrita ou através de caracteres numéricos.
	Descarte de SNs que contenham símbolos , com exceção do hífen (-).
	Descarte de SNs que possuem como núcleo um “pronome” , ou seja, quando ocorre anáfora.
	Descarte de SNs que iniciam com “advérbios” . Por exemplo, “mais frequente”.
	Deteção de SNs contidos em SNs maiores através da remoção sucessiva de adjetivos, por exemplo, na frase: “Estudos realizados mostram o perigo de doenças virais hemorrágicas”, em um processo básico de extração identificaria

Heurísticas de Inclusão	somente os seguintes SNs: “Estudos realizados”, “perigo de doenças virais hemorrágicas” e “doenças virais hemorrágicas”, no entanto, com a aplicação dessa heurística seriam gerados termos adicionais pela remoção dos adjetivos (ou verbos no particípio passado) ao fim de cada termo, gerando assim: “Estudos realizados”, “Estudos”, “perigo de doenças virais hemorrágicas”, “perigo de doenças virais”, e assim sucessivamente.
	Detecção de SNs replicados pelo uso de predicado múltiplo. Por exemplo, no SN : “pacientes idosos compram e tomam remédios mais caros”. Assim os SNs “pacientes idosos” e “remédios mais caros” serão contabilizados duas vezes, pois apesar de aparecerem implícitos, eles ocorrem duas vezes.
	Detecção de estruturas gramaticais múltiplas com o uso de conjunções, quando um substantivo é qualificado por mais de um adjetivo. Por exemplo, na sentença: os pacientes idosos e obesos [...]. Têm-se dois SNs: “pacientes idosos” e “pacientes obesos”.

Fonte: Lopes (2012, p.40)

Analisando o Quadro 10, percebe-se que as heurísticas de descarte funcionam como critérios de seleção dos SNs que podem ou não representar conceitos. Por exemplo, a partir do momento que a autora descarta os SNs que contêm numerais em sua estrutura, ela está selecionando, ou seja, subte-se que os SNs com numerais não possam funcionar como conceitos para determinado domínio. Sendo assim, nesta dissertação, tomam-se as heurísticas de descarte propostas por Lopes (2012) como critérios utilizados para a seleção de SNs. Tomam-se, também, duas heurísticas de inclusão como critérios de seleção, as quais são: a heurística de detecção de SNs contidos em outros SNs maiores e a heurística de detecção de estruturas gramaticais múltiplas com o uso de conjunções quando um substantivo é qualificado por mais de um adjetivo.

Percebe-se uma relação entre a primeira e a terceira heurística de inclusão proposta por Lopes (2012) com um dos problemas suscitados por Kuramoto (1995) em sua pesquisa, onde este autor, na identificação dos SNs, se deparou com os problemas de frases que continham sequências de palavras que precediam outro conjunto de palavras coordenadas pelas conjunções “e” e “ou”, como é o caso, por exemplo, da sentença: “**o processo de negociação dos setores privado e público**”. Nesse exemplo, com a aplicação da terceira heurística de inclusão de Lopes (2012), seriam gerados SNs que se encontram embutidos em SNs maiores.

Outra relação percebida diz respeito a uma das considerações feitas por Souza (2005) ao final de sua metodologia prospectiva com a primeira heurística de inclusão de Lopes (2012). Souza (2005) verificou nos casos de SNs embutidos em outros SNs de maior nível que existia a repetição de informação pela dupla ocorrência, acabando por gerar redundância. Como forma de amenizar essa questão o autor se utilizou da eliminação automática dos SNs de menor nível que se encontravam nos de maior nível que foram selecionados. Sendo assim,

para remediar este inconveniente, este autor levou em consideração que, para frequências semelhantes, à medida que se escolhia um SN de nível 2, 3 ou 4, eliminavam-se os de nível 1, 2 ou 3 que estivessem implícitos, aninhados, respectivamente, nos de maior nível para não gerar redundância de informação. Já a primeira heurística de inclusão de Lopes (2012) propõe a detecção de SNs implícitos contidos em SNs maiores através da remoção sucessiva de adjetivos. Assim, percebe-se uma oposição em que um autor faz uso a geração de SNs implícitos e outro faz eliminação SNs como forma de evitar redundância.

Após a extração dos SNs e de feito o refinamento da seleção dos mesmos, passa-se para a segunda etapa do processo maior que é a ordenação dos termos extraídos de acordo com a relevância. Para essa atividade a autora faz uso de um novo índice, proposto por ela mesma, que leva em conta a **frequência do termo**, mas também a **frequência de disjunção de corpora**. Esse índice é denominado *tf-dcf* (*term frequency, disjoint corpora frequency*). O referido índice baseia-se na indicação primária de relevância do termo (ocorrência no *corpus* de domínio) e de um mecanismo de recompensa/penalização (devido a ocorrência em *corpora* contrastantes). O propósito do índice *tf-dcf* é considerar a frequência absoluta do termo como indicação de importância do termo, em contrapartida, escolhe-se penalizar termos que aparece nos *corpora* contrastante, dividindo a frequência absoluta do termo no *corpus* de domínio pela composição geométrica da sua frequência absoluta em cada um dos *corpora* contrastantes.

Depois de ordenados de acordo com o *tf-dcf*, passa-se para a terceira etapa que é a identificação dos termos que funcionem como conceito para o domínio representado. Lopes (2012) faz uso de um ponto de corte híbrido para selecionar os SNs que funcionem como “conceitos”. Esse ponto de corte híbrido constitui-se de um ponto de corte por limiar e um ponto de corte relativo. Assim, os termos que não forem descartados por esse ponto de corte são considerados “conceitos” do domínio. De forma sucinta, o primeiro passo desse método híbrido foi descartar termos que tinham um índice *tf-dcf* inferior a 2, descartando assim uma grande quantidade de termos. A segunda etapa do método híbrido foi o descarte de termos por um ponto de corte relativo, onde, aplicou-se um ponto de corte intermediário, entre 13% e 18%, de 15% a ser aplicado a todas as listas de termos extraídos. Como resultado da aplicação desse ponto de corte, tem-se uma lista reduzida de termos que são considerados conceitos (LOPES, 2012).

Com os termos selecionados, passa-se para a quarta e última etapa do processo proposto por Lopes (2012), a qual se constitui das diversas possibilidades de aplicações com esses termos (conceitos) de domínio específico, como, por exemplo, “geração de listas de

termos, conceitos” que por sua vez contribuem efetivamente para a criação, por exemplo, de ontologias, glossários etc.

Apesar da pesquisa de Lopes (2012) não estar diretamente relacionada com o uso de SNs como fontes de indexação automática de documentos digitais, percebe-se contribuições pertinentes da pesquisadora para essa área. Alguns pontos levantados pela pesquisadora contribuem efetivamente para pesquisas que trabalham com a extração de SNs, como, por exemplo, as heurísticas propostas. Por exemplo, a heurística que exclui os numerais do SN pode contribuir diretamente evitando que sejam extraídos “numerais” que, em alguns casos, pouco contribui para a representação de um determinado documento. Tem-se também as heurísticas que evitam que sejam extraídos símbolos que, em alguns casos, em alguns domínios, em nada contribuem conceitualmente para a representação de um documento digital.

Torna-se evidente que algumas heurísticas não se aplicariam para o contexto da seleção de SNs para indexação e recuperação, como, por exemplo, a heurística de remoção dos artigos no início dos SNs. Talvez, esta heurística seja viável nesse contexto de Lopes (2012) que era justamente tornar SNs mais genéricos para funcionarem como conceitos de um domínio, no entanto a remoção de artigos não seria tão viável no contexto da indexação automática de textos, pois tiraria a especificidade de alguns termos para um determinado documento, o que não é interessante em alguns casos.

Em relação à metodologia de seleção de SNs representativos e potenciais conceitualmente, a pesquisa proposta por Lopes (2012), no segundo e no terceiro momento, respectivamente, ordenação de termos de acordo com sua relevância e identificação de conceitos, mostra-se pertinente para o foco deste trabalho. Na definição da relevância dos SNs, a autora se utiliza de dois critérios: a frequência do termo no documento e a disjunção desse termo no *corpus* de domínio em relação aos *corpora contrastante*. Esses critérios constituem o índice denominado *tf-dcf* (*term frequency, disjoint corpora frequency*). Esse índice proposto pela autora, assim como em outras iniciativas por outros autores, segue a linha de abordagens recentes, onde se leva em consideração a frequência do termo, mas também a sua disjunção no *corpus* de domínio em relação aos *corpora contrastante*.

Um estudo mais recente e que se relaciona diretamente com a temática proposta por esta pesquisa é o estudo de Souza e Raghavan (2014), o qual apresenta uma abordagem para extrair Sintagmas Nominais a partir de textos, levando em consideração a semântica intrínseca do texto.

Este estudo de Souza e Raghavan (2014) baseia-se na metodologia de seleção de SNs publicada por esses mesmos autores em outro trabalho (SOUZA; RAGHAVAN, 2006). Esse estudo mais recente apresenta algumas mudanças incorporadas à metodologia de seleção, as quais foram sugeridas por outros pesquisadores. Assim, a metodologia proposta nesse trabalho de Souza e Raghavan (2014) foi testada em experimentos envolvendo textos em Português utilizados por Souza e Raghavan (2006).

Foram considerados, para cálculo do *score* de cada SN, três critérios: a frequência de ocorrência do SN no documento, a frequência inversa do SN nos documentos e o valor do SN de acordo com a CNP, classificação esta baseada no nível e na estrutura do SN, a qual já foi vista nesta revisão. A pontuação para cada SN foi calculada usando a fórmula desenvolvida por Souza (2005).

Souza e Raghavan (2006) receberam comentários relacionados com a arbitrariedade de certas constantes usadas no método de atribuição de pesos aos SNs. Assim, esta pesquisa de Souza e Raghavan (2014) concebe algumas sugestões recomendadas por outros pesquisadores ao trabalho anterior.

Foi utilizado o mesmo *corpus* de documentos utilizado por Souza (2005) e Souza e Raghavan (2006) compreendido por 60 artigos de revistas científicas da área de ciência da informação. Foram utilizados o PALAVRAS e o PALAVRAS xtractor para processar os textos em linguagem natural.

As sugestões recomendadas para Souza e Raghavan (2006) giravam em torno do fato de que um fator que afeta a frequência de ocorrência de um SN em um documento é o comprimento do texto. É bem provável que um SN possa ocorrer com maior frequência em um documento que é mais extenso do que em outro documento que é bem menor. Sendo assim, ao se considerar a frequência absoluta de ocorrência de um SN para calcular sua pontuação poderia conduzir a distorções. Reconhecendo isso foram feitas sugestões para melhorar a metodologia, utilizando outros dados como, por exemplo, a frequência inversa do documento e a frequência normalizada.

Foram experimentados quatro métodos diferentes de cálculo e atribuição de pesos aos SNs: Levando em consideração a frequência normalizada, a frequência normalizada e a frequência inversa do termo no documento; Levando em consideração a frequência

normalizada, a frequência inversa e o fator CNP; e levando em consideração a função **Okapi BM25**¹⁴.

A frequência normalizada tem o intuito de corrigir qualquer distorção introduzida pelo comprimento do documento. A medida normalizada é obtida pela divisão da frequência absoluta do SN no documento pelo número total de SNs ocorridos no documento.

Os SNs selecionados por cada método foram examinados manualmente quanto ao seu valor como descritor representativo da tematicidade dos documentos. A qualidade do SN foi verificada levando em consideração a existência do SN no conjunto de palavras-chave atribuídas pelo autor do documento. Os SNs foram classificados como altamente relevantes, razoavelmente relevantes, moderadamente relevantes e não relevantes.

Percebeu-se que o método com base nos três fatores, frequência normalizada, frequência inversa e fator CNP, obteve melhores resultados em termos de sua capacidade de produzir SNs de qualidade.

Os resultados mostraram que anexar pesos aos SNs, levando em consideração a frequência por si só, não leva a resultados satisfatórios. A situação melhora um pouco quando se utiliza a frequência normalizada e a frequência inversa no documento para calcular o peso. Os resultados melhoraram ainda mais quando se utilizou a classificação CSN (Souza, 2005). A utilização da função Okapi parece resultar em um grande número de SNs selecionados. A principal razão para isso é que muitos dos SNs receberam a mesma pontuação e todos os SNs que receberam a mesma pontuação foram considerados. No entanto, uma comparação dos resultados obtidos nestes experimentos com os obtidos na pesquisa anterior (SOUZA e RAGHAVAN, 2006) indica que a metodologia anterior rendeu melhores resultados. Um dos fatores que contribui para isso poderia estar relacionado com o fato de que uma lista de *stopwords* foi utilizada para suprimir os SNs menos relevantes. (SOUZA e RAGHAVAN, 2014). A utilização da classificação CSN, demonstrou melhores resultados, levando a se pensar em uma variação dos valores da classificação CSN para assim determinar os valores ideais da CSN. Com as alterações, verificaram-se melhores resultados com os seguintes valores:

¹⁴ Em recuperação de informação, Okapi BM25 é uma função de classificação utilizada pelos motores de busca para classificar documentos correspondentes de acordo com sua relevância para uma determinada consulta de pesquisa. Ele é baseado no quadro de recuperação probabilístico desenvolvido na década de 70 e 80. Para maior esclarecimento, consultar o endereço: http://en.wikipedia.org/wiki/Okapi_BM25.

Quadro 11 – Valores ideais das Categorias do Sintagma Nominal - CSN

CATEGORIA	ESTRUTURA E NÍVEL DO SN	VALOR CSN
1 ^a	Nível 1, estrutura (D*+N)	0,2
1b	Nível 1, qualquer estrutura, exceto (D*+N)	0,8
2	Nível 2, qualquer estrutura	1.1
3	Nível 3 qualquer estrutura	1.4
4	Nível 4, qualquer estrutura	1.2
>4	Nível 5, ou superior, qualquer estrutura	0,8

Fonte: Souza e Raghavan (2014, p. 14, Tradução nossa)

Martins (2014) desenvolveu estudo voltado para o uso dos SNs na recuperação de documentos, mais precisamente, voltado para a classificação automática de documentos digitais. Em sua tese, este autor demonstra o potencial e a viabilidade dos SNs como fontes de dados para classificação temática de documentos digitais. O referido trabalho demonstra como os SNs podem ser trabalhados para que os mesmos sejam capazes de servirem como recursos para uma classificação temática de documentos mais precisa e mais correta. Martins (2014) desenvolve sua pesquisa em dois momentos, um prospectivo e um consolidado.

Na primeira etapa (escolha dos documentos) da pesquisa de Martins (2014) utilizaram-se três artigos sobre Inteligência Artificial. Em seguida tratou-se da preparação do *corpus*, onde foram resolvidos problemas com formatos dos documentos, uso de imagens e tabelas, etc. Assim, optou-se pelo o uso uniforme dos documentos no formato txt. Em seguida, trabalhou-se com a filtragem preliminar do conteúdo, onde foram removidas algumas expressões recorrentes na literatura científica, mas que nada contribuem para a qualidade dos SNs, como, por exemplo, a remoção das palavras: resumo, abstract, introdução, referencias, entre outras. Em seguida, passou-se para a extração dos SNs, onde se utilizou o PALAVRAS para a identificação e marcação dos SNs, e para extração utilizou-se um script fornecido pelo laboratório VISL, escrito na linguagem PERL. Depois de extraídos, os SNs passaram por uma “limpeza”, onde foram removidos os excessos de espaços entre palavras, os asteriscos, as aspas, os sinais matemáticos e outras estruturas que pudessem vir a afetar a qualidade do Sintagma no processo de classificação. Depois de limpos, os SNs passaram por um processo de identificação de sintagmas únicos, tendo em vista que vários SNs se repetiam ao longo dos documentos (MARTINS, 2014).

Depois das etapas mencionadas anteriormente, passou-se para a aplicação do método preliminar, onde foi feita a comparação direta entre os sintagmas dos documentos, essa comparação levou em consideração todos os sintagmas encontrados nos três primeiros documentos, independentemente da quantidade de vezes, assim não se utilizou nenhum

método de ponderação de sintagmas. Após a comparação direta entre os sintagmas, compararam-se os SNs dos documentos após o processo de remoção dos quantificadores. Essa análise foi feita de forma manual e pontual em cada sintagma modificado, tendo como objetivo avaliar o quanto a mudança na estrutura do sintagma pode impactar na comparação (MARTINS, 2014).

A primeira proposta para deixar os documentos mais semelhantes foi extrair os quantificadores encontrados nos sintagmas detectados, como, por exemplo, “o, a, os, as, etc”. Em síntese, essa remoção dos quantificadores aparentemente não gerou melhora na semelhança dos sintagmas e, assim, não foi necessariamente utilizada na etapa consolidada. Por fim, compararam-se os SNs dos documentos após um processo de *stemming*. Repetiram-se todos os testes mencionados anteriormente após o processo de *stemming* dos sintagmas. Ao utilizar o sintagma após esse processo, o grande ganho foi em relação ao uso de palavras no plural e no singular e, em alguns casos, no tempo verbal. Mas outros ganhos, como uma redução da ordem de 30% do tamanho dos documentos, também foram importantes, ao considerar questões ligadas ao custo computacional de processamento. (MARTINS, 2014). Por meio do processo de *stemming*, verificou-se um aumento perceptível na comparação de SNs semelhantes. Após a comparação com o uso do *stemming*, passou-se para a comparação após a convergência de sinônimos.

Foi realizado acesso automático à lista de sinônimos mantida pelo site <sinônimos.com.br>. Após a convergência de sinônimos realizada, percebeu-se pouca melhora na semelhança entre os sintagmas dos documentos comparados. A partir de uma análise manual, foi comprovado que diversos fatores interferem nesse processamento. O primeiro fator observado é o não tratamento, pelo software de dicionário utilizado, de palavras no plural ou no singular e também a não identificação de gêneros diferentes. Por exemplo, a palavra: “grandes” não foi identificada no dicionário, mas a palavra “grande” foi. (MARTINS, 2014).

Depois das etapas e dos testes (comparações) realizados, foi feita uma contraprova, tendo em vista a validação dos resultados encontrados. Nessa contraprova foram utilizados três documentos de outros domínios na comparação com os três de inteligência artificial, os quais já foram mencionados anteriormente.

A metodologia consolidada fez uso dos resultados alcançados na prospectiva, tendo em vista se utilizar dos resultados positivos alcançados na classificação temática de documentos digitais.

Em suma, os resultados alcançados pela pesquisa de Martins (2014) mostraram que o processo de *stemming* do sintagma, no lugar do próprio sintagma, proporciona vantagens em relação ao uso somente do sintagma no momento de treinar a máquina. Verificou-se que o processo de classificação automática alcançou um percentual de 100% de classificações corretas em relação a outras pesquisas de mesmo gênero, o que demonstra a viabilidade desse processo para sistemas de classificação automática.

Em relação à seleção de SNs, percebem-se algumas contribuições da pesquisa de Martins (2014) para este trabalho de dissertação. No que se refere à frequência de ocorrência de SNs, e sabendo que essa métrica é comumente utilizada para a seleção de SNs que possam funcionar como descritores, Martins (2014), apesar de não se utilizar dessa medida em sua pesquisa, ressalta a sua importância ao dizer que

Uma métrica importante é a quantidade de vezes em que determinados sintagmas ocorrem no corpo de um documento. Essa informação pode ser relevante para sistemas de classificação, já que sintagmas que se repetem ao longo de um ou mais documentos de um *corpus* podem frequentemente se tornar um descritor importante nesse conjunto. (MARTINS, 2014, p. 73)

Não levando a uma contradição, é pertinente salientar que, apesar de não levar em consideração a frequência de ocorrência dos SNs, nem ponderação e nem pontuação dos SNs, Martins (2014) contabiliza a frequência de ocorrência dos SNs extraídos, com vistas a demonstrar a importância dessa métrica para atividades de indexação, recuperação e, também, classificação de documentos. Nessa contabilização, o referido autor se utiliza dos SNs que apareceram pelo menos sete vezes no documento. Esse limiar foi escolhido após a observação de que o número de repetições para o sintagma que mais aparece no documento, em relação ao segundo, tem uma queda abrupta. O mesmo acontece até o sétimo sintagma que mais aparece. A partir desse ponto, existe uma tendência em variar apenas minimamente o aparecimento do próximo sintagma, em relação aos anteriores (que apareceram mais vezes).

Concluindo, segue-se mais adiante com o Quadro 12, o qual é composto pelas pesquisas analisadas nesta subseção, onde são sistematizados os critérios utilizados por cada autor para alcançar uma seleção mais consistente de SNs, ressaltando que, apesar de algumas pesquisas citadas nesta subseção não estarem relacionadas diretamente com o uso dos SNs para indexação e recuperação, fez-se pertinente mencioná-las.

Quadro 12 – Síntese de critérios utilizados por pesquisadores para se selecionar SNs.

Autores	Foco da pesquisa	Critérios utilizados na seleção de SNs
Souza (2005)	Propõe uma metodologia para indexação de documentos digitalizados de textos completos por meio da identificação, extração e seleção de SNs.	<ul style="list-style-type: none"> • Frequência de ocorrência dos SNs no texto do documento. • Inverso da frequência de ocorrência dos SNs no conjunto de documentos do <i>corpus</i> • Classificação CNP, nível e estrutura do SN. • Ocorrência do SN em tesouro da área (usado como fator primordial somente na metodologia prospectiva); • Eliminar SNs em <i>stoplist</i> de SNs não descritores.
Souza, Alvarenga Neto e Mendes (2007)	Tem como foco uma proposta analítica quádrupla de análise e compreensão de fenômenos sociais, apoiando-se no desenho de um mapa terminológico-conceitual. O objeto de estudo desse artigo foi a área “Gestão do Conhecimento”.	<ul style="list-style-type: none"> • Frequência de ocorrência dos SNs no texto/documento. • Frequência inversa de ocorrência dos SNs no conjunto de documentos do <i>corpus</i> (IDF) • Estruturas e níveis dos SNs.
Maia(2008)	Foca no uso dos SNs pontuados (pontuação de Souza, 2005) como recursos de classificação por similaridade e aglomerados de documentos eletrônicos.	<ul style="list-style-type: none"> • Frequência de ocorrência dos SNs no texto do documento • Inverso da frequência de ocorrência dos SNs no conjunto de documentos do <i>corpus</i>; • Classificação CNP, nível e estrutura do SN. <p>Obs.: o Software proposto por Maia (2008) realiza a identificação da classe gramatical do Sintagma, bem como o cálculo de pontuação do mesmo como descritor de forma automática baseando-se no método proposto por Souza (2005).</p>
Lopes (2012)	Extração de conceitos relevantes, a partir de SNs, para um determinado domínio, com vistas a execução de diversos recursos linguísticos, como, por exemplo, criação de “ontologias”	<ul style="list-style-type: none"> • Frequência de ocorrência de SNs no texto/documento. • Frequência de ocorrência de SNs em outros textos/documentos de domínios diferentes. (domínio contrastante) <p>Obs.: esses dois critérios constituem o <i>índice tf-idf</i>, criado pela própria autora.</p> <ul style="list-style-type: none"> • Heurísticas de descarte: descarte dos SNs que contêm numerais; descarte dos SNs que contenham símbolos; descarte dos SNs que possuem como núcleo um pronome; e descarte dos SNs que iniciam com advérbios. • Heurísticas de inclusão: detecção de SNs contidos em SNs maiores através da remoção sucessiva de adjetivos, detecção de estruturas gramaticais múltiplas com o uso de conjunções, quando um substantivo é qualificado por mais de um adjetivo.
Souza e Raghavan (2014)	Apresentam abordagens para a extração e seleção de SNs a partir de textos com base na semântica intrínseca do texto.	<ul style="list-style-type: none"> • Frequência de ocorrência absoluta do SN no documento. • Frequência de ocorrência normalizada/relativa do SN em um documento. • Frequência inversa de ocorrência do SN nos documentos. • Classificação CNP, nível e estrutura do SN. (SOUZA, 2005)
	Apresenta um estudo aprofundado	<ul style="list-style-type: none"> • Frequência de ocorrência dos SNS em um

<p>Martins (2014)</p>	<p>acerca dos Sintagmas Nominais e como estes podem ser utilizados na classificação automática de documentos levando em consideração o assunto tratado em cada documento, apoiando-se no Processamento de Linguagem Natural.</p>	<p>documento levando em consideração os SNs que ocorrem pelo menos sete (7) vezes no documento.</p> <p>Obs.: Martins realizou uma espécie de limpeza dos SNs, retirando expressões comuns em textos científicos do seu corpus trabalhado, onde eliminou palavras como: resumo, abstract, introdução, etc. Pode-se entender essa atividade como a utilização de uma lista de stopwords/stoplist, que no caso foi criada pelo próprio pesquisador durante a análise dos SNs.</p>
------------------------------	--	--

Fonte: desenvolvido pelo autor

Com base na síntese das pesquisas resumidas no quadro acima, percebe-se que são variados os critérios que os pesquisadores se utilizaram para a seleção de SNs, seja para a classificação automática de documentos, ou indexação, ou extração de conceitos, ou criação de ontologias, entre outros objetivos. Todavia, verifica-se que todos os critérios podem ser viáveis para a seleção de SNs para fins de indexação e recuperação, uma vez que são critérios que refinam os SNs, excluindo os que são provavelmente menos relevantes e selecionando outros que possuem um poder discriminante maior em relação a outros SNs, e, que assim possam ser utilizados como descritores documentais para fins de indexação e recuperação de documentos.

Encontra-se no Quadro 13, organizados sistematicamente, os critérios utilizados pelas pesquisas citadas no quadro anterior, ressaltando o escopo de aplicação de cada critério, bem como os autores que fizeram uso desses critérios, proporcionando uma melhor visualização do âmbito de aplicação de cada critério e quais autores fizeram uso desses critérios.

O escopo de aplicação de cada critério foi organizado em categorias, as quais são: SN, DOCUMENTO, *CORPUS* e *CORPUS CONTRASTANTE*. Essa categorização permitiu perceber em que contexto cada critério era aplicado, onde alguns critérios, como, por exemplo, a FREQUENCIA ABSOLUTA, são analisados, levando em conta apenas o próprio documento. Os diversos critérios de seleção são aplicáveis em diferentes escopos, no entanto, o escopo de aplicação que prevalece é o próprio SN, como pode ser visto no quadro 13.

Quadro 13 – Critérios de seleção, escopo de aplicação dos critérios e trabalhos que utilizaram um ou outro critério de seleção.

Critérios	Escopo de aplicação do Critério. Categorias: SN DOCUMENTO CORPUS CORPUS/CONTRASTANTE	Trabalhos que fizeram uso do critério de seleção
Frequência de ocorrência dos SNs no texto/documento.	DOCUMENTO	Souza (2005), Souza, Alvarenga Neto e Mendes (2007), Maia (2008), Lopes (2012), Souza e Raghavan (2014) e Martins (2014).
Frequência de ocorrência dos SNs no conjunto de documentos do corpus (IDF)	CORPUS	Souza (2005), Souza, Alvarenga Neto e Mendes (2007), Maia (2008) e Souza e Raghavan (2014).
Estruturas e níveis dos SNs.	SN	Souza (2005), Souza, Alvarenga Neto e Mendes (2007), Maia (2008) e Souza e Raghavan (2014).
Verificar ocorrência do SN em tesouro da área	SN	Souza (2005)
Eliminar SNs em <i>stoplist</i> de SNs não descritores	SN	Souza (2005) Martins(2014)
Frequência de ocorrência de SNs em outros textos/documentos de domínios diferentes. (<i>corpus contrastante</i>)	CORPUS/CONTRASTANTE	Lopes (2012)
Descarte dos SNs que contêm numerais;	SN	Lopes (2012)
Descarte dos SNs que possuem como núcleo um pronome;	SN	Lopes (2012)
Descarte dos SNs que iniciam com advérbios.	SN	Lopes (2012)
Deteção de SNs contidos em SNs maiores através da remoção sucessiva de adjetivos,	SN	Lopes (2012)
Deteção de	SN	Lopes (2012)

estruturas gramaticais múltiplas com o uso de conjunções, quando um substantivo é qualificado por mais de um adjetivo.		
Frequência de ocorrência normalizada do SN em um documento	SN	Souza e Raghavan (2014)

Fonte: desenvolvido pelo autor.

Verificou-se com base nos estudos analisados que a frequência de ocorrência é um dado comumente utilizado para a determinação do SN como descritor ou não, como pôde ser visto nas pesquisas da maioria dos autores. Verificou-se, também, que esse critério é fundamental para a seleção de SNs relevantes para descrição de conteúdos de documentos. Essa frequência de ocorrência demonstra em alguns casos o potencial informativo de um determinado SN, bem como também demonstra o caráter genérico de um determinado SN quando esse aparece vultosamente em vários documentos. (frequência absoluta de ocorrência e frequência inversa de documentos).

Ainda sobre a frequência de ocorrência de termo, vale salientar a pertinência da frequência normalizada e não somente a frequência absoluta, como pôde ser vista na pesquisa de Souza e Raghavan (2014), onde a ocorrência de frequência de um SN pode ser bem maior em um documento que seja mais extenso em comparação com outro bem menor. Souza e Raghavan (2014) constataram que a frequência normalizada alcançou bons resultados, o que demonstra que é um critério que deve ser levando em consideração quando da seleção de SNs.

No tocante a estrutura dos SNs, Souza (2005) evidenciou a importância de se levar em conta, paralelamente à frequência de ocorrência, a estrutura dos SNs, uma vez que SNs de nível 1, nível 2, SNs complexos possuem potencialidades diferentes, como, por exemplo, um SN complexo pode muitas vezes não ser conciso, objetivo. Este autor propõe uma tabela com diferentes pontuações para os Sintagmas com diferentes estruturas em suas formações. Propõe também uma fórmula para cálculo da pontuação de cada SN, pontuação essa que discriminará SNs mais relevantes ou não.

Em relação ao posicionamento dos SNs, Corrêa et al. (2011) levantaram a seguinte indagação: “Será que não seria o caso de se levar em consideração o posicionamento dos SNs do mesmo modo que se verifica nos sistemas de indexação baseados em “palavras isoladas”?”. Apesar de essa questão necessitar de mais estudo, pois a mesma ainda não foi

utilizada por nenhuma pesquisa como critério de seleção, acredita-se que a utilização do elemento “posicionamento” contribuirá para a indexação automática por meio de SNs do mesmo modo que na indexação manual, na qual são levadas em consideração as partes mais informativas dos documentos.

Lopes (2012) também possui sua contribuição, uma vez que além de fazer uso de alguns critérios já utilizados por outros autores, como já foi mencionado, se utiliza do critério que penaliza os SNs que ocorrem com frequência em outros documentos de outros domínios, ou seja, corpus contrastante, julgando-os como SNs que possuem pouca carga conceitual, ou seja, não são SNs relevantes.

Martins (2014), a título de ilustração e complementação, em sua tese, contabiliza a frequência de ocorrência dos SNs extraídos, com vistas a demonstrar a importância dessa métrica para atividades de indexação, recuperação, e, também, classificação de documentos. Nessa contabilização, o referido autor se utiliza dos SNs que apareceram pelo menos sete vezes no documento. Esse limiar foi escolhido após a observação de que o número de repetições para o sintagma que mais aparecia no documento, em relação ao segundo, tinha uma queda abrupta. O mesmo acontecia até o sétimo sintagma que mais aparecia. A partir desse ponto, o autor percebeu que existia uma tendência em variar apenas minimamente o aparecimento do próximo sintagma, em relação aos anteriores (que apareciam mais vezes).

Em suma, conclui-se que, seja qual for a fórmula ou o cálculo ou o modo de analisar e selecionar SNs com valor de descritores, o que se torna precípuo é a devida atenção que deve ser dada à seleção de SNs para fins de indexação e recuperação, tendo em vista que apenas a extração de SNs não é suficiente para que se tenha SNs representativos de conteúdos documentais. Vários são os critérios que podem ser aplicados à escolha de SNs com valor de Descritores, como pôde ser visto nesta revisão. No entanto, necessita-se ainda de mais estudos que se voltem especificamente para esta seleção, analisando detidamente a aplicabilidade e a eficácia de cada um dos critérios de seleção em ambientes de indexação e recuperação de informações digitais, bem como também a aplicação de múltiplos critérios de seleção como pôde ser visto em algumas pesquisas citadas nesta revisão.

2.1.2.1.2 Avaliação da Indexação Automática e da extração e seleção de SNs

Em relação à avaliação da indexação automática, os dois métodos aplicados à indexação manual, mencionados na subseção 2.1.1.3 Qualidade e Avaliação da Indexação,

também, são aplicáveis à indexação realizada de forma automática. Assim, existem medidas que são aplicadas com o intuito de comparar a indexação manual com a automática, como é o caso da pesquisa realizada por Bastos (1984 apud CÂMARA JÚNIOR, 2007, p. 43), onde a pesquisadora se utilizou da seguinte fórmula para comparar as indexações:

$$q = \frac{c}{a + m - c}$$

Com base nessa fórmula, o (q) é o coeficiente de avaliação (comparativo entre as duas indexações), (c) é a quantidade de termos extraídos pelos dois sistemas que são comuns, (a) é a quantidade de termos selecionados pelo processamento automático e (m) é a quantidade de termos selecionados pelo processamento manual. Esse coeficiente muda de valores entre 0 e 1, onde 0 significa que não coincidiu nenhum índice e 1 significa que todos os índices selecionados automaticamente são idênticos aos selecionados manualmente. A multiplicação desse resultado por cem (100) resulta na porcentagem de aderência entre uma indexação e outra. (CÂMARA JÚNIOR, 2007, p. 43)

Mais adiante, é apresentado alguns métodos utilizados para a avaliação da extração e seleção de SNs

Um ponto importante que se torna evidente nos processos desenvolvidos tanto para a extração como para a seleção de SNs diz respeito ao fato de que essas atividades são, devido a natureza do objetivo de estudo, empíricas. Sendo assim, é essencial que os pesquisadores definam as formas pelas quais eles se utilizaram para verificar a qualidade do processo, do método ou do sistema proposto. Nesse sentido, esta subseção dedica-se a expor as formas pelas quais os pesquisadores, supracitados na subseção anterior, bem como outros que não tiveram suas pesquisas sintetizadas anteriormente, avaliaram seus resultados, seus processos e sistemas desenvolvidos.

Kuramoto (1995-2002) fez uso de um índice de equivalência. São utilizados índices de equivalência também por Morellato (2010) e Miorelli (2001). Esta última se utilizou, basicamente, de três mecanismos para a avaliação. Foram eles: a extração manual dos SNs, a utilização do PALAVRAS, de Bick (2001), o qual possibilita a geração de uma árvore sintática de sentenças, demonstrando os SNs, e a utilização de um exemplo utilizado na pesquisa de Vieira et al. (2000).

Souza (2005) se utiliza das palavras-chave para a avaliação de sua metodologia como um todo, pois ele faz o cotejamento dos SNs selecionados com as palavras-chave e resumos

dos autores com o intuito de verificar a potencialidade dos SNs como recursos informacionais para representação de documentos. Essa análise baseou-se no julgamento realizado pelo próprio autor, o qual se utilizou de algumas ferramentas matemáticas para possibilitar a avaliação e a comparação.

Em relação à metodologia de avaliação do método proposto como um todo, Maia (2008) se utiliza de outros softwares (ED-CER E VISL) para avaliar sua ferramenta (o OGMA) em seus experimentos, inclusive como forma de aprimorar e ajustar sua ferramenta.

Corrêa et al. (2011), depois de submeterem seu *corpus* ao OGMA, fazem uma avaliação manual dos resultados dessa ferramenta analisando-os sob quatro perspectivas, são elas: se os SNs extraídos pelo OGMA são de fato SNs; a quantidade de SNs relevantes; os que contêm as palavras-chave; e os que são idênticos às palavras-chave dos textos submetidos ao OGMA. Esse autor fez uso das seguintes métricas: cálculo e análise dos percentuais de precisão em extrair SNs tidos como descritores; a taxa de erro ao extrair conjuntos de caracteres que não constituíam SNs e o percentual de SNs extraídos, mas que pouco contribuem para a representação do conteúdo informacional dos documentos; e precisão e abrangência.

De maneira geral, Corrêa et al. (2011) verificaram que a precisão alcançada pela ferramenta OGMA gira em torno de 50%, demonstrando que uma avaliação por parte dos profissionais indexadores seria exaustiva, pois os mesmos teriam que avaliar todos os possíveis SNs extraídos pelo software, sendo que apenas metade seria válida para a indexação e recuperação, mas, em contra partida, os autores verificaram que a qualidade dos SNs relevantes extraídos são bem pertinentes, e, além disso, verificaram também que cerca de 80% das palavras-chave atribuídas aos resumos se encontravam extraídas como SNs pelo OGMA.

No que se refere à avaliação de todo o processo proposto em sua pesquisa, Lopes (2012) fez uso de uma lista de termos relevantes do domínio previamente estabelecida como referência para o sucesso do processo (*Gold Standard*), possibilitando assim o cotejamento das listas com os termos extraídos com a lista com os termos de referência. Foram utilizados índices de precisão, abrangência e medida F. Essas medidas são utilizadas para comparar dois conjuntos, nesse caso, duas listas de termos. A *precisão* expressa a razão entre o número de termos da lista de referência que foram extraídos e considerados. A abrangência expressa a razão entre o número de termos da lista de extraídos e considerados presente na lista de referência e o tamanho da lista de referência. Por fim, a medida F expressa o equilíbrio entre

os valores de precisão e abrangência. Os valores da medida F (*inglês, f- measure*) são valores situados entre P e R (*Precision e Recall*), e quanto maior for a diferença entre esses valores, mais próxima a medida F será do menor valor entre eles.

Formula da medida F:

$$F = \frac{2 \times P \times R}{P + R}$$

A medida F demonstra o equilíbrio entre os valores de precisão e revocação. A sua representação numérica é a média harmônica entre os valores de precisão e revocação. Os índices de abrangência, precisão e medida F são medidas de qualidades utilizadas para comparar dois conjuntos, como, por exemplo, duas listas termos, ou sintagmas. Uma dessas listas será tida como referência, e a outra será comparada com essa lista de referência.

As avaliações feitas pelos pesquisadores acerca das extrações e seleções dos SNs, bem como de todo o processo ou método proposto são diretamente relacionadas com o contexto de cada pesquisa, tendo em vista que diferentes objetivos, metodologias e ambientes da pesquisa influenciam nos métodos e formas escolhidos para se avaliar um processo ou sistema.

Em suma, os métodos comumente utilizados pelos pesquisadores para avaliarem suas pesquisas, seus resultados, seus sistemas e produtos baseiam-se em: Avaliação manual pelos próprios pesquisadores; Submissão dos resultados das pesquisas a outras ferramentas/sistemas, Comparação dos resultados alcançados por meio de recursos automáticos com resultados alcançados manualmente e Comparação dos resultados com as palavras-chave contidas nos documentos.

Passa-se agora para a subseção “informação jurídica”, a qual se faz pertinente devido ao *corpus* a ser utilizado na parte empírica da presente pesquisa.

2.2 Informação Jurídica

Analisa-se nesta subseção algumas características da “Informação Jurídica”, como forma de compreender um pouco as fontes de informação jurídica, dentre as quais uma categoria específica, a qual se constitui o *corpus* desta pesquisa que é a “Informação Doutrinária”. Dentro dessa breve síntese, aborda-se também um pouco da terminologia desse domínio específico.

Inicia-se essa breve revisão com o conceito de “Informação”, que para Le Coadic (2004, p. 4) a “Informação é um conhecimento inscrito (registrado) em forma escrita

(impresa ou digital), oral ou audiovisual em um suporte.”. A palavra Direito advém do latim *Directum*, significando “muito reto”, “muito justo” (SILVA, 2010).

O Direito se manifesta através das normas jurídicas, cujas fontes são categorizadas em materiais e formais (BRAZ, 2013). De modo geral, as fontes materiais são constituídas pelos fatos sociais, pelos acontecimentos sociais, costumes etc. Já as fontes formais podem ser entendidas como a materialização de conteúdos ligados às fontes materiais. Assim, as fontes formais jurídicas podem ser apresentadas através de três modos: a legislação, a jurisprudência e a doutrina.

A “informação jurídica” se apresenta em uma das três fontes citadas anteriormente, as quais serão detalhadas a seguir:

- Fontes legislativas – são as fontes formais que atuam como instrumentos normativos aplicáveis a uma determinada região, um território. Estas fontes são criadas pela União, Estados ou Municípios.
- Fontes Jurisprudenciais – São as fontes que atuam como recursos essenciais para a constituição de sustentação e defesa de teses jurídicas. São exemplos das fontes jurisprudenciais: Decisões, Acórdãos, Sentenças.
- Fontes Doutrinárias – são as fontes que são elaboradas por profissionais especializados em determinada área do direito, como, por exemplo, professores, pesquisadores, profissionais especializados, dentre outros. As doutrinas são opiniões desses profissionais embasadas em conhecimentos científicos ou empíricos.

Um ponto relevante colocado por Braz (2013) em relação à informação jurídica é o seu caráter duradouro. A longevidade da informação jurídica é especial em relação a outros domínios científicos, pois a vida útil de determinadas fontes jurídicas, mesmo quando revogadas ainda perduram, ainda que com um caráter consultivo, mas continuam a servir aos profissionais jurídicos.

Todo Domínio Científico possui suas características, suas peculiaridades e suas nuances. A área jurídica possui uma variedade de fontes de informação, às quais devem ser dada atenção por parte dos sistemas de informação, proporcionando um tratamento adequado para cada categoria de fonte.

Silva (2010, p. 52) em relação à terminologia da área jurídica diz que “O Direito como as demais disciplinas, pertencentes ao conhecimento humano, possui uma terminologia própria que desenvolve conceitos e termos, com definições que são aplicáveis às

particularidades de suas ramificações e funções”. Torna-se perceptível a peculiaridade da área jurídica no que diz respeito a sua terminologia, a qual se torna visível nas documentações jurídicas.

Apesar de ser perceptível a terminologia “particular” da área jurídica, como todo campo do saber, verifica-se também, como em outros domínios, uma apropriação de termos comuns (cotidianos) para a área jurídica, além dos seus termos essencialmente jurídicos, como pode ser visto no quadro abaixo elaborado por Maciel (2009, p.22) inspirada em Cornu (1990 p. 61-87):

Quadro 14 – Termos Jurídicos na Linguagem especializada e na língua comum

Termos jurídicos por excelência	Termos jurídicos de dupla pertinência		
	Termos jurídicos na língua comum	Língua comum + sentido jurídico	Termos com o sentido comum e implicações legais
<i>anticrese, decujus, exequatur, exeqüenda, enfiteuse, sucumbência, habeas data.</i>	<i>usofruto, herança, salvo-conduto, quorum, tribunal, juiz, crime, julgamento.</i>	<i>despejo, servidão, julgado, imposto, sentença, testador, tombamento, tempestivo, falecer.</i>	<i>pai, mãe, filho, família, parente, cônjuge, empregado, férias, salário, benefício.</i>

Fonte: Maciel (2009, p.22) baseada em Cornu (1990, p. 61 - 87)

Com base nesse quadro fica visível a distinção entre os termos essencialmente jurídicos e os que foram “apropriados” da língua comum para a linguagem jurídica. O latim é o denominador comum de várias línguas, no entanto, a prevalência do latim na terminologia jurídica é bem maior.

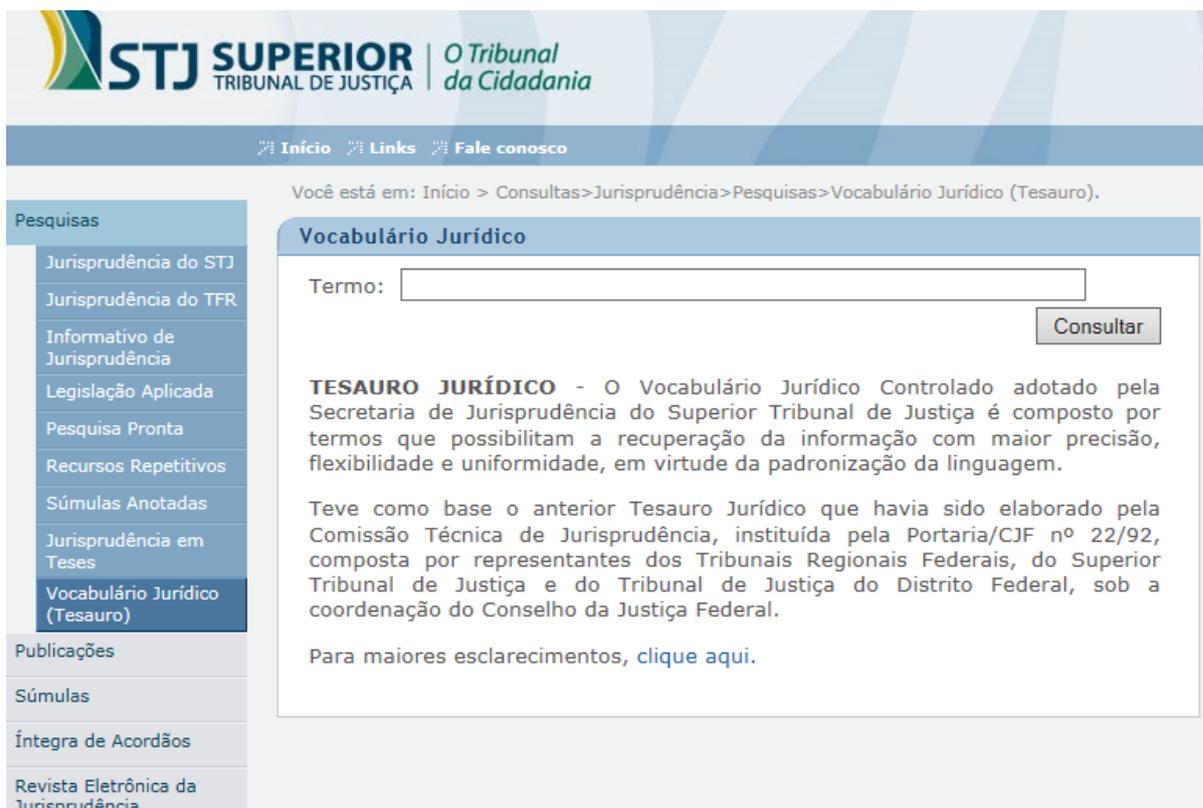
Como se pôde perceber, a área jurídica é caracterizada por conter diferentes tipologias de fontes de informação, bem como também possuir um vocabulário que abrange tanto termos técnicos especializados como também termos comuns da linguagem do cotidiano, mas que

foram apropriados para outra esfera semântica. Como toda informação, a informação jurídica precisa ser organizada e tratada para que seja disseminada e utilizada por indivíduos da sociedade. Nesse sentido, as instituições de memória (unidades de informação jurídica) precisam ater-se para questões relacionadas com as formas, os recursos, os instrumentos e as políticas necessárias para que se tenha uma informação jurídica organizada e apta a ser encontrada e utilizada.

Instrumentos essenciais para a organização da informação são as Linguagens Documentárias, aquelas vistas anteriormente nesse referencial, as quais têm como intuito proporcionar um controle terminológico de um determinado domínio, além de mediar a ligação entre a necessidade de informação de um indivíduo com a representação de um determinado documento em uma sistema de recuperação de informação. Um exemplo de linguagem Documentária é o Tesouro. Pode-se considerar o Tesouro como um conjunto de termos semântica e genericamente relacionados, cobrindo uma determinada área específica do conhecimento. É um instrumento fundamental para a indexação e recuperação de informação. (GOMES, 2007).

As instituições de memória especializadas, como, por exemplo, bibliotecas jurídicas, escritórios jurídicos, centros de documentação especializados, arquivos de tribunais, entre outras, comumente se utilizam de instrumentos de controle terminológico (tesauros) para organização das fontes de informação jurídica. A seguir, encontra-se, a título de exemplo, uma demonstração de uma busca com a expressão “legislação brasileira” no Tesouro especializado disponível online do Superior Tribunal de Justiça - Brasil:

Figura 1 - Página de busca no tesauro jurídico do site do Superior Tribunal de Justiça-Brasil.



Fonte: (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014)

A página inicial do tesauro de busca do Superior Tribunal de Justiça-Brasil, como pode ser visto na Figura 1, possui um leiaute simples e que trás várias possibilidades de buscas, como, por exemplo, Jurisprudência do STJ, Jurisprudência do TFR, entre outros.

Logo abaixo, pode-se ver que, ao inserir uma expressão de busca, o sistema já evidencia outro termo de busca por meio do **USE**, demonstrando os relacionamentos existentes entre os termos.

Figura 2 - Página de busca no tesauro jurídico do site do Superior Tribunal de Justiça-Brasil com a expressão de busca: legislação brasileira

STJ SUPERIOR TRIBUNAL DE JUSTIÇA | O Tribunal da Cidadania

Início Links Fale conosco

Você está em: Início > Consultas>Jurisprudência>Pesquisas>Vocabulário Jurídico (Tesouro).

Pesquisas

- Jurisprudência do STJ
- Jurisprudência do TRF
- Informativo de Jurisprudência
- Legislação Aplicada
- Pesquisa Pronta
- Recursos Repetitivos
- Súmulas Anotadas
- Jurisprudência em Teses
- Vocabulário Jurídico (Tesouro)

Vocabulário Jurídico

LEGISLAÇÃO BRASILEIRA

USE	LEI BRASILEIRA
CAT	CIV/01, DC/DCTC, DP/DP01, DPP/DPP02, DTR/DTR03

Termo:

Publicações

Fonte: (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014)

Na Figura 2, pode-se ver que o Tesouro já demonstra a possibilidade de usos de outros termos, como sinônimos do que foi utilizado na busca inicial.

Na figura 3, verificam-se outros relacionamentos com a expressão de busca “lei brasileira”, a qual foi recomendada pelo tesouro a se utilizar ao invés de “legislação brasileira”.

Figura 3 - Página de busca no tesouro jurídico do site do Superior Tribunal de Justiça-Brasil já com os relacionamentos dos termos do vocabulário com o termo de busca inicial

The screenshot shows the website of the Superior Tribunal de Justiça (STJ) with the logo 'STJ SUPERIOR TRIBUNAL DE JUSTIÇA | O Tribunal da Cidadania'. The navigation bar includes 'Início', 'Links', and 'Fale conosco'. The breadcrumb trail reads 'Você está em: Início > Consultas > Jurisprudência > Pesquisas > Vocabulário Jurídico (Tesouro)'. On the left, a sidebar lists search options: 'Pesquisas' (with sub-items: Jurisprudência do STJ, Jurisprudência do TFR, Informativo de Jurisprudência, Legislação Aplicada, Pesquisa Pronta, Recursos Repetitivos, Súmulas Anotadas, Jurisprudência em Teses, Vocabulário Jurídico (Tesouro)), 'Publicações', and 'Súmulas'. The main content area is titled 'Vocabulário Jurídico' and displays a search for 'LEI BRASILEIRA'. The results are as follows:

Termo	Relacionamentos
UP	LEGISLAÇÃO BRASILEIRA
UP	LEGISLAÇÃO NACIONAL
TG1	LEI
TR	JURISDIÇÃO BRASILEIRA
TR	LEI ESTRANGEIRA
TR	LICC
CAT	CIV/01, DC/DCTC, DP/DP01, DPP/DPP02, DTR/DTR03

At the bottom, there is a search input field containing 'lei brasileira' and a 'Consultar' button.

Fonte: (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014)

Com base no simples exemplos acima, percebe-se que os tesouros exercem uma função essencial na organização de determinado domínio do conhecimento, uma vez que os mesmos controlam a terminologia da área, estabelecendo relações semânticas e sintáticas entre os termos/conceitos de um domínio.

No âmbito brasileiro, várias instituições vêm reunindo esforços para o desenvolvimento de seus tesouros e suas linguagens com vistas a conseguir controlar ao máximo o vocabulário da área, bem como proporcionar um bom relacionamento com os usuários desses sistemas.

A seguir, encontra-se seção dedicada à metodologia desta pesquisa, a qual demonstra qual o percurso utilizado para compor o que até aqui se fez, bem como também o caminho percorrido para a elaboração do experimento realizado.

3 METODOLOGIA DA PESQUISA

A fundamentação teórica da metodologia desta pesquisa tem como referência os autores Gil (2002) e Gonsalves (2007), os quais explicitam os procedimentos sistemáticos dos métodos científicos, sintetizando a metodologia, as técnicas e os procedimentos do trabalho científico.

Quanto a sua natureza, esta pesquisa caracteriza-se como básica, ao passo que, pretende, com base em estudos na literatura da área e estudo de caso envolvendo experimento, gerar novos conhecimentos acerca da *indexação automática por meio de SNs*, mais especificamente, acerca da etapa de *seleção de SNs* na indexação automática de resumos de teses e dissertações da área de Direito.

Em relação à forma de abordagem do problema, essa pesquisa caracteriza-se como quantiquantitativa, uma vez que se utiliza de dados quantitativos e qualitativos acerca do processo em estudo, a *Indexação Automática por meio de SNs*. O seu caráter quantitativo fica perceptível quando se trabalha com a quantificação de SNs extraídos do texto do documento, bem como, também, a quantidade de SNs que realmente representam “Descritores documentais”. Em relação ao caráter qualitativo, Gonsalves (2007, p 68) diz que a pesquisa qualitativa preocupa-se “com a compreensão, com a interpretação do fenômeno, considerando o significado que os outros dão às suas práticas”. Assim, no presente trabalho, o caráter qualitativo evidencia-se no momento em que se faz a análise reflexiva dos dados quantitativos obtidos, proporcionando estabelecer relações de variáveis, permitindo, quando possível, uma melhor compreensão dos fenômenos e, também, o surgimento de possíveis soluções para problemas encontrados. Esse caráter qualitativo torna-se perceptível tanto na elaboração do referencial teórico do trabalho, como também na execução do experimento desenvolvido.

No tocante ao seu objetivo, o presente trabalho classifica-se como exploratório, pois como afirma Gonsalves (2007) a pesquisa exploratória caracteriza-se pelo entendimento e esclarecimento de ideias, com o intuito de fornecer uma visão mais ampla de um determinado fenômeno que não é muito explorado. A referida autora informa que esse tipo de pesquisa também é conhecido por “pesquisa de base”. De acordo com Gil (2007, p. 41), a pesquisa exploratória “têm como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses. Essas pesquisas têm como objetivo principal o aprimoramento de ideias [...]”.

Nos procedimentos, foi utilizada a pesquisa bibliográfica, uma vez que a mesma proporciona uma maior familiaridade com o tema tratado, bem como com os pesquisadores que direta e indiretamente escreveram acerca da temática proposta. Esse estudo obtém informações pertinentes ao objetivo proposto com base em um experimento realizado, em que serão obtidos alguns dados empíricos, contribuindo para o conhecimento mais aprofundado acerca da indexação automática que faz uso dos SNs como recursos de acesso à informação, bem como conhecimento detalhado acerca da etapa de seleção de SNs.

No que se refere ao seu caráter bibliográfico, a pesquisa faz um levantamento da temática abordada revisando a literatura acerca dos seguintes temas: **o processo de indexação tradicional (manual)**, e **o processo de indexação automática por meio de sintagmas nominais**. Revisam-se, também, **as metodologias de seleção de SNs** propostas por alguns pesquisadores, as quais se constituem o foco deste trabalho. São, também, revisadas **as fontes de informação jurídica e a terminologia jurídica** em uso no Brasil, uma vez que se mostram pertinentes pelo fato de constituírem o domínio escolhido para a realização do experimento empírico do presente trabalho, ou seja, a área jurídica. Dentro da área jurídica, o experimento empírico referido se utiliza especificamente de um tipo específico de informação jurídica, que é a informação doutrinária composta por dissertações e teses da Biblioteca Digital de Teses e Dissertações da UFPE.

Como forma de situar e contextualizar o problema desse estudo, a pesquisa bibliográfica permitiu identificar as informações publicadas acerca do tema em questão. Para o levantamento bibliográfico, foram usados alguns recursos de busca na *web*. Foram utilizados, principalmente, dissertações, teses e artigos científicos. Os artigos científicos foram recuperados nas seguintes bases de dados:

- BRAPCI¹⁵;
- PERI¹⁶;
- SciELO¹⁷; e
- Google Acadêmico¹⁸.

¹⁵ Universidade Federal do Paraná – UFPR. Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação – BRAPCI. Disponível em: www.brapci.ufpr.br. Acesso em: 18 set. 2014.

¹⁶ Universidade Federal De Minas Gerais – UFMG. Base de Dados que contém artigos de periódicos e trabalhos publicados em anais de eventos Técnico-Científicos, refletindo a literatura nacional nas áreas de Biblioteconomia, Ciência da Informação, Arquivística e outras disciplinas. Disponível em: <http://bases.eci.ufmg.br/peri.htm>. Acesso em: 20 set. 2014.

¹⁷ Scientific Eletronic Library Online – SciELO. Disponível em: <http://www.scielo.org/php/index.php>. Acesso em> 14 out. 2014.

Também foram utilizados alguns livros que se faziam essenciais para a fundamentação basilar do tema tratado neste trabalho. Para se recuperar os materiais informacionais que serviram de base para a construção da fundamentação teórica foram utilizados vários termos de busca com vistas a encontrar o maior número de materiais informacionais acerca da temática tratada. Dentre as expressões de busca utilizadas, tem-se: “análise documentária”, “análise conceitual”, “análise de assunto”, “indexação automática”, “representação temática da informação”, “tematicidade”, “indexação de assuntos”, “catalogação de assuntos”, “sintagmas nominais”, “indexação automática *and* sintagmas nominais” e “sintagma nominal *and* indexação”.

Foi utilizada, também, a Biblioteca Digital de Teses e Dissertações do IBICT (BDTD-IBICT), na qual foi recuperado um significativo número de trabalhos que direta e indiretamente tratavam da temática proposta nesta pesquisa. É interessante ressaltar que muitos outros materiais que serviram de base para este trabalho foram identificados com o decorrer das aulas ministradas no PPGCI-UFPE, nas quais sempre surgia uma ou outra referência pertinente que era(m) indicada(s) pelos docentes do programa.

Utilizou-se a BDTD-IBICT, tendo em vista que boa parte da literatura voltada para os estudos da indexação automática baseada em SNs se encontra publicada em dissertações e teses.

Antes de detalhar o experimento realizado por esta pesquisa, faz-se necessário que se contextualize a pesquisa empírica, bem como o método de estudo de caso. A pesquisa empírica é aquela que busca respostas e soluções através de observação e prática dos fenômenos que embasam suas conclusões (MICHEL, 2009). Nessa tipologia de pesquisa, apenas o aporte teórico não é suficiente para melhor compreensão e entendimento do fenômeno estudado.

Nesse contexto, o estudo em questão se caracteriza como uma pesquisa empírica justamente pelo fato de realizar um experimento, por meio do qual se busca verificar a utilidade ou não de determinados critérios de seleção identificados na literatura no contexto da indexação automática de um conjunto de Teses e Dissertações da área jurídica. Esse experimento procura verificar se os critérios contribuem ou não para a seleção de SNs, ao passo em que selecionam SNs descritores e eliminam SNs não descritores. Para alcançar este propósito, faz-se uso da comparação de um conjunto de SNs com outro conjunto de palavras-

¹⁸ Disponível em: <<https://scholar.google.com.br/>>. Acesso em: 6 nov. 2014.

chave, identificando assim os SNs descritores, ou seja, os que são semelhantes as palavras-chave ou as contêm. Depois de identificados os SNs descritores, procedem-se com a aplicação dos critérios de seleção, identificando o comportamento de cada critério ao selecionar ou eliminar SNs descritores. São elaborados cálculos de revocação e precisão para cada critério aplicado, para que assim levante-se julgamento de cada um como útil ou não.

Essa parte empírica da pesquisa constituiu-se de nove (9) etapas gerais, as quais podem ser vistas logo a seguir:

1. Escolha dos documentos – seleção de trinta resumos de teses e dissertações da área de Direito;
2. Coleta dos documentos – os documentos são os Títulos e Resumos de cada dissertação e tese que foram transcritos para arquivos texto;
3. Indexação manual dos documentos coletados (*corpus*) - Indexação realizada por quatro bibliotecários/indexadores.
4. Determinação dos descritores de cada documento - Análise comparativa da indexação feita pelos quatro bibliotecários e o autor para cada documento visando determinar os descritores documentais;
5. Submissão dos documentos coletados ao software PALAVRAS para identificação dos SNs;
6. Extração manual dos SNs identificados e marcados pelo software PALAVRAS;
7. Definição dos SNs descritores – marcação dos SNs com valor de descritor, definidos nesta pesquisa como os semelhantes aos descritores documentais;
8. Aplicação dos critérios de seleção encontrados na literatura aos SNs extraídos - verificar a viabilidade de cada critério para a seleção de SNs que funcionem como Descritores Documentais. Os critérios encontram-se explicitados no quadro 13 na seção “2.1.2.2.1 Seleção de Sintagmas Nominais com valor de descritores”.
9. Análise e comparação dos valores de revocação e precisão na seleção de SNs descritores mensurados após a aplicação de cada critério em cada documento e no *corpus* como um todo.

A seguir, detalham-se essas nove etapas:

1. Escolha dos documentos

Foram selecionados 30 resumos de dissertações e teses da área de Direito que se encontram indexadas na Biblioteca Digital de Teses e Dissertações - BDTD da Universidade Federal de Pernambuco, atualmente, disponível no Repositório Institucional da mesma instituição. Optou-se por trabalhar com um único domínio, a área jurídica, e a escolha das dissertações e teses levou em consideração os documentos da área mais recente inseridos na BDTD.

2. Coleta dos documentos

Depois de selecionados os trinta documentos, foi feita uma preparação dos mesmos para que pudessem ser repassados para a próxima etapa. Essa preparação constituiu-se da coleta do título e resumo de cada dissertação e tese e a cópia destes para um arquivo de texto, para que assim, pudessem ser manuseados mais facilmente. Os títulos e resumos coletados foram utilizados tanto para a indexação manual como para a identificação automática dos SNs. Inicialmente, retiraram-se as palavras-chave que seguiam o resumo de cada documento, tendo em vista proporcionar uma indexação por parte dos indexadores mais imparcial, pois as palavras-chave dos autores poderiam influenciar de alguma forma os Bibliotecários na indexação.

Assim, construiu-se um Questionário no Microsoft Word contendo os títulos e os resumos coletados das dissertações e teses, bem como espaços para a inserção das palavras-chave por parte dos bibliotecários. A cada documento foi atribuído um identificador único sequencial no formato DOC. 1, DOC. 2, até DOC. 30. Esse Questionário foi definido como **Atividade de indexação manual**. Além da construção dessa atividade, foi elaborado um arquivo contendo cada título e seu respectivo resumo em formato txt nomeados, na mesma forma da atividade de indexação, de DOC. 1, DOC. 2, até DOC. 30 para que assim pudessem ser submetidos ao software PALAVRAS para a identificação dos SNs.

Durante a coleta dos títulos e resumos das dissertações e teses, percebeu-se o uso frequente dos autores de expressões em *latim*, em alguns casos sendo comum encontrar períodos completamente em *latim* dentro do resumo. Essa simples verificação condiz com o que foi visto na literatura da área e que foi exposto na subseção “2.2 informação jurídica” deste trabalho.

3. Indexação manual dos documentos coletados

Nesta etapa, a **Atividade de indexação manual** foi entregue a quatro bibliotecários de duas instituições para que procedessem com a indexação manual dos títulos e resumos. Desses quatro bibliotecários, dois pertenciam a Faculdade de Direito da Universidade Federal de Pernambuco - UFPE, e outros dois a Universidade Federal de Campina Grande – UFCG. Todos os quatro bibliotecários lidam diariamente com materiais informacionais da área jurídica, no entanto, os bibliotecários da Faculdade de Direito da Universidade Federal de Pernambuco – UFPE têm uma aproximação maior com a documentação jurídica, pois pertencem a uma biblioteca especializada em Direito, enquanto que os outros dois bibliotecários pertencem a uma biblioteca que agrega várias áreas, inclusive a área de Direito. Todavia, ambas as bibliotecas lidam com as mesmas fontes de informação, fontes doutrinárias.

Depois de alguns contatos com as duas Bibliotecas, as quais os referidos Bibliotecários estão ligados, conseguiu-se que as mesmas autorizassem a participação dos Bibliotecários nesta pesquisa, salientando que foram tomadas as providências cabíveis no que se refere ao Comitê de Ética ¹⁹da UFPE onde esta pesquisa foi desenvolvida, uma vez que ao lidar com seres humanos, mesmo que indiretamente, como é o caso desse trabalho, faz-se necessário que se consulte e obtenha-se o parecer do referido Comitê. A escolha dessas bibliotecas se deu pelo fato de serem de fácil acesso ao autor desta dissertação e também porque disponibilizaram seus profissionais bibliotecários para participarem desta pesquisa.

Apesar de os bibliotecários participantes desta pesquisa possuírem conhecimentos e experiências distintas, principalmente pelo fato de uns serem de uma biblioteca especializada e outros não, acreditou-se que esse fato não afetaria consideravelmente essa etapa da pesquisa, a qual buscou construir uma lista de palavras-chaves atribuídas por profissionais bibliotecários que lidassem com essa atividade rotineiramente, não voltando-se para uma análise do processo de indexação, ou características dessa atividade, por exemplo.

Sabendo das diferenças que cada bibliotecário possuía no que diz respeito à atividade de indexação, tomou-se o cuidado em verificar a existência de alguns pontos em comuns na prática da indexação das duas instituições. Assim, verificou-se que as duas bibliotecas fazem uso de vocabulário controlado para a indexação no dia a dia, bem como também o uso da linguagem livre. Verificou-se que ambas as bibliotecas utilizam em média um quantitativo de cinco palavras-chave/descriptores para cada documento indexado, salientando que em alguns

¹⁹ Número do processo. **CAAE:** 43024515.3.0000.5208. Encontra-se nos anexos desta dissertação comprovante de abertura de processo perante o comitê de ética da UFPE.

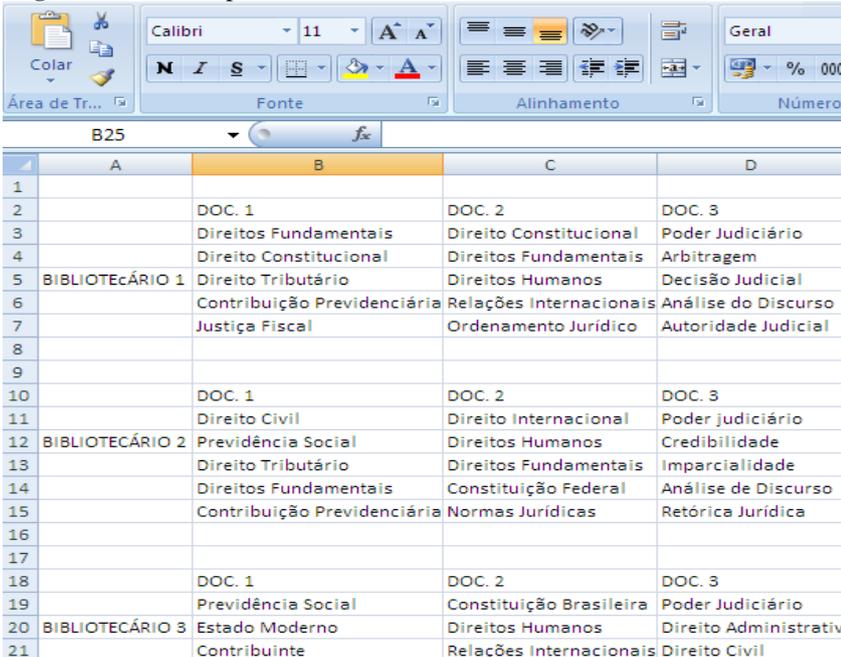
casos, de acordo com a complexidade do material indexado, esse número varia. Verificou-se, também, que os bibliotecários fazem uso de outras bases de dados quando da escolha de descritores para assuntos mais complexos, o que por sua vez demonstra uma preocupação por parte destes profissionais para com a atividade de indexação.

Depois da obtenção da autorização das Bibliotecas, marcaram-se datas específicas para ser entregue pessoalmente a atividade a cada bibliotecário, onde na ocasião explicou-se a pesquisa em desenvolvimento, retiraram-se algumas dúvidas dos participantes, bem como foram expostos alguns esclarecimentos. Depois de entregue a atividade, foi dado um prazo de 30 dias para cada bibliotecário para que devolvessem as atividades. No entanto, três dos quatro bibliotecários participantes retornaram as atividades em um prazo médio de 45 a 50 dias.

4. Determinação dos descritores de cada documento

Após a devolução das atividades por parte dos quatro bibliotecários, procedeu-se com a análise de cada atividade (etapa 3), onde foram transcritas para uma planilha do Excel cada palavra-chave atribuída a cada documento. Abaixo, tem-se uma amostra das palavras-chave transcritas da atividade manual de indexação para a planilha eletrônica. A referida planilha com todas as palavras-chave atribuídas pelos cinco indexadores (4 bibliotecários e 1 autor) encontra-se como “Apêndice D” desta Dissertação.

Figura 4 – Lista de palavras-chave



	A	B	C	D
1				
2		DOC. 1	DOC. 2	DOC. 3
3		Direitos Fundamentais	Direito Constitucional	Poder Judiciário
4		Direito Constitucional	Direitos Fundamentais	Arbitragem
5	BIBLIOTECÁRIO 1	Direito Tributário	Direitos Humanos	Decisão Judicial
6		Contribuição Previdenciária	Relações Internacionais	Análise do Discurso
7		Justiça Fiscal	Ordenamento Jurídico	Autoridade Judicial
8				
9				
10		DOC. 1	DOC. 2	DOC. 3
11		Direito Civil	Direito Internacional	Poder judiciário
12	BIBLIOTECÁRIO 2	Previdência Social	Direitos Humanos	Credibilidade
13		Direito Tributário	Direitos Fundamentais	Imparcialidade
14		Direitos Fundamentais	Constituição Federal	Análise de Discurso
15		Contribuição Previdenciária	Normas Jurídicas	Retórica Jurídica
16				
17				
18		DOC. 1	DOC. 2	DOC. 3
19		Previdência Social	Constituição Brasileira	Poder Judiciário
20	BIBLIOTECÁRIO 3	Estado Moderno	Direitos Humanos	Direito Administrativo
21		Contribuinte	Relações Internacionais	Direito Civil

Fonte: desenvolvido pelo autor.

Nessa etapa, foi feita uma comparação entre as indexações feitas por cada Bibliotecário, bem como também com as palavras-chave atribuídas pelos próprios autores dos documentos. O propósito dessa comparação foi selecionar cinco descritores para cada documento, levando em consideração a frequência de ocorrência de cada palavra-chave, ou seja, o número de vezes que cada uma foi atribuída ao documento levando em conta as 5 indexações realizadas pelos quatro bibliotecários e autor.

Escolheu-se um quantitativo de cinco descritores para cada documento, tendo em vista que é o número comumente utilizado de descritores na prática da indexação manual. Essa escolha se deu, também, pelo fato de esse quantitativo de palavras-chave já ser comumente utilizados pelos Bibliotecários pesquisados.

Durante a análise das palavras-chave para a definição dos cinco descritores para cada documento percebeu-se vários casos de palavras que diferiam apenas no número (singular e plural), como também se verificou que em alguns casos os indexadores atribuíram siglas como palavras-chave, enquanto que outros indexadores atribuíram não a sigla e sim a frase por extenso referente à sigla. Perceberam-se, também, casos em que as palavras-chave atribuídas não se repetiam o suficiente para formar o conjunto de cinco descritores, demonstrando assim a necessidade de atentar para critérios de escolha de uma ou de outra palavra-chave para formar o quantitativo de cinco palavras-chave para cada documento. Todos esses inconvenientes, encontrados no início da análise das palavras-chave, levou à tomada de algumas decisões, com vistas a propor uma análise mais uniforme possível dessas palavras-chave, seguem os critérios utilizados:

- A escolha das palavras-chave baseou-se na **frequência de ocorrência em comum**, como pode ser visto no Quadro 15, mais adiante, onde levou-se em consideração as palavras-chave que ocorriam no mínimo duas vezes indo até cinco vezes.
- Nos casos de palavras-chave idênticas, diferindo somente no que diz respeito ao número (singular e plural), contabilizou-se como palavras-chave iguais para a verificação da frequência de ocorrência, todavia escolheu-se a forma que mais apareceu dentro do grupo de palavras-chave para funcionar como descritor. Ou seja, se a palavra “informação doutrinária” ocorreu duas vezes e a palavra “informações doutrinárias” apareceu três vezes, escolheu-se para funcionar como descritor a forma mais frequente, que, no caso, foi o *plural*. Já

se a palavra “informação doutrinária” e “informações doutrinárias” ocorressem uma mesma quantidade de vezes, optar-se-ia pela palavra-chave que ocorresse no título ou resumo do documento, e se as duas tivessem ocorrência no título ou resumo, como critério de escolha de uma forma ou de outra forma (singular ou plural), optar-se-ia pela escolha da forma no singular.

- Em casos de palavras que continham preposições ligando o substantivo ao complemento (adjetivo ou palavra adjetivada, como, por exemplo, “Direito de família” e “Direito da família”, contabilizou-se como sendo uma mesma palavra-chave, optando-se, da mesma forma que no item anterior, pela forma mais frequente para servir como descritor. Nesse mesmo contexto, inserem-se as palavras que representavam um mesmo conceito, no entanto mudavam a forma do complemento, tendo em alguns casos o complemento (adjetivo) ligado por preposição e em outros casos sendo ligado sem preposição, como, por exemplo, o termo “Direito do trabalho” e “Direito trabalhista”. Nesses casos contabilizou-se como sendo uma mesma palavra-chave, mas optou-se pela forma mais frequente para compor os descritores de cada documento.
- Em casos de documentos que só tiveram duas palavras-chaves com duas repetições, a escolha das outras três para compor os cinco descritores de cada documento levou em consideração as palavras-chave de ocorrência “um” que se encontravam no corpo do texto do título e do resumo, uma vez que, acreditava-se que as palavras que ocorrem no corpo do texto são mais representativas do conteúdo desse texto.
- Em casos de documentos que tiveram mais de cinco grupos de palavras-chave repetidas, a escolha de uma ou de outra palavra-chave levou em consideração a ocorrência dessa palavra no título ou resumo do documento.
- Nos casos de atribuição de siglas como palavras-chave, verificou-se a existência da sua forma por extenso como palavra-chave e contabilizou-se como sendo uma mesma palavra-chave, todavia a escolha da forma (por extenso ou sigla) para funcionar como descritor levou em consideração a que aparecia na forma por extenso.

Mais adiante, segue um quadro, a título de exemplificação, de palavras-chave semelhantes que ocorreram tanto no singular como no plural:

Quadro 15 – Palavras-chave semelhantes resultantes da Indexação Manual.

Indexador	Palavras-chaves	Documento	Ordem	Descritores Doc. 15	Frequência de ocorrência
BIBLIOTECÁRIO 1	Administração Pública	DOC. 15	1	Agências Reguladoras	5
BIBLIOTECÁRIO 3	Administração Pública	DOC. 15	2	Administração Pública	2
PRÓPRIO AUTOR	Administrativo	DOC. 15	3	Cirurgião Plástico	2
BIBLIOTECÁRIO 1	Agência Reguladora	DOC. 15	4	Ordenamento Jurídico	2
BIBLIOTECÁRIO 2	Agências Reguladoras	DOC. 15	5	Regulamentação	2
BIBLIOTECÁRIO 3	Agências Reguladoras	DOC. 15			
BIBLIOTECÁRIO 4	Agências Reguladoras	DOC. 15			
PRÓPRIO AUTOR	Agências Reguladoras	DOC. 15			
BIBLIOTECÁRIO 4	Cirurgia Plástica	DOC. 15			
BIBLIOTECÁRIO 3	Cirurgião Plástico	DOC. 15			
BIBLIOTECÁRIO 4	Cirurgião Plástico	DOC. 15			

Fonte: desenvolvido pelo autor.

Assim, considerou-se que o termo “agência(s) reguladora (s)” ocorreu cinco vezes, todavia, optou-se por utilizar o termo no plural “agências reguladoras” para compor os descritores do DOC. 15, pois foi a forma no plural que mais ocorreu nas indexações. Em casos de empate de frequência de ocorrência, a escolha foi pela forma que aparecia no resumo ou título do documento. Existindo as duas formas no resumo, tanto no singular como no plural, optou-se pelo singular. Dodebei (2002, p. 103) no que se refere à escolha de se utilizar um termo no singular ou no plural na construção de Linguagens de Indexação, ressalta que “se a representação do conceito, na sua forma simbólica, for mais geral quando a palavra estiver no singular, tal forma deve ser a escolhida.”. Ainda a esse respeito, tomou-se o cuidado com o estabelecimento das definições conceituais.

Como resultado da análise comparativa, tem-se uma listagem composta por descritores documentais resultantes de um processo de análise de coerência. Esta listagem funcionou como lista de referência para a comparação com a Lista de SNs extraídos do PALAVRAS nas etapas posteriores.

Ao finalizar todo o procedimento de análise e de definição dos descritores para cada documento, percebeu-se, que em alguns poucos casos, as indexações feitas pelos quatro bibliotecários e os autores possuíam 100% de coincidência na atribuição de um determinado termo. No entanto, verificou-se também que, na maior parte das indexações, as palavras atribuídas pelo próprio autor eram mais específicas, enquanto que nas indexações dos bibliotecários, evidenciou-se mais termos gerais, como por exemplo, termos como “Direito”,

“Ciências Humanas” atribuídos a documento que tratavam de assuntos específicos dentro da área de Direito.

A título de complementação, tendo em vista o aproveitamento dos dados das indexações, elaborou-se uma análise de consistência quantitativa da indexação feita pelos quatro bibliotecários. A referida análise constituiu-se de uma análise rígida (GIL LEIVA, 2001), onde compararam-se os termos que coincidiam completamente. Utilizou-se uma variação da fórmula matemática de Hooper (1965), a qual se encontra mais adiante:

$$\frac{100C}{C + A + B}$$

Onde, (C) se refere aos termos comuns nas duas indexações; (A) refere-se aos termos utilizados na indexação A, mas não na B; e (B) refere-se aos termos utilizados na indexação B, mas não na A.

Realizaram-se os cálculos por pares de indexadores da mesma instituição, ou seja, calculou-se primeiro a consistência dos Bibliotecários 1 e 2 (par 1), e em seguida calculou-se dos Bibliotecários 3 e 4 (par 2). Os índices verificados na medida de consistência rígida foram calculados com o intuito de se encontrar uma média geral de consistência. O resultado foi 18,66% para o par 1 e 18,03% para o par 2, demonstrando assim um baixo nível de consistência, não estando inclusos na média dos índices que oscilam entre os 25 % e os 60 % de coincidências obtidos em resultados de estudos desenvolvidos por Gil Leiva (2008).

Apesar de saber que vários são os fatores que interferem no alcance de uma consistência satisfatória, acredita-se que os índices deveriam ser maiores, uma vez que o par A pertence a uma biblioteca que faz uso de um vocabulário controlado, onde, também, foram solicitados um mesmo quantitativo de termos de indexação, no caso, cinco termos, a cada bibliotecário. Vale ressaltar aqui, que para estudos mais aprofundados nessa vertente, outras variáveis deveriam ser controladas, buscando ao máximo controlar a subjetividade inerente à atividade de indexação, atentando-se para elementos já verificados por Gil Leiva, Rubi e Fujita (2008), como, padronização e atualização das linguagens de indexação utilizadas, fatores referentes ao ambiente de indexação de cada indexador, bem como a existência de políticas de indexação que norteiem essa atividade.

5. Submissão dos documentos coletados ao software PALAVRAS para identificação dos SNs;

Nesta etapa os mesmos documentos que foram submetidos aos indexadores humanos foram, agora, submetidos à indexação automática para que fossem identificados os SNs que se encontravam no corpo do resumo e do título de cada documento.

Silva (2014) realizou estudo onde comparou o desempenho de três softwares, dentre os quais se encontrava o PALAVRAS. Apesar de apresentar alguns inconvenientes, o PALAVRAS alcançou um melhor desempenho em relação aos dois outros softwares analisados. Dessa forma, o parser PALAVRAS foi escolhido para realizar a identificação dos SNs. Essa escolha se deu pelo fato de, além de estar disponível publicamente via internet, o mesmo já vem sendo utilizado por diversos pesquisadores, como, Vieira et al.(2000), Miorelli (2001), Souza (2005), Santos (2005), Arcoverde (2007), Maia (2008), Lopes, (2012), Silva (2014), Martins (2014), Souza e Raghavan (2014) entre outros, o que demonstra , por sua vez, uma viabilidade no que diz respeito ao seu desempenho.

O PALAVRAS foi desenvolvido pela Southern University of Denmark. Esse software pode ser acessado e utilizado livremente via internet, apesar de possuir algumas limitações quando utilizado online. O PALAVRAS recebe o documento em formato textual e analisa-o, levando em consideração o caráter morfológico, sintático e semântico do texto, onde cada palavra é etiquetada em uma classe gramatical, e as orações são marcadas de acordo com as várias possibilidades de classificação sintática e semântica. Esse software analisa as orações e suas classificações, tendo em vista alcançar uma classificação exata dos elementos que compõem o texto, evitando assim ao máximo a existência de ambiguidades. Depois de submetido o texto ao software, o mesmo retorna-o com os SNs marcados, bem como com todas as palavras componentes do texto com suas classificações. A saída do texto analisado pode ser visualizada de várias formas, inclusive na forma de visualização arbórea.

A título de exemplificação da análise feita pelo PALAVRAS, segue na Figura 5 um trecho do DOC. 1 analisado por este software:

Figura 5 – Análise feita pelo PALAVRAS.

a fiscalidade, fenômeno essencial à existência do estado, também tem sofrido mudanças, de entre as quais se destaca o grande aumento de contribuições percentualmente em relação a a receita estatal tributária, que, devido a inúmeras razões, passam a substituir gradativamente os impostos diretos.

```

- S: g(np)
  | -D: pron(det "o" &lt;artd&gt; DET F S)      a
  | -H: n("fiscalidade" F S)      fiscalidade
- A: cl(fcl)
  | -D: g(np)
    | -H: n("fenômeno" M S)      fenômeno
    | -D: adj("essencial" M S)  essencial
    | -D: g(pp)
      | -H: prp("a" &lt;sam&gt; &lt;np-long&gt;) a
      | -D: g(np)
        | -D: pron(det "o" &lt;-sam&gt; &lt;artd&gt; DET F S)      a
        | -H: n("existência" F S) existência
        | -D: g(pp)
          | -H: prp("de" &lt;sam&gt; &lt;np-close&gt;) de
          | -D: g(np)
            | -D: pron(det "o" &lt;-sam&gt; &lt;artd&gt; DET M S) o
            | -H: n("estado" &lt;prop&gt; M S) estado
          | -H: n("transformação" F P) transformações
      | -A: adv("também" &lt;left&gt;) também
  | -P: vp
    | -VD: v(fin "ter" &lt;fmc&gt; PR 3S IND VFIN) tem
    | -H: v(pcp "sofrer" &lt;mv&gt;) sofrido
  | -Od: n("mudança" F P) mudanças
  | -A: g(pp)
    | -H: prp("de" &lt;sam&gt; &lt;right&gt;) de

```

Fonte: desenvolvido pelo autor.

Como pode ser visto na Figura 5 têm-se os termos marcados com símbolos que identificam as classes morfossintáticas de cada termo. Com base no trecho acima analisado, tem-se o lexema “existência”, onde o software marcou-o com sua classificação morfológica, utilizando o símbolo n que indica um nome, o F e o S entre parênteses indicam, respectivamente, o gênero e o número, ou seja, feminino e singular, além de mostrar a forma canônica de cada palavra, no caso da palavra tida como exemplo, “existência” mesmo. As marcações “np” sinalizam Sintagmas Nominiais.

Todo o processo de submissão dos documentos ao PALAVRAS iniciou-se primeiramente com a inserção do título do DOC. 1 ao software, depois de extraídos os SNs do título, inseriu-se o resumo, e foram extraídos os SNs do resumo. Esse procedimento foi feito em todo o *corpus* da pesquisa, DOC. 1, DOC. 2, DOC 3 até DOC. 30.

Apesar da capacidade robusta do software PALAVRAS na identificação de SNs, suas funcionalidades disponíveis online não incluem a extração dos SNs, permitindo apenas a

marcação desses. Dessa forma, a extração dos SNs identificados foi feita manualmente, a qual será detalhada mais adiante.

6. Extração manual dos SNs identificados e marcados pelo software PALAVRAS;

Depois de identificados, os SNs foram extraídos para que assim se tornassem passíveis de manipulação. A extração foi bem exaustiva, uma vez que o PALAVRAS apresentou alguns problemas, como, estruturas que não se constituíam em SNs, ou ocultando SNs que deveriam ser identificados por ele, ou, ainda, a ocultação de nomes próprios, não os identificando como SNs, entre outros casos.

O processo de extração dos SNs pode ser considerado o procedimento mais cansativo e exaustivo de toda essa pesquisa, uma vez que exigiu do pesquisador uma análise cuidadosa para que não fosse omitido nenhum SN identificado pelo PALAVRAS, bem como também não passasse despercebido algumas falhas do próprio software que acarretava ora na omissão de SNs ora na identificação de estruturas que não constituíam SNs.

Tanto o **título** quanto o **resumo** de cada documento foram inseridos por completo de um de cada vez, seguidos cada um de um ponto final (.), para que não causasse confusão no momento de análise textual por parte do software.

Apesar de ter sido submetido o título e o resumo por completo um de cada vez ao software, o mesmo analisava o texto por partes (trechos). Nesse contexto, no momento em que se extraíam os SNs de um determinado trecho do texto, o pesquisador voltava para o trecho analisado, tendo em vista verificar se a análise sintática cobria todo o trecho ou não. Foi por meio dessa análise mais cuidadosa de conferência do texto corrido submetido com a própria análise feita pelo PALAVRAS que se verificou que em alguns casos o software cometia alguns erros de análise.

A extração constitui-se basicamente da transcrição de cada estrutura identificada como SN para um editor de texto e um editor de planilha. Ao passo que iam sendo extraídos os SNs, o pesquisador ia registrando alguns erros de análise do software, tendo em vista verificar estruturas que não constituíam SNs, mas que foram identificadas como tal, como também verificar a omissão, em alguns casos, de SNs.

Uma falha relativamente frequente encontrada nas análises feitas pelo software diz respeito à classificação errônea de palavras, onde adjetivos eram marcados como substantivos “nomes”, ou verbos eram considerados como palavra substantivada quando não deveriam ser.

A busca de solução para essa falha é crucial para melhoramentos futuros, uma vez que a partir dela várias sentenças são analisadas incorretamente, ou seja, erros na etapa inicial, na classificação das palavras, acarretaram em erros na identificação de SNs posteriormente. Um exemplo desse tipo de falha pode ser visto na Figura 6.

Figura 6 – Falha na identificação de SNs pelo PALAVRAS.

Título do DOC. 11 : Parto anônimo e principio da afetividade.	
SOURCE: Running text	
1. parto anônimo e principio da afetividade.	
A1	
UTT:par	
.	
-CJT:cl(fcl)	
-P:v(fin "partir" <fmc> PR 1S IND VFIN)	parto
-fCs:adj("anônimo" M S)	anônimo
-CO:conj("e" <f:2349620>)	e
-CJT:x	
-P:v(fin "principiar" <fmc> PR 1S IND VFIN)	principio
-A:g(pp)	
-H:prp("de" <sam-> <right>)	de
-D:g(np)	
-D:pron(det "o" <-sam> <artd> DET F S)	a
-H:n("afetividade" F S)	afetividade

Fonte: desenvolvido pelo autor.

Como pode ser visto na Figura 6, o software marcou o termo “parto” como sendo verbo, no entanto, o referido termo não constitui verbo e sim um nome, um substantivo. Devido a essa etiquetagem errada, o software não identificou o SN “parto anônimo”, onde o termo “parto” é substantivo e o termo “anônimo” é um adjetivo que qualifica o nome “parto”.

Outra falha percebida diz respeito à omissão de algumas palavras do texto analisado, onde o software por algum motivo omitia palavras fundamentais para a formação de SNs, deixando os possíveis SNs quebrados ou sem serem identificados.

Figura 7 – Falha do PALAVRAS ao omitir termos.

<p>Trecho do DOC. 14: o contrato de concessão estabelecido na lei 9.478/97, da criação da Pré-Sal Petróleo</p>	
<pre> .-CJT:g(np) -D:pron(det "o" &lt;artd&gt; DET M S) o -H:n("contrato" M S) contrato -D:prp("de" &lt;np-close&gt;) de -D:cl(icl) -P:v(pcp "estabelecer" M S) estabelecido -A:g(pp) -H:prp("em" &lt;sam-&gt; &lt;right&gt;) em -D:g(np) -D:pron(det "o" &lt;-sam&gt; &lt;artd&gt; DET F S) a -H:n("lei" F S) lei -CJT:adjp -H:num("9.478/97" &lt;card&gt; M/F P) 9.478/97 -D:g(pp) -H:prp("de" &lt;sam-&gt; &lt;np-close&gt;) de -D:g(np) -D:pron(det "o" &lt;-sam&gt; &lt;artd&gt; DET F S) a -H:n("criação" F S) criação -D:g(pp) -H:prp("de" &lt;sam-&gt; &lt;np-close&gt;) de -D:g(np) -D:pron(det "o" &lt;-sam&gt; &lt;artd&gt; DET F S) a -H:prop("Pré-Sal" F S) Pré-Sal -D:g(np) -H:prop("Petróleo" M S) Petróleo </pre>	

Fonte: desenvolvido pelo autor.

A Figura 7 demonstra a falha do software ao omitir o termo “concessão”, que viria logo após a primeira preposição “de”, complementando a expressão “contrato de...”. Percebeu-se esse comportamento do software tanto quando foi submetido o resumo por completo de uma única vez, como também quando foi submetido trechos, parágrafos do resumo.

Além do exposto até aqui acerca das falhas encontradas na análise textual, por parte do PALAVRAS, verificou-se também que o mesmo, em alguns momentos, classificava os numerais isolados como sendo SNs, como também algumas palavras apenas como “nomes”, mas não como SNs. A maior parte das falhas encontradas no PALAVRAS pode ser resultado da etiquetagem errada, uma vez que, como já mencionado, os erros na categorização das palavras afetam todo o procedimento posterior de identificação dos SN, resultando em identificações erradas ou omissão de SNs.

Os mesmos problemas identificados nesta fase de identificação e extração dos SNs foram percebidos por Silva (2014), onde este autor sugeriu possíveis soluções para alguns

problemas detectados por ele à época. Deve-se levar em consideração também que alguns recursos da linguagem natural, como, por exemplo, as construções anafóricas e as elipses, são complexas de serem interpretadas pela máquina, o que tende a resultar nessas falhas.

A extração de todas as estruturas marcadas como SN possibilitou a verificação do percentual de acerto do PALAVRAS em relação ao seu próprio resultado, ou seja, a razão do total da quantidade de estruturas identificadas pelo software como sendo SNs, e a quantidade de estruturas que realmente constituíam SNs, ou verdadeiros SNs. Sendo assim, depois de extraídos todas as estruturas identificadas como sendo SNs, selecionaram-se as que realmente constituíam SNs, sendo descartadas as estruturas que iniciavam por preposição ou conjunção, e as que continham verbo em sua estrutura. Os SNs omitidos pelo PALAVRAS foram registrados apenas com o intuito de explicitar as falhas encontradas no software no momento de identificação de SNs.

Mesmo com o exposto anteriormente acerca das falhas na identificação de SNs pelo PALAVRAS, esse software, considerado pela literatura como um dos melhores para identificação de SNs (MAIA, 2008; MORELLATO, 2010; LOPES, 2012; SILVA, 2014; MARTINS, 2014), de modo geral, alcançou um considerável desempenho na identificação dos Sintagmas. Em suma, o PALAVRAS identificou 2089 estruturas como sendo SNs, sendo que deste total, 1953 constituíam expressões que realmente eram SNs, denominado, neste trabalho, de verdadeiros SNs e 136 não constituíam Sintagmas Nominais, representando, assim, um total de 94% de acerto em relação aos seus próprios resultados, o que vem confirmar a taxa obtida pelo seu idealizador, Bick (2000) que foi de 98% de acerto na identificação de SNs, como também a taxa obtida por Silva (2014) que foi de 94% de acerto em relação aos seus próprios resultados.

Posteriormente, esses 1953 SNs foram transcritos para planilhas eletrônicas, sendo uma para cada documento, compondo uma *Lista com SNs verdadeiros* para cada documento do *corpus* desta pesquisa. A reunião destas listas com os SNs identificados automaticamente e extraídos manualmente deu-se o nome de **Lista com SNs verdadeiros**.

De posse desta listagem com todos os SNs verdadeiros, prosseguiu-se com a junção da mesma com a Lista de palavras-chave (produto da indexação manual) em uma única pasta do Excel, onde foram criadas 30 planilhas, uma para cada documento do *corpus*, para que assim pudessem ser manuseados mais facilmente.

7. Determinação dos SNs descritores

Antes de iniciar a aplicação de cada critério ao conjunto de SNs de cada documento, realizou-se a comparação entre os SNs que eram semelhantes às palavras-chave ou que as continham em suas estruturas, para que assim fossem marcados como SNs descritores. O método utilizado para julgamento de cada critério de seleção foi a comparação dos SNs com as palavras-chave atribuídas pelos indexadores a cada documento. Assim, os SNs semelhantes às palavras-chave ou que continham as palavras-chave em suas estruturas eram considerados SINTAGMAS NOMINAIS DESCRITORES e os que não se encaixavam nessa situação foram considerados apenas Sintagmas Nominais.

8. Aplicação dos critérios de seleção encontrados na literatura aos SNs extraídos

Nesta etapa foi feita a aplicação dos critérios de seleção expostos no Quadro 13 desta dissertação, com exceção de dois critérios, o critério de verificação de ocorrência de SN em tesauro da área e o critério de verificação de ocorrência de SNs em outros textos/documentos diferentes (corpus contrastante). Esses critérios não foram utilizados neste experimento, devido à baixa eficácia reportada por Souza (2005) no uso de tesauro para seleção de SNs e à inexistência de corpus contrastante disponível, respectivamente.

Os critérios de seleção foram organizados em um arquivo do Excel. Dentro desse arquivo organizaram-se planilhas referentes a cada documento, sendo nomeadas de DOC. 1, DOC. 2 até DOC. 30. Além de conter os SNs, cada planilha continha também as palavras-chave resultantes da análise comparativa da indexação manual, para que assim pudessem ser comparadas com os SNs de cada documento. Juntamente com os SNs e as palavras-chave, em cada planilha, continha os critérios de seleção já mencionados nesta pesquisa. Todo o processo de comparação entre SNs e palavras-chave, bem como a aplicação dos critérios de seleção a cada SN iniciou-se com um único arquivo do Excel, subdividido em trinta planilhas. Alguns critérios, como, por exemplo, frequência de ocorrência no texto e frequência de ocorrência no conjunto de documentos, foram aplicados a todos os sintagmas, já outros critérios, como, por exemplo, o descarte de SNs que possuíam numerais, o descarte de SNs que possuíam pronomes como núcleo, não foram aplicados a todos os SNs, tendo em vista que eram critérios que ora eram aplicados ora não. Esta etapa de aplicação e análise dos critérios foi uma atividade exaustiva, tendo em vista que foi desenvolvida manualmente em sua maior parte. Buscou-se verificar, neste momento, a viabilidade de cada critério no que se

refere à seleção de SNs descritores e eliminação de SNs não descritores, ou seja, aqueles sintagmas que pouco contribuíam para a descrição de conteúdos documentais.

9. Análise e comparação dos valores de revocação e precisão na seleção de SNs descritores mensurados após a aplicação de cada critério, em cada documento e no *corpus* como um todo.

Para a mensuração da aplicação de cada critério, bem como a aplicação de forma útil ou não, se utilizou as medidas de revocação e precisão, tomando sempre como referência os SNs denominados descritores, ou seja, aqueles que continham as palavras-chave da indexação manual ou eram semelhantes a estas.

Os critérios utilizados nesse experimento encaixam-se nas seguintes categorias: Critérios de eliminação, Critérios de adição, Critérios de ordenação, critério de nível e critério de posição, porém todos analisados com um único propósito, o de contribuir para a escolha de SNs descritores.

Na aplicação dos critérios de eliminação, verificou-se se, ao aplicar o critério, o mesmo eliminaria os SNs que não continham as palavras-chave (não descritores), mostrando-se assim um critério útil.

Na aplicação dos critérios de adição, verificou-se se ao aplicar os critérios, os mesmos possibilitavam a obtenção de SNs que continham as palavras-chave (descritores), se sim então esses critérios eram identificados como úteis.

Já para os critérios de ordenação do SN por importância, verificou-se se, ao ordenar os SNs pelos critérios, os SNs descritores encontravam-se dentro de determinados pontos de corte definidos por meio da análise do comportamento de cada critério de ordenação.

No critério de nível, procurou-se verificar a relação do nível do SN com a sua capacidade de funcionar como descritor documental. Essa análise foi feita por meio de ponto de corte.

O critério de posição do SN procurou verificar se os primeiros SNs extraídos de cada documento eram SNs descritores, procurando verificar se existe relação entre os SNs descritores e determinadas partes dos documentos.

Mais adiante, analisa-se cada critério, expondo um quadro resumo de aplicação dos critérios aos trinta documentos que compuseram o *corpus* desta pesquisa. Alguns critérios, para que fossem passíveis de serem analisados, foram verificados em conjunto com outros,

como é o caso dos critérios: frequência no documento e frequência normalizada; e frequência na coleção e frequência IDF.

Para cada critério, calcularam-se as medidas de revocação e precisão, as quais se encontram no decorrer da análise de cada critério.

4 ANÁLISE DOS RESULTADOS EXPERIMENTAIS

Nesta seção, expõem-se os resultados alcançados com a aplicação de cada critério, e por fim, uma análise dos critérios individuais e em conjunto.

Para cada critério de seleção são elaborados cálculos de revocação e precisão, tendo em vista obter dados que contribuam para o julgamento dos mesmos. Assim, fez-se necessário que se levantasse as taxas de revocação e precisão sem a aplicação de cada critério para que assim, fosse possível realizar comparações entre as taxas de revocação e precisão com e sem a aplicação de cada critério.

Na análise por sintagmas nominais, ou seja, a análise da utilidade dos critérios na separação dos sintagmas nominais descritores dos não descritores, levando em conta todos os sintagmas nominais extraídos dos documentos da coleção, tem-se que o corpus é constituído de 423 sintagmas nominais descritores e 1358 sintagmas nominais não descritores. O que resulta numa precisão de 23,75% e revocação de 100% quando da não aplicação de nenhum critério.

Nesse contexto, segue, mais adiante, um quadro onde são expostas as taxas de revocação e precisão sem a aplicação de nenhum critério na análise por documento. Os valores de precisão e cobertura levantados no parágrafo anterior e no Quadro 16 servem de referência para todos os critérios analisados nesta seção.

Quadro 16 – Médias de revocação e precisão sem aplicação de critério

MÉDIAS DE REVOCAÇÃO E RPECISÃO SEM APLICAÇÃO DE CRITÉRIOS			
DOCUMENTO	SNs SELECIONADOS/ EXTRAÍDOS	REVOCAÇÃO	PRECISÃO
DOC. 1	34	100,0%	26,4%
DOC. 2	37	100,0%	27,0%
DOC. 3	56	100,0%	21,4%

DOC. 4	50	100,0%	26,0%
DOC. 5	45	100,0%	11,1%
DOC. 6	51	100,0%	54,9%
DOC. 7	70	100,0%	11,4%
DOC. 8	51	100,0%	31,3%
DOC. 9	29	100,0%	68,9%
DOC. 10	57	100,0%	26,3%
DOC. 11	72	100,0%	19,4%
DOC. 12	33	100,0%	18,0%
DOC. 13	60	100,0%	20,0%
DOC. 14	60	100,0%	25,0%
DOC. 15	61	100,0%	16,3%
DOC. 16	72	100,0%	15,2%
DOC. 17	69	100,0%	36,2%
DOC. 18	72	100,0%	34,7%
DOC. 19	47	100,0%	38,2%
DOC. 20	83	100,0%	18,0%
DOC. 21	79	100,0%	26,5%
DOC. 22	68	100,0%	10,2%
DOC. 23	54	100,0%	25,9%
DOC. 24	68	100,0%	22,0%
DOC. 25	56	100,0%	14,2%
DOC. 26	87	100,0%	9,1%
DOC. 27	88	100,0%	20,4%
DOC. 28	57	100,0%	31,5%
DOC. 29	68	100,0%	26,4%
DOC. 30	47	100,0%	17,0%
MEDIA	59,3	100,0%	25,0%
DESVIO PADRÃO	15,53967854	0,0%	12,7%
MEDIANA	0,585	100,0%	23,5%

Fonte: desenvolvido pelo autor.

4.1 Descarte dos SNs que Contêm Numerais

A aplicação desse critério ao *corpus* buscou verificar a viabilidade do mesmo ao contribuir ou não para a seleção de SNs Descritores. Conforme foi exposto no referencial deste trabalho, Lopes (2012) fez uso desse critério para a seleção de conceitos tomando por base os SNs. Na pesquisa dessa autora, esse critério se mostrou eficiente ao eliminar SNs como: “três meses”, “ano 2000”. No entanto, nesse momento buscou-se verificar a eficiência desse critério no *corpus* utilizado nesta pesquisa, verificando se o mesmo contribuía para a

seleção de sintagmas nominais descritores, eliminando sintagmas que realmente não eram pertinentes e não eliminando sintagmas relevantes. Esse critério elimina os sintagmas que contenham numerais, seja na forma extensa ou por meio de caracteres numéricos (dígitos).

Esse critério foi aplicado em 46% do *corpus* utilizado nesta pesquisa, correspondendo a exatamente 14 dos 30 documentos analisados. Durante a aplicação desse critério, verificou-se que o mesmo contribuiu para o descarte de SNs como, por exemplo, “três prestações”, “três vias processuais”, “três modos”, “os últimos seis anos”, “essas duas teorias”, “duas ou mais personalidades”, “cinco (5) audiências públicas”, demonstrando assim um comportamento semelhante ao encontrado por Lopes (2012), excluindo sintagmas referentes a datas, períodos, etc. Todavia, um detalhe percebido durante a aplicação desse critério foi que o domínio estudado possui vários sintagmas com números de leis, incisos e artigos, característicos da própria área, em suas estruturas. Assim, nesse momento, verificou-se que o critério em questão descartou sintagmas como “a Constituição Brasileira de 1988”, “o art. 1º da lei 8.137/1990”, “a lei 8.137/1990”.

O critério de descarte de SNs que contenham números foi aplicado, em média, a 2% do total de 1781 sintagmas únicos, um percentual bem pequeno, quando comparado com o mesmo critério aplicado por Lopes (2012), onde a autora verificou um percentual de aproximadamente 16% de aplicação desse critério. No entanto, o *corpus* utilizado na presente pesquisa é bem menor e possui características próprias.

Mais adiante, encontra-se um Quadro resumo de aplicação desse critério, ressaltando a quantidade de SNs descritores e não descritores que atendem ou não o critério. Esse quadro foi elaborado com base na análise individual de cada documento. Levantou-se, também, as taxas de revocação e precisão, refletindo, respectivamente, o percentual de SNs descritores selecionados do total de SNs descritores e o percentual de SNs descritores selecionados do total de SNs selecionados. Esses dados podem ser verificados mais adiante.

Quadro 17 – Quantitativo de SNs descritores e não descritores eliminados pelo critério: descarte de SNs com numerais.

CRITÉRIO DE ELIMINAÇÃO (NUMERAIS) COLEÇÃO	TOTAL DE SNs DESCRITORES	TOTAL DE SNs não DESCRITORES	Percentual de SNs DESCRITORES
Quantidade de SNs que não atendem ao critério	408	1336	23,3%
Quantidade de SNs que atendem ao critério	15	22	40,5%
TAXAS DE REVOCAÇÃO E PRECISÃO			
REVOCAÇÃO: 96,4%			

PRECISÃO: 23,3%			
------------------------	--	--	--

Fonte: desenvolvido pelo autor.

Com base nos dados demonstrados no Quadro 17, verifica-se que esse critério de eliminação alcançou uma boa taxa de revocação, no entanto, a precisão ficou abaixo da média quando da não aplicação que é de 23,75% na análise por sintagma nominal. Como pode ser visto na quarta coluna do Quadro 17, o percentual de SNs descritores que foram eliminados pelo referido critério foi de 40,5%, ou seja, de todos os SNs eliminados por este critério, quase metade deles eram descritores, demonstrando, assim, que o critério não se mostrou tão viável. Se, com a aplicação desse critério, tivessem sido eliminados uma quantidade menor de SNs descritores com a aplicação desse critério, poderia-se vê-lo como um critério que ajudou a eliminar SNs não descritores.

Com base nos dados do Quadro 18, mais adiante, procurou-se verificar a existência de algumas características em comum dos 15 SNs descritores (terceira linha e segunda coluna) que foram eliminados por este critério. Mais adiante, segue uma amostra dos SNs descritores que foram eliminados por esse critério.

Quadro 18 – Exemplos de SNs descritores eliminados pelo critério de descarte de SNs com numerais.

SNS DESCRITORES QUE FORAM ELIMINADOS COM A APLICAÇÃO DO CRITÉRIO DE ELIMINAÇÃO DE SNs COM NUMERAIS
apenas três modos de entidades familiares
art. 5º, § 2º da Constituição Federal
a administração pública brasileira nos últimos seis anos
o instituto processual da audiência pública introduzido no direito brasileiro pelas leis nº 9.868/99 e 9.882/99
a audiência pública introduzido no direito brasileiro pelas leis nº 9.868/99 e 9.882/99
o crime de sonegação fiscal em conformidade com o art. 1º da lei nº 8.137/1990
o direito penal de duas velocidades
a constituição brasileira de 1988

Fonte: desenvolvido pelo autor.

No Quadro 18 são expostos alguns SNs descritores que foram eliminados de todos os 15 que foram eliminados neste critério. De todos os 15 SNs descritores eliminados, apenas um se referia a data e dois referiam-se a períodos. Percebe-se que a maior parte dos SNs descritores eliminados se constitui de SNs que possuem números de leis e artigos. Vale ressaltar que essa característica é bem particular de SNs desse domínio específico, o qual faz uso excessivo de números para referenciar leis específicas e artigos de determinadas leis.

Depois de feitas as análises citadas anteriormente, prosseguiu-se com uma análise por documento, onde calculou-se novamente os índices de precisão e revocação, demonstrando essas taxas de cada documento e em seguida de toda a coleção utilizada nesta pesquisa. Mais adiante segue um quadro resumo de aplicação desse critério com os percentuais de revocação e precisão da aplicação do mesmo.

Quadro 19 – Médias de revocação e precisão / Critério: Descarte de SNs com numerais.

QUADRO RESUMO DE APLICAÇÃO DO CRITÉRIO DE ELIMINAÇÃO DE SNs COM NUMERAIS				
DOCUMENTO	QUANTIDADE DE SNs ELIMINADOS	PERCENTUAL DE APLICAÇÃO DO CRITÉRIO EM RELAÇÃO SNs UNICOS	REVOCAÇÃO	PRECISÃO
DOC. 1	0	0,0%	100,0%	26,0%
DOC. 2	1	2,7%	90,0%	25,0%
DOC. 3	0	0,0%	100,0%	21,4%
DOC. 4	1	1,0%	92,3%	21,4%
DOC. 5	0	0,0%	100,0%	11,1%
DOC. 6	0	0,0%	100,0%	54,9%
DOC. 7	3	4,2%	100,0%	11,9%
DOC. 8	0	0,0%	100,0%	31,3%
DOC. 9	5	17,2%	62,5%	75,0%
DOC. 10	0	0,0%	100,0%	26,3%
DOC. 11	0	0,0%	100,0%	19,4%
DOC. 12	0	0,0%	100,0%	18,0%
DOC. 13	0	0,0%	100,0%	20,0%
DOC. 14	4	6,6%	100,0%	26,7%
DOC. 15	3	4,9%	80,0%	13,7%
DOC. 16	1	1,3%	100,0%	15,4%
DOC. 17	0	0,0%	100,0%	36,2%
DOC. 18	2	2,7%	100,0%	35,7%
DOC. 19	0	0,0%	100,0%	38,2%
DOC. 20	10	12,0%	86,6%	17,8%
DOC. 21	1	1,2%	100,0%	26,9%
DOC. 22	0	0,0%	100,0%	10,2%
DOC. 23	0	0,0%	100,0%	25,9%
DOC. 24	0	0,0%	100,0%	22,0%
DOC. 25	0	0,0%	100,0%	14,2%
DOC. 26	4	4,5%	75,0%	7,2%
DOC. 27	1	1,1%	94,0%	19,5%
DOC. 28	0	0,0%	100,0%	31,5%

DOC. 29	1	1,4%	94,4%	25,3%
DOC. 30	0	0,0%	100,0%	17,0%
MEDIA	1,2333333	2,0%	96,1%	24,8%
DESVIO PADRÃO	2,1922014	3,9%	8,9%	13,7%
MEDIANA	0	0,0%	100,0%	21,7%

Fonte: desenvolvido pelo autor.

Com base no Quadro 19, mais especificamente na terceira coluna, tem-se o percentual de aplicação do critério em relação à quantidade de SNs únicos. Esses dados permite ver a existência de uma oscilação de aplicação, onde têm-se documentos com 0,0% de aplicação e outros documentos com 4,0% e 12% de aplicação, entre outros. Os dados expostos na terceira coluna do Quadro 19, bem como também dos quadros dos outros critérios foram levantados mais com o intuito de permitir a visualização de contratos de aplicação de cada critério em documentos distintos.

Por meio do Quadro 19, pode-se perceber a taxa de revocação e precisão dos sintagmas quando da aplicação desse critério de eliminação dos SNs que possuíam numerais. Ao comparar os SNs descritores que não foram eliminados pelo critério de descarte com o total de SNs descritores (com ou sem numerais), verificou-se, em média, uma boa taxa de revocação, chegando a 96,1%. Em contrapartida, a precisão alcançada quando da comparação dos SNs descritores que não foram eliminados pelo critério com o total de sintagmas (descritores ou não), excetuando apenas os excluídos pelo critério, foi, em média, de 24,8%. A média de precisão sem aplicação do critério é de 25%, isso demonstra que a precisão alcançada com esse critério se mostrou semelhante à média de precisão sem uso do critério, não alcançando uma precisão acima da média de 25% na análise por documento.

A aplicação desse critério se mostrou uniforme na grande maioria dos documentos analisados, com ressalva, apenas para o DOC. 9 e o DOC. 26, onde, o referido critério obteve taxas mais baixas de revocação, pois eliminou de forma considerável SNs descritores, como por exemplo, o SN: “a Constituição Federal de 1988”.

Boa parte dos SNs eliminados, como foi mencionado, são sintagmas que possuem em suas estruturas números de leis e incisos. Uma saída para essa questão de descarte dos sintagmas com números em suas estruturas para esse domínio estudado seria a utilização de uma heurística que não descartasse os SNs que contivessem números após as expressões “lei(s)”, “artigo (s)”, “inciso(s)(em algarismo romanos)” e “parágrafos”, tendo em vista que o critério eliminou Sintagmas que possuíam essas características.

Apesar de ter alcançado boas taxas de revocação, considerou-se que esse critério, de modo geral, não se mostrou útil na seleção de SNs descritores, uma vez que alcançou precisões abaixo da média alcançada sem aplicação do critério. Sendo assim, com base nas taxas de revocação e precisão alcançadas tanto na análise por SN quanto por documento, verifica-se que o critério em questão não se mostrou tão útil para a seleção de sintagmas nominais descritores, ao passo que não conseguiu eliminar uma boa quantidade de SNs não descritores e evitar o descarte de SNs descritores. Por fim, pode-se concluir que o critério não é útil para a seleção de SNs, uma vez que, de todas as eliminações com esse critério, 40,5 % eram SNs descritores.

É pertinente mencionar que o comportamento desse critério de forma não tão eficiente, ao passo que eliminou uma boa quantidade de SNs descritores, se deu pelo fato de existir algumas características próprias desse domínio. Isso não significa que esses critérios se comportará da mesma forma em outros domínios, uma vez que nem todas as disciplinas (áreas de conhecimento) fazem uso exacerbado de números como o estudado desta pesquisa.

Nesse domínio específico, as buscas por documentos são feitas muitas vezes com bases em números, ou seja, os usuários fazem buscas por meio de números de leis. Assim, o fato de esse critério eliminar todos os SNs com números, seja na forma por extenso ou por meio de dígitos numéricos, conduziria a uma restrição nas possibilidades de buscas para usuários deste domínio, pois não seria possível que um determinado indivíduo fizesse uma busca, por exemplo, de um documento por meio da expressão “lei 8.112/90”.

4.2 Descarte de SNs que possuem Pronome como Núcleo

A aplicação desse critério ao *corpus* buscou verificar a viabilidade do mesmo ao contribuir ou não para a seleção de SNs, refinando assim as listas de sintagmas para fins de seleção dos melhores sintagmas. Lopes (2012) fez uso desse critério para a seleção de conceitos tomando por base os SNs. Na pesquisa dessa autora, esse critério descartou 6,109 sintagmas de um total de 189,146 sintagmas nominais, abrangendo um percentual de 3,2% do total de sintagmas.

Levando em consideração os 30 documentos utilizados nesta pesquisa, a aplicação desse critério alcançou um percentual de 10%, correspondendo a exatamente 3 dos 30 documentos.

A baixa frequência de aplicação desse critério pode ser compreendida quando se leva em consideração os elementos textuais dos quais os sintagmas foram extraídos, ou seja, utilizaram-se os títulos e os resumos das dissertações e teses para compor o *corpus* desta pesquisa, e, comumente, a utilização de pronomes como núcleo de sintagmas é encontrada no corpo do trabalho, na parte textual do trabalho, onde os autores dissertam e fazem uso das construções anafóricas, utilizando, assim, os pronomes como núcleos dos sintagmas.

No Quadro 20 são expostos os quantitativos de SNs descritores e não descritores que atenderam a esse critério, bem como as taxas de revocação e precisão, levando em consideração o percentual de Sintagmas descritores excluídos.

Quadro 20 – Quantitativo de SNs descritores e não descritores eliminados pelo critério: descarte de SNs com pronomes.

CRITÉRIO DE ELIMINAÇÃO (PRONOMES) COLEÇÃO	TOTAL DE SNs DESCRITORES	TOTAL DE SNs não DESCRITORES	Percentual de SNs DESCRITORES
Quantidade de SNs que não atendem ao critério	423	1355	23,7%
Quantidade de SNs que atendem ao critério	0	3	0%
TAXAS DE REVOCAÇÃO E PRECISÃO			
REVOCAÇÃO: 100%			
PRECISÃO: 23,7%			

Fonte: desenvolvido pelo autor.

Por meio do Quadro 20, pode-se ver uma excelente taxa de revocação e uma taxa de precisão pouco acima da média encontrada quando da não aplicação de nenhum critério na análise por sintagma nominal que é de 23,75%. No entanto, apesar de apresentar uma precisão baixa, os SNs que foram eliminados por este critério, demonstrado no Quadro 20, são sintagmas não descritores, o que demonstra, por sua vez que o critério em questão eliminou apenas sintagmas que não eram descritores. Os sintagmas eliminados por este critérios são: “aquela idealizada”, “aqueles previstos” e “eles equiparadas”. Com base na estrutura dos referidos sintagmas, percebe-se que os mesmos fazem referencia algum termo que já foi expresso em outro momento do texto, não sendo, assim, autocontido, ou seja, não possuem sentido específico, não conseguindo eles sós significar algo.

Com base no exposto, é notório que esse critério se comportou de forma efetiva, ao passo que eliminou SNs irrelevantes. Por meio dos dados demonstrados no quadro acima, acredita-se que esse critério se mostrou útil, contribuindo para a seleção de SNs nesse experimento específico.

Mais adiante, no Quadro 21, são expostas as médias de revocação e precisão para cada documento do corpus deste trabalho, bem como o percentual de aplicação desse critério em relação à quantidade de SNs únicos de cada documento.

Quadro 21 – Médias de revocação e precisão / Critério: Descarte de SNs com pronome no núcleo.

QUADRO RESUMO DE APLICAÇÃO D CRITÉRIO DE ELIMINAÇÃO DE SNs COM PRONOMES				
DOCUMENTO	QUANTIDADE DE SNs ELIMINADOS	PERCENTUAL DE APLICAÇÃO DO CRITÉRIO EM RELAÇÃO SNs UNICOS	REVOCAÇÃO	PRECISÃO
DOC. 1	0	0,0%	100,0%	26,4%
DOC. 2	0	0,0%	100,0%	27,0%
DOC. 3	1	1,7%	100,0%	21,8%
DOC. 4	1	2,0%	100,0%	26,5%
DOC. 5	0	0,0%	100,0%	11,1%
DOC. 6	0	0,0%	100,0%	54,9%
DOC. 7	0	0,0%	100,0%	11,4%
DOC. 8	0	0,0%	100,0%	31,3%
DOC. 9	0	0,0%	100,0%	68,9%
DOC. 10	0	0,0%	100,0%	26,3%
DOC. 11	0	0,0%	100,0%	19,4%
DOC. 12	1	3,0%	100,0%	18,7%
DOC. 13	0	0,0%	100,0%	20,0%
DOC. 14	0	0,0%	100,0%	25,0%
DOC. 15	0	0,0%	100,0%	16,3%
DOC. 16	0	0,0%	100,0%	15,2%
DOC. 17	0	0,0%	100,0%	36,2%
DOC. 18	0	0,0%	100,0%	36,1%
DOC. 19	0	0,0%	100,0%	38,2%
DOC. 20	0	0,0%	100,0%	18,0%
DOC. 21	0	0,0%	100,0%	26,5%
DOC. 22	0	0,0%	100,0%	10,2%
DOC. 23	0	0,0%	100,0%	25,9%
DOC. 24	0	0,0%	100,0%	22,0%
DOC. 25	0	0,0%	100,0%	14,2%
DOC. 26	0	0,0%	100,0%	9,1%
DOC. 27	0	0,0%	100,0%	20,4%
DOC. 28	0	0,0%	100,0%	31,5%
DOC. 29	0	0,0%	100,0%	26,4%
DOC. 30	0	0,0%	100,0%	17,0%

MEDIA	0,1	0,2%	100,0%	25,1%
DESVIO PADRÃO	0,3051286	0,7%	0,0%	12,8%
MEDIANA	0	0,0%	100,0%	23,5%

Fonte: desenvolvido pelo autor.

Verificou-se uma taxa de revocação média de 100%, ao comparar os sintagmas descritores que não foram eliminados pelo critério com o total de SNs descritores. Já a precisão, em média, foi de 25,1% quando da comparação dos sintagmas descritores que não foram eliminados pelo critério com o total de sintagmas (descritores ou não). As taxas de revocação e precisão alcançadas com esse critério são as que mais se aproximam das taxas obtidas quando da não aplicação desse critério. Em suma, o critério se mostrou viável e de utilidade para a seleção de SNs com valor de descritores, ao passo que manteve a revocação de 100% e a precisão na média da não aplicação do critério.

4.3 Descarte de SNs que Iniciam com Advérbios

Esse critério tem como intuito, segundo Lopes (2012), eliminar os sintagmas que iniciam com advérbios, levando em consideração que alguns sintagmas que possuem advérbio, geralmente, fazem referência a um termo já mencionado não possuindo autonomia significativa, não sendo útil para servir como conceito. Dessa forma, buscou-se verificar a viabilidade desse critério no que se refere a seleção ou exclusão de Sintagmas Descritores.

Durante a análise de aplicação desse critério em todos os sintagmas, percebeu-se que o mesmo teria uma frequência de aplicação maior se os artigos (determinantes) fossem excluídos dos sintagmas como foi feito por Lopes (2012). Sendo assim, procedeu-se com a reanálise de aplicação desse critério novamente, no entanto, considerando, desta vez, além dos SNs que iniciavam diretamente com advérbio, os SNs que iniciavam com determinante seguido de um advérbio, simulando, assim, que os determinantes tinham sido removidos igualmente ao que foi feito por Lopes (2012). No contexto desta pesquisa, durante a aplicação desse critério, considerou-se os “SNs que iniciavam com advérbio” todos os SNs que iniciaram diretamente com advérbio ou que iniciaram com determinante mais advérbio.

Mais adiante, no Quadro 22, são expostos os quantitativos de SNs descritores e não descritores que atenderam a esse critério, bem como as taxas de revocação e precisão, levando em consideração o percentual de Sintagmas descritores excluídos. A taxa de revocação do

quadro que segue reflete o percentual de SNs descritores selecionados do total de SNs descritores e a precisão reflete o percentual de SNs descritores selecionados do total de SNs selecionados.

Quadro 22 – Quantitativo de SNs descritores e não descritores eliminados pelo critério: descarte de SNs que iniciam com advérbio.

CRITÉRIO DE ELIMINAÇÃO (ADVÉRBIO) COLEÇÃO	TOTAL DE SNs DESCRITORES	TOTAL DE SNs não DESCRITORES	Percentual de SNs DESCRITORES
Quantidade de SNs que não atendem ao critério	420	1358	23,6%
Quantidade de SNs que atendem ao critério	3	0	100%
TAXAS DE REVOCAÇÃO E PRECISÃO			
REVOCAÇÃO: 99,2%			
PRECISÃO: 23,6%			

Fonte: desenvolvido pelo autor.

Por meio dos dados expostos no Quadro 22, pode-se perceber que o critério de eliminação foi aplicado em apenas três sintagmas, os quais eram SNs descritores, ou seja, de todos os SNs eliminados pelo referido critério, 100% era de sintagmas descritores, o que vem a demonstrar que o critério eliminou SNs que não deveriam ser eliminados, pois apesar de iniciarem com advérbios, eles eram descritores. As taxas de revocação e precisão, principalmente a de precisão confirmam que o critério em questão se comportou de forma negativa, onde a taxa de precisão ficou em 23,6%, ou seja, abaixo da média obtida sem a aplicação de critério que foi de 23,75%.

Conclui-se que, apesar de obter uma baixa frequência de aplicação, esse critério se mostrou falho ao eliminar SNs descritores. Logo abaixo, são expostos os SNs descritores que foram eliminados por este critério:

Quadro 23 – SNs descritores eliminados pelo critério de descarte de SNs com advérbio no início.

SINTAGMAS ELIMINADOS POR INICIAREM DIRETAMENTE POR ADVÉRBIO OU POR MEIO DE DETERMINANTE MAIS ADVÉRBIO.
apenas três modos de entidades familiares
a mais flagrante violação da isonomia tributária
mais variados setores da administração pública brasileira nos últimos seis anos

Fonte: desenvolvido pelo autor.

Pode-se perceber, por meio do Quadro 23, que os três SNs descritores se referem a conceitos pertinentes, mesmo iniciando por advérbio. Apesar de esses SNs fazerem referência

a conceitos, o fato de eles iniciarem por um advérbio contribuiu para a eliminação dos mesmos. Apesar de saber que, convencionalmente, os SNs que iniciam com advérbios fazem referência a termos previamente mencionados no texto, não possuindo sentido próprio, neste caso específico, os três sintagmas não se encaixaram nesse parâmetro, pois como pode ser analisado apenas o terceiro sintagma faz referência a algo que já foi mencionado no texto que é justamente em relação aos “mais variados setores da administração pública...”.

Depois de levantados os dados quantitativos acerca dos SNs descritores ou não eliminados pelo critério de descarte de SNs que iniciam com advérbio, como forma de complementar a análise feita desse critério, elaborou-se um quadro resumo contendo a quantidade de SNs eliminados por este critério em cada documento, o percentual em relação à quantidade de sns únicos de cada documento, bem como as taxas de revocação e precisão alcançadas com a aplicação desse critério em cada documento de forma individual.

No que se refere às taxas de revocação e precisão expostas no quadro a seguir, para o cálculo da revocação dividiu-se a quantidade de sintagmas descritores que não tinham sido eliminados por este critério pelo total de sintagmas descritores no documento. A precisão foi alcançada dividindo-se a quantidade de SNs descritores que não tinha sido eliminados pelo critério pelo total de sintagmas (descritores ou não) do documento, excluindo-se os eliminados pelo critério. No Quadro 24 são expostas as taxas de revocação e precisão, bem como as médias alcançadas.

Quadro 24 – Médias de revocação e precisão / Critério: Descarte de SNs que iniciam com advérbio.

QUADRO RESUMO DE APLICAÇÃO D CRITÉRIO DE ELIMINAÇÃO DE SNs COM ADVÉRBIOS				
DOCUMENTO	QUANTIDADE DE SNs ELIMINADOS	PERCENTUAL DE APLICAÇÃO DO CRITÉRIO EM RELAÇÃO SNs UNICOS	REVOCAÇÃO	PRECISÃO
DOC. 1	0	0,0%	100,0%	26,0%
DOC. 2	0	0,0%	100,0%	27,0%
DOC. 3	0	0,0%	100,0%	21,4%
DOC. 4	1	2,0%	92,3%	24,4%
DOC. 5	0	0,0%	100,0%	11,1%
DOC. 6	0	0,0%	100,0%	54,9%
DOC. 7	0	0,0%	100,0%	11,4%
DOC. 8	0	0,0%	100,0%	31,3%

DOC. 9	0	0,0%	100,0%	68,9%
DOC. 10	1	1,7%	93,3%	26,3%
DOC. 11	0	0,0%	100,0%	19,4%
DOC. 12	0	0,0%	100,0%	18,0%
DOC. 13	0	0,0%	100,0%	20,0%
DOC. 14	0	0,0%	100,0%	25,0%
DOC. 15	1	1,6%	90,0%	15,0%
DOC. 16	0	0,0%	100,0%	15,2%
DOC. 17	0	0,0%	100,0%	36,2%
DOC. 18	0	0,0%	100,0%	36,1%
DOC. 19	0	0,0%	100,0%	38,2%
DOC. 20	0	0,0%	100,0%	18,8%
DOC. 21	0	0,0%	100,0%	26,5%
DOC. 22	0	0,0%	100,0%	10,2%
DOC. 23	0	0,0%	100,0%	25,9%
DOC. 24	0	0,0%	100,0%	22,0%
DOC. 25	0	0,0%	100,0%	14,2%
DOC. 26	0	0,0%	100,0%	9,1%
DOC. 27	0	0,0%	100,0%	20,2%
DOC. 28	0	0,0%	100,0%	31,5%
DOC. 29	0	0,0%	100,0%	26,4%
DOC. 30	0	0,0%	100,0%	17,0%
MEDIA	0,1	0,2%	99,2%	24,9%
DESVIO PADRÃO	0,305129	0,5%	2,5%	12,8%
MADIANA	0	0,0%	100,0%	23,2%

Fonte: desenvolvido pelo autor.

O Quadro 24, logo acima, confirma o que foi visto no Quadro 22, onde identificou-se uma precisão abaixo da média alcançada quando da não aplicação desse critério. Em suma, o critério de eliminação de SNs que iniciam com advérbios não foi útil nesse experimento, pois eliminou SNs descritores, onde não o deveria ter feito.

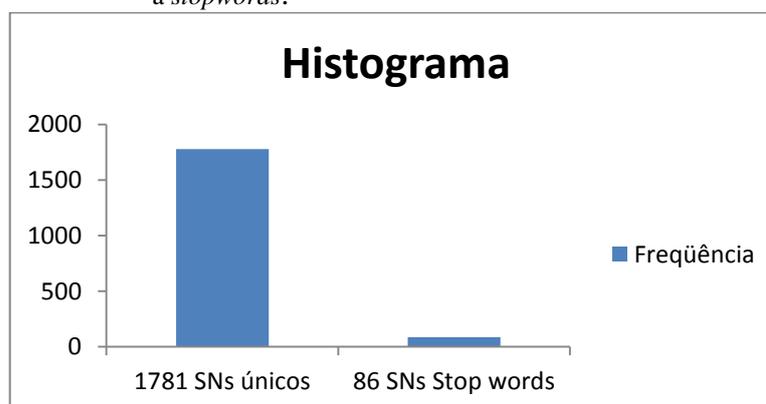
4.4 Eliminação de SNs Não Descritores categorizados como *Stop Words*

Termos muito frequentes que ocorrem comumente em diversos documentos tendem a constituir expressões que são generalistas e possuem pouco poder discriminatório. Geralmente essas expressões são palavras ou conjunto de palavras que são comumente encontradas em textos científicos, mas que não servem como descritores. Assim, essas expressões convencionalmente usuais na língua constituem uma lista de termos que devem ser

desconsiderados nas pesquisas voltadas para a indexação automática. Essas expressões podem ser consideradas como termos de pouco potencial discriminatório, e que é recorrente na literatura científica.

Durante o processo de aplicação dos critérios utilizados nesta pesquisa, cada SN foi analisado manualmente com vistas a verificar a sua relevância para fins de construção de uma lista de *stopwords*, ou seja, ao passo que se aplicavam cada critério de seleção, ia-se marcando as expressões que comporiam as *stopwords*, ou seja, termos comuns em textos científicos, e que são irrelevantes para indexação e recuperação de documentos. De um total de 1781 sintagmas únicos, foram identificadas 86 expressões que eram frequentes e bastante comuns em textos científicos e em textos do domínio estudado nesta pesquisa, e, que, conseqüentemente, em nada contribuiriam para a representação temática dos documentos analisados. Assim 4,94% dos sintagmas únicos constituíam-se de sintagmas não relevantes que compuseram a *stopwords*, como pode ser visto mais adiante na Figura 8.

Figura 8 – Quantidade de SNs únicos e de expressões que compuseram a *stopwords*.



Fonte: desenvolvido pelo autor.

Segue no apêndice C desta dissertação uma lista com esses sintagmas que foram eliminados por serem comuns e possuírem pouco poder discriminatório, salientando que estas *stopwords* devem ser consideradas apenas no contexto deste trabalho. Esta lista, apesar de ser pequena, devido ao corpus trabalhado, serve como ponto de partida para a elaboração de uma lista de *stopwords* mais extensa e completa e que possa ser utilizada por pesquisas que utilizem esse domínio específico, uma vez que a eliminação de *stopwords* em métodos de indexação contribui efetivamente para a eliminação de palavras ou expressões sem relevância conceitual.

Mais adiante, no Quadro 25, são expostos os quantitativos de SNs descritores e não descritores que atenderam a esse critério, bem como as taxas de revocação e precisão, levando em consideração o percentual de Sintagmas descritores excluídos. A taxa de revocação e precisão expressa no quadro abaixo refletem, respectivamente, o percentual de SNs descritores selecionados do total de SNs descritores e o percentual de SNs descritores selecionados do total de SNs selecionados

Quadro 25 – Quantitativo de SNs descritores e não escritores eliminados pelo critério: descarte de SNs Stopwords.

CRITÉRIO DE ELIMINAÇÃO (STOP WORDS) COLEÇÃO	TOTAL DE SNs DESCRITORES	TOTAL DE SNs não DESCRITORES	Percentual de SNs DESCRITORES
Quantidade de SNs que não atendem ao critério	423	1272	25%
Quantidade de SNs que atendem ao critério	0	86	0%
TAXAS DE REVOCAÇÃO E PRECISÃO			
REVOCAÇÃO: 100%			
PRECISÃO: 25%			

Fonte: desenvolvido pelo autor.

Por meio dos dados explicitados no Quadro 25, pode-se perceber a taxa de revocação igual e a de precisão pouco superior à taxa alcançada com a não aplicação de nenhum critério (23,75%), o que confirma por sua vez o que a literatura diz acerca do uso de lista de stopwords em sistemas de indexação automática, ou seja, a utilização de listas de *stopwords* só contribui para a seleção de termos de indexação. Por meio da quantificação dos SNs excluídos, verifica-se que todos os SNs eliminados constituíam realmente sintagmas não descritores. Assim, esse critério eliminou apenas SNs não descritores, mostrando-se assim de total utilidade.

Por meio do Quadro 25 fica visível que de todos os SNs eliminados, todos eram categorizados como *stopwords*, ou seja, não serviam como descritores.

Além dos dados alcançados por meio dos cálculos e da quantificação realizada anteriormente, procedeu-se com a elaboração de um quadro resumo constituído pela quantidade de SNs eliminados de cada documento por esse critério, o percentual de aplicação do referido critério em relação ao quantitativo de SNs únicos de cada documento, bem como a taxa de revocação e precisão de cada documento. Esses dados são expostos mais adiante no Quadro 26.

Quadro 26 – Médias de revocação e precisão / Critério: uso de Stop words.

QUADRO RESUMO DE APLICAÇÃO D CRITÉRIO DE ELIMINAÇÃO DE SNs /STOPWORDS				
DOCUMENTO	QUANTIDADE DE SNs ELIMINADOS	PERCENTUAL DE APLICAÇÃO DO CRITÉRIO EM RELAÇÃO SNs UNICOS	REVOCAÇÃO	PRECISÃO
DOC. 1	0	0,0%	100,0%	26,4%
DOC. 2	4	10,8%	100,0%	30,3%
DOC. 3	3	5,3%	100,0%	22,6%
DOC. 4	1	2,0%	100,0%	26,5%
DOC. 5	3	66,0%	100,0%	11,9%
DOC. 6	1	1,9%	100,0%	56,0%
DOC. 7	1	1,4%	100,0%	11,5%
DOC. 8	3	5,8%	100,0%	33,3%
DOC. 9	3	10,3%	100,0%	76,9%
DOC. 10	2	3,5%	100,0%	27,2%
DOC. 11	9	12,5%	100,0%	22,2%
DOC. 12	1	3,0%	100,0%	18,7%
DOC. 13	8	13,3%	100,0%	23,0%
DOC. 14	3	5,0%	100,0%	26,3%
DOC. 15	4	6,5%	100,0%	17,5%
DOC. 16	1	1,3%	100,0%	15,4%
DOC. 17	2	2,8%	100,0%	35,4%
DOC. 18	2	2,7%	100,0%	37,1%
DOC. 19	4	8,5%	100,0%	41,8%
DOC. 20	7	8,4%	100,0%	19,2%
DOC. 21	0	0,0%	100,0%	26,5%
DOC. 22	1	1,4%	100,0%	10,4%
DOC. 23	5	9,2%	100,0%	28,5%
DOC. 24	2	2,9%	100,0%	22,7%
DOC. 25	5	8,9%	100,0%	15,6%
DOC. 26	1	1,1%	100,0%	9,3%
DOC. 27	2	22,0%	100,0%	20,9%
DOC. 28	4	7,0%	100,0%	33,9%
DOC. 29	3	4,4%	100,0%	27,6%
DOC. 30	1	2,1%	100,0%	17,3%
MEDIA	2,866666667	7,7%	100,0%	26,4%
DESVIO PADRÃO	2,224213667	12,0%	0,0%	13,9%
MEDIANA	2,5	4,7%	100,0%	24,7%

Fonte: desenvolvido pelo autor.

É possível perceber que esse critério alcançou um dos melhores resultados, tanto no que se refere à revocação quanto à precisão. A média da taxa de revocação ficou igual à taxa alcançada quando da não aplicação do critério e a precisão alcançou uma média 26,4%, ou seja, acima da taxa de 25%.

Em suma, verifica-se que a utilização desse critério é fundamental em qualquer sistema que trabalhe estatisticamente com seleção de termos. Vários autores fazem uso de *stopwords/stoplist*, como, por exemplo, Martins (2014), onde este autor em sua metodologia prospectiva fez uma filtragem preliminar dos documentos que compuseram seu *corpus*, eliminando algumas expressões recorrentes na literatura científica, mas que nada contribuíam para a qualidade dos SNs. A eliminação de *stopwords* é fundamental, pois permite a previa eliminação de termos usuais e bastante comuns em textos científicos, contribuindo para que fique os SNs que realmente contribuam para a indexação ou classificação, ou qualquer outro fim. A utilização desse recurso refina os sintagmas, retirando os que não serviriam como termos de indexação, limpando consideravelmente os sintagmas irrelevantes de um grupo de sintagmas.

Com base em todos os dados alcançados, mencionados anteriormente, conclui-se que esse critério não é só útil como também essencial em sistemas de indexação automática.

4.5 Detecção de SNs contidos em SNs maiores por meio da remoção sucessiva de Adjetivos

No contexto desse experimento, a aplicação desse critério buscou verificar se o mesmo contribuía ou não para a seleção de sintagmas descritores. Assim, o intuito desse critério, neste experimento, foi identificar a relação dos sintagmas que possuíam sintagmas menores em suas estruturas que eram SNs descritores, inferindo que esses sintagmas além de gerarem novos sintagmas, possuem potencial discriminatório.

Este critério não é de eliminação, assim, os cálculos para obtenção das medidas de revocação e precisão foram diferenciados, pois neste caso os SNs retornados são os que o critério é aplicado, sendo assim, critérios de adição. A análise feita deste critério também difere um pouco da análise feita dos critérios anteriores, os quais eram de eliminação.

Esse critério leva em consideração os sintagmas que possuem adjetivos ou verbos no participio passado ao final de cada termo, possuindo assim sintagmas implícitos em suas estruturas. Tomando como exemplo o SN: “a receita estatal tributária”, onde seriam gerados os seguintes SNs: “a receita tributária” e a “a receita estatal”. No Sintagma “a receita estatal

tributária” tem-se dois adjetivos para o nome “receita”, gerando assim sintagmas implícitos por meio da remoção desses adjetivos após o nome (substantivo). Pode-se ver no quadro a seguir a quantidade de SNs que seriam gerados com a aplicação desse critério.

Quadro 27 – Quantidade de SNs que seriam gerados com a aplicação do critério de adição de SNs com múltiplos adjetivos.

QUANTIDADE DE SNs ÚNICOS	QUANTIDADE DE VEZES QUE O CRITÉRIO FOI APLICADO	PORCENTAGEM	QUANTIDADE DE SNs QUE SERIAM GERADOS
1781	494	27,7%	550

Fonte: desenvolvido pelo autor.

Mais adiante, no Quadro 28, são expostos os quantitativos de SNs descritores e não descritores que atenderam a esse critério, bem como as taxas de revocação e precisão. Essas taxas refletem, respectivamente, o percentual de SNs descritores selecionados do total de SNs descritores e o percentual de SNs descritores selecionados do total de SNs selecionados.

Quadro 28 – Quantitativo de SNs descritores e não descritores selecionados pelo critério: detecção de SNs por meio da remoção adjetivos múltiplos.

CRITÉRIO DE ADIÇÃO (MÚLTIPLOS ADJETIVOS) COLEÇÃO	TOTAL DE SNs DESCRITORES	TOTAL DE SNs não DESCRITORES	Percentual de SNs DESCRITORES
Quantidade de SNs que não atendem ao critério	301	986	23,3%
Quantidade de SNs que atendem ao critério	122	372	24,7%
TAXAS DE REVOCAÇÃO E PRECISÃO			
REVOCAÇÃO: 28,8%			
PRECISÃO: 24,7%			

Fonte: desenvolvido pelo autor.

Com base nos dados expostos no Quadro 28, percebe-se que de todos os SNs selecionados por este critério, ou seja, de todos os SNs que possuíam outros SNs implícitos em suas estruturas, que atenderam ao critério de adição, 24,7% deles eram SNs descritores, o que deixa inferir que uma parte razoável dos SNs descritores possui potencial para gerar novos SNs. Ainda com base no quadro anterior, pode-se perceber que as taxas de revocação e precisão foram baixas, onde a revocação ficou em torno de 28,8%, ou seja, de todos os SNs descritores, apenas 28,8% eram SNs descritores que possuíam outros SNs implícitos em suas

estruturas. A precisão, da mesma forma que a revocação, também ficou baixa, em torno de 24,7%, apesar de superior à média da não aplicação deste critério que foi de 23,75%.

Como forma de complementar os dados obtidos no quadro anterior, prosseguiu-se com a elaboração de um Quadro resumo, onde foram expostos a quantidade de SNs selecionados em cada documento pelo critério em questão, bem como também o percentual de aplicação desse critério em cada documento, levando em consideração a quantidade de SN únicos de cada documento e também as taxas de revocação e precisão em cada documento. Para o cálculo da revocação, dividiu-se o quantitativo de sintagmas descritores que se encaixavam nesse critério, ou seja, que possuíam mais de um adjetivo qualificando um mesmo nome, pela quantidade total de sintagmas descritores do documento (independente de se encaixarem ou não nesse critério). Já a precisão foi alcançada, dividindo-se a quantidade de SNs descritores que se encaixavam nesse critério pela quantidade total de sintagmas (descritores ou não) que se encaixavam nesse critério. No quadro 29, logo abaixo, são expostos os referidos percentuais.

Quadro 29 – Médias de revocação e precisão / Critério: Detecção de SNs com múltiplos adjetivos.

QUADRO RESUMO DE APLICAÇÃO D CRITÉRIO DE ADIÇÃO. ADJETIVOS MÚLTIPLOS				
DOCUMENTO	QUANTIDADE DE SNs SELECIONADOS	PERCENTUAL DE APLICAÇÃO DO CRITÉRIO EM RELAÇÃO SNs UNICOS	REVOCAÇÃO	PRECISÃO
DOC. 1	11	32,3%	44,4%	36,3%
DOC. 2	16	43,2%	50,0%	31,2%
DOC. 3	21	23,2%	16,6%	15,3%
DOC. 4	17	34,0%	38,4%	29,4%
DOC. 5	6	13,3%	20,0%	16,6%
DOC. 6	13	25,4%	21,4%	46,1%
DOC. 7	18	25,7%	25,0%	11,1%
DOC. 8	10	19,6%	18,7%	30,0%
DOC. 9	6	20,6%	25,0%	83,3%
DOC. 10	22	38,5%	20,0%	13,6%
DOC. 11	22	30,5%	28,5%	18,1%
DOC. 12	9	27,2%	33,3%	22,2%
DOC. 13	15	38,3%	83,3%	43,4%
DOC. 14	14	23,3%	0,0%	0,0%
DOC. 15	22	36,0%	70,0%	31,8%

DOC. 16	17	23,6%	27,2%	17,6%
DOC. 17	15	21,7%	20,0%	33,3%
DOC. 18	15	20,8%	19,2%	33,3%
DOC. 19	11	23,4%	27,7%	45,4%
DOC. 20	23	27,7%	53,3%	34,7%
DOC. 21	21	26,5%	23,8%	23,8%
DOC. 22	20	29,4%	42,8%	15,0%
DOC. 23	17	31,4%	14,2%	11,7%
DOC. 24	11	16,1%	6,6%	9,0%
DOC. 25	19	33,9%	50,0%	21,0%
DOC. 26	26	29,8%	62,5%	19,2%
DOC. 27	24	27,2%	33,3%	25,0%
DOC. 28	22	38,5%	27,7%	22,7%
DOC. 29	13	19,1%	16,6%	23,0%
DOC. 30	18	38,2%	37,5%	16,6%
MÉDIA	16,46666667	27,9%	31,9%	26,0%
DESVIO PADRÃO	5,322258617	7,4%	18,6%	15,5%
MEDIANA	17	27,2%	27,5%	22,9%

Fonte: desenvolvido pelo autor.

A taxa de revocação alcançada com a aplicação desse critério chega a 31,9%. Em contrapartida, a precisão se manteve próximo à taxa alcançada quando da não aplicação do critério, a qual gira em torno de 25%. Dos trinta (30) documentos, em que esse critério foi aplicado, apenas em um (1), no DOC. 14, a taxa de revocação e precisão foi de 0%. Isso se deu pelo fato de que esse critério não foi aplicado a nenhum SN descritor. Esse documento foi a exceção de todos os outros. Outro documento específico que se destacou, não por ter revocação e precisão “0%”, e sim por ter alcançado uma alta precisão foi o DOC. 9, por meio do qual verificou-se uma taxa de precisão de 83,3%. Esse fato se deu, pois, dos seis SNs descritores selecionados desse documento, cinco eram SNs descritores que possuíam outros sintagmas em suas estruturas, ou seja, cinco SNs descritores foram marcados pelo critério de adição de SNs implícitos.

Levando em consideração as taxas de revocação e precisão dos dois quadros anteriores, verificou-se que tanto a revocação quanto a precisão foram baixas, demonstrando que o percentual de SNs descritores que atendem e não atendem ao critério são próximos, fazendo inferir que esse critério não ajuda a separar SNs descritores de SNs não descritores.

Em suma, este critério não se mostrou útil ao passo que obteve um resultado relativamente ruim, demonstrando assim que não influencia tanto à seleção de SNs descritores.

4.6 Identificação de SNs por meio de Conjunção entre Adjetivos

A aplicação desse critério, no contexto desse experimento, tem o intuito de verificar a incidência de Sintagmas que possuem outros sintagmas implícitos embutidos em suas estruturas por meio de adjetivos ligados por conjunção, subtendendo-se que esses sintagmas podem possuir poder discriminatório para fins de indexação.

A frequência de aplicação desse critério foi bem mais baixa do que a do critério anterior, como pode ser vista no quadro a seguir:

Quadro 30 – Quantidade de SNs que seriam gerados com a aplicação do critério de adição de SNs com múltiplos adjetivos ligados por conjunção.

QUANTIDADE DE SNs ÚNICOS	QUANTIDADE DE VEZES QUE O CRITÉRIO FOI APLICADO	PORCENTAGEM	QUANTIDADE DE SNs QUE SERIAM GERADOS
1781	22	1,19%	22

Fonte: desenvolvido pelo autor.

Esse critério, bem como o anterior, como demonstrado por Lopes (2012) se mostram pertinentes quando o interesse é a geração de novos SNs. Mais adiante, no Quadro 31, são expostos os quantitativos de SNs descritores e não descritores que atenderam a esse critério, bem como as taxas de revocação e precisão. Essas taxas expressas no quadro abaixo refletem, respectivamente, o percentual de SNs descritores selecionados do total de SNs descritores e o percentual de SNs descritores selecionados do total de SNs selecionados.

Quadro 31 – Quantitativo de SNs descritores e não descritores selecionados pelo critério: detecção de SNs por meio de múltiplos adjetivos ligados por conjunção.

CRITÉRIO DE ADIÇÃO (MÚLTIPLOS ADJETIVOS LIGADOS POR CONJUNÇÃO) COLEÇÃO	TOTAL DE SNs DESCRITORES	TOTAL DE SNs não DESCRITORES	Percentual de SNs DESCRITORES
Quantidade de SNs que não atendem ao critério	419	1340	23,8%
Quantidade de SNs que atendem ao critério	4	18	18,1%
TAXAS DE REVOCAÇÃO E PRECISÃO			
REVOCAÇÃO: 0,9%			
PRECISÃO: 18,1%			

Fonte: desenvolvido pelo autor.

O Quadro 31 demonstra que o percentual de SNs descritores que atenderam ao critério foi muito baixo, bem como também os SNs não descritores que atenderam a esse critério. A maior parte dos SNs descritores não atende ao critério, onde a taxa de revocação foi muito baixa ficando em torno de 0,9%. A precisão também se mostrou baixa, ao passo que alcançou apenas 18,1%, ou seja, de todos os SNs que atenderam ao critério, apenas quatro eram SNs descritores, demonstrando assim que esse critério não se mostrou tão efetivo em conseguir separar SNs descritores de SNs não descritores.

Como forma de complementar os dados obtidos com o Quadro 31, elaborou-se um quadro resumo, no qual foram expostos a quantidade de SNs selecionados em cada documento do corpus com a aplicação desse critério, o percentual de aplicação do critério em relação à quantidade de SNs únicos de cada documento, bem como também as taxas de revocação e precisão de cada documento. Essas taxas foram alcançadas, respectivamente, dividindo-se a quantidade de SNs descritores que atenderam ao critério pelo total de SNs descritores de cada documento e dividindo a quantidade de SNs descritores que atenderam ao critério pelo total de SNs selecionados pelo critério em questão.

Quadro 32 – Médias de revocação e precisão / Critério: Detecção de SNs com múltiplos adjetivos ligados por conjunção.

DOCUMENTO	QUANTIDADE DE SNs SELECIONADOS	PERCENTUAL DE APLICAÇÃO DO CRITÉRIO EM RELAÇÃO SNs UNICOS	REVOCAÇÃO	PRECISÃO
DOC. 1	0	0,0%	0,0%	0,0%
DOC. 2	0	0,0%	0,0%	0,0%
DOC. 3	0	0,0%	0,0%	0,0%
DOC. 4	0	0,0%	0,0%	0,0%
DOC. 5	0	0,0%	0,0%	0,0%
DOC. 6	0	0,0%	0,0%	0,0%
DOC. 7	0	0,0%	0,0%	0,0%
DOC. 8	0	0,0%	0,0%	0,0%
DOC. 9	0	0,0%	0,0%	0,0%
DOC. 10	0	0,0%	0,0%	0,0%
DOC. 11	1	1,3%	0,0%	0,0%
DOC. 12	3	9,0%	0,0%	0,0%
DOC. 13	1	1,6%	0,0%	0,0%
DOC. 14	4	6,6%	0,0%	0,0%
DOC. 15	0	0,0%	0,0%	0,0%

DOC. 16	0	0,0%	0,0%	0,0%
DOC. 17	0	0,0%	0,0%	0,0%
DOC. 18	1	1,3%	3,8%	100,0%
DOC. 19	1	2,1%	0,0%	0,0%
DOC. 20	1	1,2%	0,0%	0,0%
DOC. 21	1	1,2%	0,0%	0,0%
DOC. 22	0	0,0%	0,0%	0,0%
DOC. 23	1	1,8%	0,0%	0,0%
DOC. 24	0	0,0%	0,0%	0,0%
DOC. 25	0	0,0%	0,0%	0,0%
DOC. 26	6	6,8%	12,5%	16,6%
DOC. 27	0	0,0%	0,0%	0,0%
DOC. 28	0	0,0%	0,0%	0,0%
DOC. 29	2	2,9%	11,1%	100,0%
DOC. 30	0	0,0%	0,0%	0,0%
MEDIA	0,733333333	1,2%	0,9%	7,2%
DESVIO PADRÃO	1,387961376	2,3%	3,0%	25,4%
MEDIANA	0	0,0%	0,0%	0,0%

Fonte: desenvolvido pelo autor.

Ratificando os dados expostos no quadro 31, verifica-se uma baixa taxa de revocação e precisão, todavia há de se levar em consideração que o referido critério foi aplicado pouquíssimas vezes. Como pode ser visto, esse critério foi aplicado em onze documentos, sendo que desse total, dois se destacaram por terem alcançado altas taxas de precisão. Esses dois documentos foram o DOC. 18 e 29 com uma taxa de 100% de precisão. O DOC. 26 obteve uma precisão de 16,6%. No caso daqueles dois documentos, a precisão ficou em 100%, pois o critério foi aplicado a um único sintagma e este era um SN descritor, ou seja, de todos os SNs existentes nesse documento, apenas um possuía adjetivos ligados por conjunção, o qual era descritor. No DOC. 29 ocorreu do mesmo modo, onde o critério foi aplicado a dois SNs e estes eram descritores, ou seja, de todos os SNs existentes nesse documento, apenas dois possuem múltiplos adjetivos ligados por conjunção, e esses dois eram SNs descritores. Já no DOC. 26 ocorreu o contrário, onde de um total de seis (6) sintagmas que possuíam múltiplos adjetivos ligados por conjunção, ou seja, que atendiam ao critério, apenas um era SN descritor, resultando, assim, em uma precisão de 16,6%. Mesmo com alguns documentos alcançando individualmente resultados satisfatórios de precisão, não se pode abranger todos os outros documentos, considerando-os como um todo, até porque o desvio padrão foi bastante alto, o que demonstra, por sua vez, um alto grau de dispersão entre as taxas obtidas em cada documento.

Com base nas medias de revocação e precisão alcançadas com a aplicação desse critério, torna-se perceptível que o mesmo não contribuiu efetivamente para a seleção de SNs descritores. A taxa de precisão ficou bem abaixo da média quando da não aplicação desse critério e a revocação mais ainda, indicando que este critério não ajuda a separar SNs descritores de SNs não descritores. Conclui-se que esse critério não se mostrou útil para a seleção de SNs.

4.7 Estrutura e Nível do SN

A aplicação desse critério permitiu perceber a diferença entre sintagmas de nível 1, 2, 3, 4 e assim por diante, no que se refere ao poder discriminatório que cada sintagma possui. Não foi atribuído nenhum valor a cada SN, levando em consideração esses níveis, como proposto por Souza (2006). A atribuição do nível a cada Sintagma contribuiu para a verificação do potencial discriminatório que um sintagma de nível 2 ou 3, por exemplo, possui em comparação com um Sintagma de nível 1 formado por determinante e um nome, que é, na maioria dos casos, bem geral e comum, possuindo pouco poder discriminatório.

Mais adiante, encontram-se no Quadro 33 os dados quantitativos referentes aos SNs descritores e não descritores e seus respectivos níveis. Separaram-se os SNs em seis categorias, as quais são: nível 1a, nível 1b, nível 2, nível 3, nível 4 e nível 5 ou mais (+).

Quadro 33 – Quantitativo de SNs descritores e não descritores que atenderam ao critério de nível, sendo categorizados em nível 1a, 1b, 2, 3, 4 ou 5 ou mais.

NÍVEL	TOTAL DE SNS DESCRITORES	TOTAL DE SNS NÃO DESCRITORES	PERCENTUAL DE SNS DESCRITORES
1a	27	400	6,3%
1b	82	362	18,4%
2	135	378	26,3%
3	90	136	39,8%
4	41	51	44,5%
5+	48	31	60,7%

Fonte: desenvolvido pelo autor.

Como pode ser visto no Quadro 33, a maior parte dos SN descritores encontra-se nos SNs de nível 2, seguidos dos de nível 3 e nível 1b, demonstrando, assim, que os SNs de nível 2 tendem a possuir maior poder discriminatório do que SNs de outros níveis. Ao levar em consideração o percentual de SNs descritores, comparando a quantidade de SNs descritores e

não descritores de cada nível, verifica-se que os SNs de nível maior ou igual a cinco possuem um maior percentual de SNs descritores, alcançando um percentual de 60,7%.

Comparando a quantidade de SNs descritores de nível 1a e 1b, verifica-se, ainda com base no quadro anterior, que entre os SNs de nível 1a e 1b, a maior parte dos SNs descritores se encaixam nos SNs de nível 1b. A esse respeito, Souza (2006) em sua CSN (Categoria do SN) atribui um peso maior a esses Sintagmas de nível 1b. A Tabela 2 mostra alguns Sintagmas e seus respectivos níveis, demonstrando o potencial que um sintagma de nível 1 sem ter a estrutura “determinante + nome” possui em relação a um de nível 1 com estrutura “determinante + nome”, bem como também o potencial de sintagmas de maior nível, em alguns casos.

Tabela 2 – Exemplo de SNs e níveis.

o princípio da legalidade	n2
<i>a legalidade</i>	n1
a produção normativa das agências reguladoras brasileiras	n2
as agências reguladoras brasileiras	n1
a ascensão de um estado regulador	n2
um estado regulador	n1
a prestação de serviços públicos	n2
serviços públicos	n1
o exercício de atividades econômicas à iniciativa privada	n3
atividades econômicas à iniciativa privada	n2
a iniciativa privada	n1
a organização e a fiscalização do desempenho dessas funções	n4
a fiscalização do desempenho dessas funções	n3
o desempenho dessas funções	n2
<i>essas funções</i>	n1
o sistema brasileiro	n1
as agências reguladoras	n1
a prestação daquelas atividades administrativas pelos particulares	n3
aquelas atividades administrativas pelos particulares	n2
<i>os particulares</i>	n1
o fundamento jurídico-constitucional para o exercício dessa função normativa	n3
o exercício dessa função normativa	n2
<i>essa função normativa</i>	n1
o real limite dessa atuação	n2
<i>essa atuação</i>	n1
a ordem jurídica	n1
os princípios da separação dos poderes e da legalidade	n4
a separação dos poderes e da legalidade	n3
os poderes e da legalidade	n2

a legalidade	n1
uma ampla margem	n1
atuação transferida àquelas autarquias	n2
aquelas autarquias	n1
as leis instituidoras	n1
<i>o tema</i>	n1
a análise doutrinária	n1
<i>os discursos</i>	n1

Fonte: desenvolvido pelo autor.

A tabela anterior se constitui de uma pequena parte dos sintagmas que foram extraídos do DOC. 22 do *corpus* utilizado nesse experimento. O documento em questão trata dos limites da produção normativa das agências reguladoras brasileiras, tomando por base o princípio da legalidade. Além de abordar o fundamento jurídico-constitucional para o exercício da função normativa dessas instituições, bem como qual o real limite dessa atuação. Levando em consideração essa pequena síntese do conteúdo do documento em questão e analisando os sintagmas marcados em itálico e em cinza da tabela anterior, percebe-se a distinção no quesito potencial discriminatório dos referidos sintagmas, uma vez que os sintagmas marcados em cinza são mais específicos e conseqüentemente tendem a ser bons candidatos a descritores documentais do documento do qual foram extraídos. Já os sintagmas escritos em itálico são mais gerais e menos específicos. O sintagma de nível 2, marcado em negrito, possui alto poder discriminatório para o documento em exemplo, uma vez que, sozinho ele abrange a questão da função normativa e das agências reguladoras brasileiras.

A análise dos diferentes níveis dos SNs permitiu a confirmação do que Souza (2006) já havia ressaltado acerca da relação que o nível do Sintagma possui com a sua densidade informacional, ou seja, a estrutura e o nível do sintagma são proporcionais a sua carga semântica.

Mais adiante, são expostas as taxas de revocação e precisão alcançadas com a análise de cada nível individualmente. Para elaboração dos cálculos levou-se em consideração os SNs de nível 1a, 1b, 2, 3, 4 e 5 ou mais. Esses dados encontram-se no Quadro 34, a seguir:

Quadro 34 – Taxas de revocação e precisão alcançadas com a análise de cada nível de SN.

TAXAS DE REVOCAÇÃO E PRECISÃO DE CADA NÍVEL DE SN		
Nível	Revocação	Precisão

1a	6,3%	6,3%
1b	19,3%	18,4%
2	31,9%	26,3%
3	21,2%	39,8%
4	9,6%	44,5%
5+	11,3%	60,7%

Fonte: desenvolvido pelo autor.

O Quadro 34 permite uma visualização do comportamento de cada nível de SN no que se refere à taxa de revocação e precisão. A revocação foi alcançada dividindo-se o total de SNs descritores de cada nível individualmente pelo total de SNs descritores de toda a coleção (soma de todos os SNs descritores de cada documento). Já a precisão foi alcançada, dividindo-se o total de SNs descritores de cada nível pelo total de SNs descritores e não descritores que cada nível. Levando em consideração a revocação, verifica-se que o nível 2 foi o que alcançou um melhor índice, e, no que se refere a precisão, os SNs de nível igual ou maior a 5 se mostraram com um índice melhor. De modo geral, os SNs de nível 2 ou mais apresentaram melhores taxas de precisão.

Os dados demonstrados nos Quadros 33 e 34, os dois anteriores, conduziram para a escolha de um ponto de corte para a análise do critério nível de modo geral. Esse ponto de corte foi necessário para o levantamento da taxa de revocação e precisão de cada documento de forma individual e de toda a coleção, como forma de contribuir para o julgamento da utilidade do critério de nível para a seleção de SNs descritores. Com base nesse ponto de corte, calculou-se a revocação e precisão geral para todos os SNs de nível maior ou igual a dois (2), onde se verificou uma taxa de revocação de 74,2% e precisão de 41,5%.

No Quadro 35, a seguir, pode-se perceber os percentuais de revocação e precisão obtidos quando da aplicação desse critério em cada documento individualmente, bem como as médias e o desvio padrão. Para a quantificação das taxas de revocação e precisão selecionou-se os sintagmas que possuíam nível maior ou igual a dois (2), ponto de corte já mencionado. Depois de aplicado esse ponto de corte, calculou-se a precisão, dividindo-se a quantidade de SNs descritores que se encontravam dentro dos SNs de nível dois ou maior pela quantidade total de SNs que passaram pelo ponto de corte mencionado anteriormente. A revocação foi calculada, dividindo-se a quantidade de SNs descritores que passaram pelo ponto de corte mencionado pela quantidade total de SNs descritores do documento inteiro.

Quadro 35 – Médias de Revocação e Precisão / Critério: nível.

QUADRO RESUMO DE APLICAÇÃO DO CRITÉRIO NÍVEL			
DOCUMENTO	QUANTIDADE DE SNs SELECIONADOS	REVOCAÇÃO	PRECISÃO
DOC. 1	11	66,7%	54,5%
DOC. 2	11	60,0%	54,5%
DOC. 3	25	66,7%	32,0%
DOC. 4	18	61,5%	44,4%
DOC. 5	11	40,0%	18,2%
DOC. 6	32	82,1%	71,9%
DOC. 7	38	62,5%	13,2%
DOC. 8	30	93,8%	50,0%
DOC. 9	18	80,0%	88,9%
DOC. 10	27	80,0%	44,4%
DOC. 11	34	71,4%	29,4%
DOC. 12	16	83,3%	31,2%
DOC. 13	32	75,0%	28,1%
DOC. 14	33	80,0%	36,3%
DOC. 15	35	60,0%	17,1%
DOC. 16	45	90,9%	22,2%
DOC. 17	36	72,0%	50,0%
DOC. 18	46	76,9%	43,4%
DOC. 19	28	77,7%	50,0%
DOC. 20	52	86,6%	25,0%
DOC. 21	40	66,6%	35,0%
DOC. 22	35	85,7%	17,1%
DOC. 23	30	71,4%	33,3%
DOC. 24	32	80,0%	37,5%
DOC. 25	24	62,5%	20,8%
DOC. 26	41	37,5%	7,3%
DOC. 27	46	66,6%	26,0%
DOC. 28	28	83,3%	53,5%
DOC. 29	37	77,7%	37,8%
DOC. 30	19	62,5%	26,3%
MEDIA	30,33333333	72,0%	36,6%
DESVIO PADRÃO	10,89236789	13,1%	17,7%
MEDIANA	32	73,5%	34,2%

Fonte: desenvolvido pelo autor.

Por meio das médias de revocação e precisão, expostas no quadro 35, é possível perceber uma taxa de revocação relativamente aceitável, onde, ao comparar os sintagmas descritores que possuíam nível maior ou igual a dois com todos os sintagmas descritores, fossem de nível 1, 2, 3, 4 ou mais, alcançou-se uma taxa de revocação de 72%. Este fato confirma o que já foi identificado na literatura da área, onde a densidade informacional do Sintagma Nominal é relativa ao nível do sintagma. Grande parte dos sintagmas descritores encontra-se nos sintagmas de nível dois ou mais. Apesar da precisão não ter sido tão alta, essa se mostrou acima da média obtida quando da não aplicação do critério, onde a taxa é de 25%, a qual encontra-se no Quadro 16. Ao comparar os sintagmas descritores que possuíam nível maior ou igual a dois com o total de sintagmas (descritores ou não) que superavam esse ponto de corte, verificou-se uma precisão, em média, de 36,6%. Essas taxas são as médias das taxas de revocação e precisão de cada documento individualmente. Verifica-se que essas médias são próximas das taxas de revocação e precisão gerais para todos os SNs de nível maior ou igual a dois (2), onde se verificou uma taxa de revocação de 74,2% e precisão de 41,5%.

Essas medidas são relativas, ao passo que ao aumentar uma, outra diminui. Assim, o interessante é alcançar um equilíbrio entre essas medidas. Com base em todos os dados expostos, e por meio das taxas mencionadas mais acima, verifica-se que esse critério se mostrou útil ao passo que alcançou boas médias, demonstrando assim, e, ratificando o que Souza (2006) já ressaltou, que o nível do SN deve ser levado em conta como recurso para subsidiar a seleção de SNs descritores dentre uma variedade de SNs descritores e não descritores.

4.8 Posição do Sintagma Nominal no Documento

A aplicação desse critério, bem como a possível utilidade do mesmo surgiu no decorrer das análises dos SNs nos critérios anteriores, onde foi possível perceber que grande parte dos sintagmas descritores encontrava-se entre os primeiros sintagmas extraídos de cada documento. Assim, buscou-se verificar se existia uma relação entre relevância e o quesito posição dos sintagmas, ou seja, se os primeiros sintagmas extraídos, os quais vieram do título e do primeiro trecho do resumo de cada documento se constituíam de SNs descritores.

Como forma de embasar a aplicação desse critério, bem como o ponto de corte utilizado, elaborou-se um Quadro, onde foi exposto o total de SNs descritores e não descritores, bem como o percentual de SNs descritores das partes dos documentos utilizados.

Assim, os SNs extraídos de cada documento foram organizados na ordem em que apareciam no texto e procedeu-se com a divisão desses SNs em quatro partes do resumo. Essas quatro partes correspondiam, em ordem, aos SNs extraídos do título e do resumo de cada documento. Assim, procedeu-se com a análise dos SNs descritores e não descritores que se encontravam em cada uma das quatro partes de cada documento, como pode ser visto no Quadro 36, a seguir:

Quadro 36 – Quantitativo de SNs descritores ou não que se encontraram em uma das quatro partes dos documentos.

Parte de Documento	Total de SNs descritores	Total de SNs não descritores	Percentual de SNs descritores
Quantidade de SNs no primeiro quartil (0-25% dos SNs)	170	270	38,6%
Quantidade de SNs no segundo quartil (26 a 50% dos SNs)	77	366	17,3%
Quantidade de SNs no terceiro quartil (51 a 75% dos SNs)	78	366	17,5%
Quantidade de SNs no quarto quartil (76 a 100% dos SNs)	98	356	21,5%

Fonte: desenvolvido pelo autor.

Por meio do Quadro 36, pode-se constatar que a maior parte dos SNs descritores se encontram, no contexto desta pesquisa especificamente, na primeira parte dos documentos utilizados neste experimento, ou seja, nos SNs extraídos do título e do início do resumo de cada documento. Na quarta parte, ou seja, nos SNs extraídos do final do resumo de cada documento, verificou-se, também, uma maior quantidade de SNs descritores quando comparado com o segundo e terceiro quartil de cada documento. Nesse contexto, o percentual de SNs descritores se mostrou bem melhor no primeiro quartil, ficando em 38,6%. Infere-se, com base nos dados expostos no quadro anterior, que os SNs descritores se encontram mais no título, no início do resumo, e no fim de cada resumo. Isso se deve pelo fato de que o título possui, por sua própria natureza, ou, talvez deva possuir, expressões que resumam o conteúdo do documento, ou seja, termos com alto poder discriminatório. Assim, os SNs que se encontram no título de cada documento possuem maior poder discriminatório do que outros SNs que se encontram em outras partes de cada documento. No que se refere às partes do resumo, verifica-se que o início do resumo e o fim são as partes que mais possuem SNs descritores. O que justifica isso é que, convencionalmente, as primeiras frases de um resumo

possuem mais termos específicos e conceitos mais precisos, o que gera SNs mais relevantes do que outros de outras partes do resumo.

O melhor comportamento deste critério se deu para os SNs extraídos do título e do início do resumo e também para os SNs extraídos do final do resumo. Esse comportamento pode ser relacionado com a superestrutura do texto utilizado nesta pesquisa, onde os resumos, por exemplo, iniciam com a definição do propósito do estudo, o agente da ação, o objeto de estudo, bem como a ação desenvolvida para alcançar os objetivos. Já a última parte dos resumos contém os elementos conclusivos, onde expõem-se os resultados alcançados e as considerações acerca de toda a pesquisa. Essas macroestruturas são as partes mais informativas e que conseqüentemente possuem um maior número de SNs relevantes e que conseguem representar o conteúdo geral do documento ao qual o resumo pertence.

Nesse contexto, calcularam-se as taxas de revocação e precisão de cada uma das quatro partes de cada documento, bem como de toda a coleção. A revocação foi obtida por meio da divisão do total de SNs descritores de cada parte pelo total de NS descritores de toda a coleção. Já a precisão foi alcançada por meio da divisão entre o total de SNs descritores de cada parte pelo total de SNs descritores e não descritores de cada parte. Mais adiante, esses dados são expostos no Quadro 37, a seguir:

Quadro 37 – Taxas de revocação e precisão alcançadas com a análise de cada uma das quatro partes dos documentos.

TAXAS DE REVOCAÇÃO E PRECISÃO DE CADA PARTE DO DOCUMENTO		
Parte do documento	Revocação	Precisão
Primeiro quartil	40,1%	38,6%
Segundo quartil	18,2%	17,3%
Terceiro quartil	18,4%	17,5%
Quarto quartil	23,1%	21,5%

Fonte: desenvolvido pelo autor.

Os dados do Quadro 37 ratificam, por meio de percentuais, o que demonstra o Quadro 36, onde se verifica maior revocação e precisão, respectivamente, 40,1% e 38,6% para o primeiro quartil dos documentos utilizados neste experimento.

Depois de elaborados os cálculos anteriores e de feita uma análise mais ampla das partes dos documentos que compuseram o *corpus* desta pesquisa, procedeu-se com o levantamento das taxas de revocação e precisão de cada documento levando em consideração os primeiros sintagmas extraídos de cada documento. Como ponto de corte, escolheu-se os

10 primeiros sintagmas de cada documento. A escolha desse ponto de corte se deu pelo fato de que, como demonstrado anteriormente, a maior parte dos SNs descritores se encontra entre os primeiros SNs extraídos. Assim, a escolha do quantitativo de 10 SNs se deu, buscando um melhor desempenho do critério de posição em termos de precisão. No cálculo da revocação, dividiu-se a quantidade de sintagmas descritores que se encontravam entre os dez primeiros sintagmas extraídos de cada documento pela quantidade total de sintagmas descritores do documento inteiro. A precisão foi obtida comparando-se os sintagmas descritores que se encontravam entre os dez primeiros sintagmas extraídos com todos os dez primeiros sintagmas.

No Quadro 38, são expostas as taxas de revocação e precisão, bem como as médias alcançadas:

Quadro 38 – Médias de revocação e precisão / critério: posição do SN.

QUADRO RESUMO DE APLICAÇÃO DO CRITÉRIO: POSIÇÃO/ORDEM		
DOCUMENTO	REVOCAÇÃO	PRECISÃO
DOC. 1	44,4%	40,0%
DOC. 2	40,0%	40,0%
DOC. 3	16,7%	20,0%
DOC. 4	15,4%	30,0%
DOC. 5	60,0%	30,0%
DOC. 6	32,1%	90,0%
DOC. 7	37,5%	30,0%
DOC. 8	12,5%	20,0%
DOC. 9	30,0%	60,0%
DOC. 10	53,3%	80,0%
DOC. 11	21,4%	30,0%
DOC. 12	50,0%	30,0%
DOC. 13	33,3%	40,0%
DOC. 14	6,6%	10,0%
DOC. 15	40,0%	40,0%
DOC. 16	27,2%	30,0%
DOC. 17	24,0%	60,0%
DOC. 18	23,0%	60,0%
DOC. 19	50,0%	90,0%

DOC. 20	26,6%	40,0%
DOC. 21	28,5%	60,0%
DOC. 22	42,8%	30,0%
DOC. 23	57,1%	80,0%
DOC. 24	40,0%	60,0%
DOC. 25	50,0%	40,0%
DOC. 26	50,0%	40,0%
DOC. 27	16,6%	30,0%
DOC. 28	27,7%	50,0%
DOC. 29	22,2%	40,0%
DOC. 30	62,5%	50,0%
MEDIA	34,7%	45,0%
DESVIO PADRÃO	15,1%	20,5%
MEDIANA	32,7%	40,0%

Fonte: desenvolvido pelo autor.

Por meio do Quadro 38, é possível perceber uma taxa de revocação relativamente baixa de 34,7%, quando se comparou os sintagmas descritores que se encontram entre os dez (10) primeiros sintagmas extraídos com o total de sintagmas descritores que se encontravam em qualquer posição do documento. Em contrapartida alcançou-se uma precisão média de 45,0% ao comparar os sintagmas descritores que se encontravam entre os dez (10) primeiros sintagmas com o total de dez sintagmas selecionados. A precisão alcançada se mostrou bem acima da taxa quando da não aplicação desse critério, a qual gira em torno de 25% e que se encontra no Quadro 16.

Com base em todas as análises feitas acerca do critério de “posição do SN” como recurso para a seleção de SNs, conclui-se que o elemento “posição” tende a ser pertinente para a seleção de SNs descritores, ressaltando o que Corrêa et al. (2011) afirmaram sobre essa questão, ao dizerem que o elemento “posição”, apesar de carecer de mais estudo, pode sim contribuir para a seleção de SNs. Esse critério se mostrou útil, pois permitiu inferir-se que a posição do SN tem relação com o seu potencial discriminatório, ou seja, SNs que se encontram em partes mais relevantes de documentos, tendem a ser melhores candidatos a descritores documentais. Além disso, esse critério alcançou um bom desempenho no quesito precisão em selecionar SNs descritores.

A análise desse critério deixa clara a importância que deve ser dada à superestrutura dos documentos no momento de representação da informação, seja a representação manual ou automática. Esse critério coloca em relevo a questão da importância de determinadas partes (estruturas) do documento para a atividade de representação temática. Conhecer a estrutura de

um documento, bem como a ordem global do texto/documento, ou seja, as partes que compõem um ou outro texto. Analisando a superestrutura textual dos textos utilizados neste experimento, pôde-se perceber que determinadas partes do texto possuem de fato maior potencial informativo, por conterem a maior parte dos SNs descritores.

4.9 Frequência de ocorrência dos SNs no Texto/Documento e Frequência Normalizada

O critério de frequência de ocorrência de um sintagma foi aplicado a todos os SNs extraídos de cada documento, uma vez que se constitui um critério básico de seleção de termos candidatos a descritores documentais. A frequência de ocorrência de termos, ou frequência absoluta, representa a medida da quantidade de vezes que um determinado termo aparece em um documento. Essa é a medida de peso mais simples que existe, sendo bem utilizada em sistemas de indexação automática que utilizam palavras constantes nos textos para a recuperação dos mesmos. As várias palavras existentes em um documento possuem valores distintos para fins de indexação e recuperação desse documento, o que demonstra que algumas palavras por ocorrem com maior frequência tendem a ser melhores candidatos a termos de indexação, bem como também a ocorrência demasiada de um determinado termo pode representar um termo muito usual, comum, e que não seja pertinente para servir como termo de indexação.

Por ser uma medida simples, é notória a necessidade de uso desta em conjunto com outras metodologias em sistemas de indexação e recuperação de informação. Nesse contexto, procedeu-se com a contabilização de quantas vezes cada SN ocorria no texto/documento. No Quadro x39, a seguir, observa-se a representação das frequências de ocorrências de SNs percebidas para as várias frequências. Nesse quadro não se faz distinção de SNs descritores e não descritores, enfatizando apenas a repetição.

Quadro 39 – Ocorrências de SNs percebidas para várias frequências.

Frequência de ocorrência no texto	Quantidade de SNs
1 ocorr.	1664
2 ocorr.	85
3 ocorr.	20
4 ocorr.	6
5 ou +	6

Fonte: desenvolvido pelo autor.

A frequência que sobressai no Quadro 39 é a que ocorre apenas uma vez, seguido dos SNs que ocorrem duas vezes, três, quatro e cinco ou mais. Depois de feito o levantamento das diferentes frequências de ocorrências, procedeu-se com a diferenciação dos SNs descritores e não descritores que possuíam uma dada frequência ou outra. Esses dados podem ser vistos no Quadro 40, logo abaixo:

Quadro 40 – Quantitativo de SNs descritores ou não de acordo com a frequência de ocorrência no texto/documento.

FREQUÊNCIA DE OCORRÊNCIA DOS SNs	TOTAL DE SNs DESCRITORES	TOTAL DE SNs NÃO DESCRITORES	PERCENTUAL DE SNs DESCRITORES
Quantidade de SNs que aparecem somente 1 vez	376	1288	22,5%
Quantidade de SNs que aparecem 2 vezes	26	59	30,5%
Quantidade de SNs que aparecem 3 vezes	13	7	65,0%
Quantidade de SNs que aparecem 4 vezes	5	1	83,3%
Quantidade de SNs que aparecem mais de 5 vezes	3	3	50,0%

Fonte: desenvolvido pelo autor.

Por meio do Quadro 40 fica perceptível que a maior parte dos SNs descritores aparecem apenas uma única vez, resultando em um percentual de 22,5% de SNs descritores para esses SNs que aparecem uma só vez. Isso demonstra que, nesse experimento específico, a frequência de ocorrência absoluta não fez tanta diferença, pois o que se espera, como já visto na literatura, é que os SNs que se repitam com frequência tendam a ser descritores documentais ou candidatos a descritores. Depois dos SNs que apareceram apenas uma vez, a frequência que se mostrou um pouco melhor foi a frequência de duas ocorrências, onde de um total de 85 SNs que se repetiram duas vezes, 30,5% deles eram descritores documentais. O Quadro 39 deixa claro que o percentual de SNs descritores se mostra bem melhor para as frequências de duas vezes pra cima, a partir da qual alcança um percentual acima de 30,5%, semelhante à precisão alcançada quando da não aplicação de nenhum critério.

Verifica-se que, apesar de a maior parte dos SNs descritores se encontrar entre os SNs que aparecem apenas um única vez, deve-se atentar para os percentuais de SNs descritores para cada faixa de frequência, onde as frequências de 2, 3, 4 e 5 ou mais vezes alcançaram melhores resultados, com os respectivos percentuais, 30,5%, 65,0%, 83,3% e 50%. Isso

ratifica o que se encontra na literatura acerca da tendência de SNs que se repetem em um determinado documento em possuírem mais características de descritores documentais.

As taxas de revocação e precisão de cada uma das frequências expostas no Quadro 40 são demonstradas no Quadro 41, logo abaixo.

Quadro 41 – Taxas de revocação e precisão para cada frequência de ocorrência dos SNs no texto/documento.

TAXAS DE REVOCAÇÃO E PRECISÃO DE CADA FREQUÊNCIA DE OCORRÊNCIA		
	REVOCAÇÃO	PRECISÃO
Quantidade de SNs que aparecem somente 1 vez	88,8%	22,5%
Quantidade de SNs que aparecem 2 vezes	6,1%	30,5%
Quantidade de SNs que aparecem 3 vezes	3,0%	65,0%
Quantidade de SNs que aparecem 4 vezes	1,1%	83,3%
Quantidade de SNs que aparecem mais de 5 vezes	0,7%	50,0%

Fonte: desenvolvido pelo autor.

A faixa de frequência de ocorrência que demonstrou melhor revocação, com base no quadro acima, foi a frequência de ocorrência de apenas uma vez, alcançando uma revocação de 88,8%. Já as faixas de frequências que apresentaram melhores índices de precisão foram as que apareceram 2, 3, 4 e 5 ou mais vezes, demonstrando precisão acima da taxa de precisão alcançada sem a aplicação de nenhum critério, a qual é de 23,75% para sintagmas nominais e 25% para documentos. Levando em consideração mais o valor da precisão, verifica-se que a frequência de ocorrência absoluta acima de 1 demonstra melhores resultados em comparação com a frequência de ocorrência de apenas uma única vez. Ressaltando a precisão, elaborou-se o cálculo de revocação e precisão acumulado dos SNs que possuíam frequência absoluta maior que um (1). Como resultado, obteve-se: 11,1% de revocação e 40,1% de precisão. Essa baixíssima revocação se deu pelo fato de que grande parte dos SNs descritores, como já ressaltado, encontravam-se na faixa de frequência de ocorrência de apenas uma vez. Já a precisão desse acumulado (com frequência absoluta maior que 1) se mostrou bem melhor.

Em paralelo à frequência absoluta de cada termo dentro de um documento, encontra-se a frequência normalizada, a qual se constitui da frequência absoluta de um determinado termo dividida pela quantidade total de Sintagmas extraídos do documento. Essa frequência normalizada leva em consideração o tamanho do documento, tendo em vista que o comprimento do mesmo influencia a frequência de ocorrência de um determinado termo.

Assim, analisaram-se essas duas frequências (absoluta e normalizada) juntas, para que fossem calculadas as taxas de precisão e revocação de cada documento individualmente, tendo em vista verificar o comportamento de cada documento individual, bem como levantamento de uma média de toda a coleção. Inicialmente, aplicou-se um ponto de corte onde se selecionou os sintagmas que possuíam frequência no documento maior que um (1) e frequência normalizada maior que 0,02739. O ponto de corte da frequência normalizada variava, tendo em vista que em alguns documentos o menor valor era 0,05555556, ou 0,0333579. Em suma, o ponto de corte efetivo foi a frequência absoluta, pois percebeu-se que a frequência normalizada parecia não se mostrar útil no corpus trabalhado. Depois de aplicado esse ponto de corte, calculou-se a precisão, dividindo-se a quantidade de SNs descritores que se encontravam acima desse ponto de corte pela quantidade total de sintagmas que se encontravam acima do referido ponto de corte. A revocação foi calculada, dividindo-se a quantidade de SNs descritores que se encontravam acima do referido ponto de corte pela quantidade total de SNs descritores do documento inteiro. Mais adiante, expõem-se, no Quadro 42, um resumo das taxas de revocação e precisão, bem como as médias e o desvio padrão.

Quadro 42 - Médias de Revocação e Precisão / Critério: Freq. Absoluta e normalizada.

QUADRO RESUMO DE APLICAÇÃO DO CRITÉRIO: FREQ. NORM.			
DOCUMENTO	QUANTIDADE DE SNs SELECIONADOS	REVOCAÇÃO	PRECISÃO
DOC. 1	2	0,0%	0,0%
DOC. 2	0	0,0%	0,0%
DOC. 3	1	0,0%	0,0%
DOC. 4	1	0,0%	0,0%
DOC. 5	1	0,0%	0,0%
DOC. 6	6	17,9%	83,3%
DOC. 7	3	0,0%	0,0%
DOC. 8	5	6,3%	20,0%
DOC. 9	3	15,0%	100,0%
DOC. 10	4	20,0%	75,0%
DOC. 11	3	7,1%	33,3%
DOC. 12	2	0,0%	0,0%
DOC. 13	3	8,3%	33,3%

DOC. 14	3	6,7%	33,3%
DOC. 15	3	10,0%	33,3%
DOC. 16	8	27,2%	37,5%
DOC. 17	4	16,0%	100,0%
DOC. 18	5	11,5%	60,0%
DOC. 19	2	11,1%	100,0%
DOC. 20	4	6,6%	25,0%
DOC. 21	3	9,5%	66,6%
DOC. 22	3	14,2%	33,3%
DOC. 23	5	21,4%	60,0%
DOC. 24	4	26,6%	100,0%
DOC. 25	4	12,5%	25,0%
DOC. 26	8	0,0%	0,0%
DOC. 27	12	27,7%	41,6%
DOC. 28	5	11,1%	40,0%
DOC. 29	2	0,0%	0,0%
DOC. 30	8	0,0%	0,0%
MEDIA	3,9	9,6%	36,7%
DESVIO PADRÃO	2,537307832	8,9%	35,1%
MEDIANA	3	8,9%	33,3%

Fonte: desenvolvido pelo autor.

Por meio das médias das taxas de revocação e precisão, postas no quadro anterior, percebe-se uma baixa taxa de revocação em contrapartida de uma maior taxa precisão. Comparando-se os sintagmas descritores que possuíam frequência absoluta maior que um (1) e frequência normalizada maior que 0,027027027 com todos os sintagmas descritores (com qualquer frequência absoluta e normalizada), verificou-se uma taxa de revocação média de 9,6%, muito abaixo da taxa de revocação quando da não aplicação desse critério, e uma precisão média 36,7%, ou seja, acima da média de precisão quando da não aplicação desse critério. Essas taxas se assemelham às obtidas anteriormente para os sintagmas nominais, onde se calculou a revocação e a precisão acumulada dos SNs que possuíam frequência absoluta acima de um (1), onde os resultados ficaram em torno de 11,1% de revocação e 40,1% de precisão.

Há de se levar em consideração, quando da análise do Quadro 41, que o desvio padrão para a precisão e revocação foi alto, demonstrando alto grau de dispersão de valores para com a média, isto causado por grande oscilação de revocação e precisão, como pode ser vista no quadro. Enquanto que um documento alcançava 100% de precisão, outro não passava dos 0%. Dos trinta documentos analisados, dez tiveram taxas de revocação e precisão 0%, pelo que foi

percebido, isso se deve ao fato de que muitos dos SNs descritores (ou seja, que eram semelhantes as palavras-chave ou as continham em suas estruturas) não possuíam frequência maior que um (1), isto é, abaixo do ponto de corte. Por exemplo, no DOC. 4, todos os SNs descritores eram de frequência absoluta um (1), ou seja, as taxas de revocação e precisão para este DOC. 4 já eram de 0%. Isso faz perceber que a frequência absoluta, neste momento, não se fez tão pertinente. Essa verificação pode ser explicada pela baixa frequência de repetição de sintagmas em todo o *corpus*, como pôde ser visto anteriormente as repetições de sintagmas foi baixa no corpus trabalhado, talvez isso se dê pelo fato de se ter utilizado apenas o resumo e o título de cada documento, ao invés do documento na íntegra para a extração de sintagmas.

Em suma, apesar de ter verificado que a frequência não foi um ponto crucial no corpus trabalhado devido às próprias características dos elementos textuais utilizados, pôde-se perceber que a frequência de ocorrência de um determinado termo tem muito a ver com o poder discriminatório do SN. Levando em consideração as taxas de precisão alcançadas com todos os cálculos, pode-se inferir que o critério de frequência absoluta se mostrou útil, ao passo que consegui, de modo geral, ser mais preciso em selecionar SNs descritores e evitar os SNs não descritores. Apesar de ter verificado que a frequência normalizada não pareceu útil neste experimento, é importante que se ressalte que isto foi percebido neste experimento específico, não invalidando a sua utilidade em outros contextos. Retomando aqui a pesquisa de Souza e Raghavan (2014), é pertinente ressaltar que a utilização da frequência normalizada na pesquisa desses autores se mostrou viável ao passo que alcançou melhores resultados do que quando da não aplicação da frequência normalizada.

4.10 Frequência de ocorrência dos SNs no Conjunto de Documentos – *Frequência Inversa de Documento – (Idf)*

A verificação da frequência de ocorrência dos SNs no conjunto de documentos e não somente no próprio documento é uma medida importante para a verificação e seleção de SNs que possam funcionar como descritores, uma vez que a ocorrência de um determinado SN em demasia em vários documentos pode demonstrar que esse SN é comum e possui pouco poder discriminatório. Segue o Quadro 43 que demonstra alguns sintagmas nominais que apresentaram uma frequência de ocorrência alta no conjunto do *corpus* desta pesquisa:

Quadro 43 – Exemplos de SNs com frequências altas em todo o *corpus*

SINTAGMAS NOMINAIS	FREQUÊNCIA DE OCORRENCIA EM TODO O <i>CORPUS</i>
o direito	11
o estado	6
o presente trabalho	6
o trabalho	5
o tema	5
a dissertação	5

Fonte: desenvolvido pelo autor.

Dentre os Sintagmas que se encontram no quadro anterior, o sintagma “*o direito*” apareceu em 11 documentos de um total de trinta, demonstrando assim o seu caráter genérico, possuindo pouco poder discriminatório. Os sintagmas “o direito” e “o estado” são exemplos de termos que ocorreram com frequência no *corpus*. Isso demonstra a viabilidade de se levar em conta para a seleção de SNs valores relacionados ao inverso da frequência de ocorrência no *corpus*. As pesquisas nessa vertente fazem uso dessa medida em conjunto com outros critérios para uma seleção mais eficaz de SNs que funcionem como descritores documentais.

Alguns pesquisadores propuseram fórmulas que levassem em conta tanto a frequência de um termo no documento em que ocorre como também a frequência desse mesmo termo em vários documentos. A frequência do termo, conhecida como “*tf*” e o inverso da frequência do termo em vários documentos diferentes, conhecida como “*idf*”, juntas, formam o índice *tf-idf*, o qual possui várias formas de representação, porém todas com a ideia de que um termo vai ser relevante para um determinado documento se ele for muito frequente nesse documento e, em contra partida, aparecer em poucos documentos. A aplicabilidade desse índice é bem comum, pois permite que termos frequentes que se encontram em vários documentos não sejam considerados tão pertinentes para representação de determinados documentos, uma vez que possuem uma popularidade extensa, possuindo pouco poder discriminatório. Apesar da função desempenhada, esse índice apresenta em alguns momentos alguns inconvenientes, pontuando sintagmas que não deveriam ser pontuados. Lopes (2012), em sua pesquisa, demonstra claramente o desempenho desse índice em comparação com outros índices.

No Quadro 44 são expostos dados quantitativos referentes à quantidade de SNs descritores e não descritores que aparecem em apenas um documento da coleção, em dois documentos da coleção, três, quatro e em cinco ou mais documentos.

Quadro 44 – Quantitativo de SNs descritores ou não que apareceram em 1, 2, 3, 4 ou 5 ou mais documentos da coleção.

COLEÇÃO	TOTAL DE SNs DESCRITORES	TOTAL DE SNs NÃO DESCRITORES	PERCENTUAL DE SNs DESCRITORES	IDF
Quant. de SNs que aparecem em 1 documento	394	1218	24,40%	1,477121
Quant. de SNs que aparecem em 2 documentos	23	73	23,90%	1,176091
Quant. de SNs que aparecem em 3 documentos	5	22	18,50%	1
Quant. de SNs que aparecem em 4 documentos	0	8	0,00%	0,875061
Quant. de SNs que aparecem em 5 ou mais documentos	1	37	2,60%	[0,43472857 0,778155125]

Fonte: desenvolvido pelo autor.

Com base no Quadro 43, verifica-se que de todos os SNs descritores de toda a coleção, a maior parte deles aparece em apenas um documento dos trinta que compõe o corpus. Analisando a quarta coluna do Quadro 43, pode-se perceber que a medida que os SNs aparecem em mais documentos, o percentual de SNs descritores diminui, demonstrando assim o que foi visto na literatura acerca da frequência demasiada de SNs, ou seja, SNs que aparecem com frequência em muitos documentos, tendem a possuir pouco poder discriminatório não servindo como descritores documentais. Verifica-se também que os percentuais de SNs descritores apresentam melhores taxas para as faixas de ocorrência em 1, 2, e 3 documentos, as quais possuem, respectivamente, os seguintes percentuais de SNs descritores: 24,1%, 23,9% e 18,5%. Levando em consideração o IDF, percebe-se que o IDF igual ou acima de um (1) é o que tem um melhor rendimento em termos de selecionar SNs descritores.

A seguir, são demonstradas as taxas de revocação e precisão alcançadas para cada faixa de ocorrência dos SN em um, dois, três, quatro e em cinco ou mais documentos. Esses dados encontram-se no Quadro 45.

Quadro 45 – Taxas de revocação e precisão alcançadas com as diferentes frequências de ocorrências dos SNs na coleção.

TAXAS DE REVOCAÇÃO E PRECISÃO PARA CADA FREQUÊNCIA DE OCORRÊNCIA NA COLEÇÃO – IDF		
COLEÇÃO	REVOCAÇÃO	PRECISÃO
Quant. de SNs que aparecem em 1 documento	93,1%	24,4%
Quant. de SNs que aparecem em 2 documentos	5,4%	23,9%
Quant. de SNs que aparecem em 3 documentos	1,1%	18,5%
Quant. de SNs que aparecem em 4 documentos	0,0%	0,0%
Quant. de SNs que aparecem em 5 ou mais documentos	0,2%	2,6%

Fonte: desenvolvido pelo autor.

Ao analisar as taxas de revocação e precisão do Quadro 45 percebe-se que tanto a revocação quanto a precisão se mostraram com melhor desempenho para os SNs que apareceram em apenas um documento. Se olhar especificamente para a precisão, verifica-se um desempenho razoável nos SNs que aparecem em dois (2) documentos de toda a coleção.

Levando em consideração os SNs que tiveram frequência na coleção maior que um (1) e idf igual ou maior que um (1), tem-se as seguintes taxas de revocação e precisão: respectivamente, 6,6% e 22,7%. Essas taxas se mostraram baixas devido ao fato de a maior parte dos SNs descritores aparecerem apenas em um documento de toda a coleção, como foi colocado no Quadro 43. Assim, os cálculos foram refeitos, levando em consideração, ao invés dos SNs que possuem frequência na coleção maior que um (1) e idf maior ou igual a um (1), os SNs que tiveram frequência na coleção maior ou igual a um (1) e idf maior ou igual a um (1). Os resultados com essa mudança de ponto de corte foram bem melhores, pois alcançaram 99,76% de revocação e 24,3% de precisão.

No Quadro 46, encontra-se, a título de exemplificação, a aplicação do índice *idf* ao DOC.10 do corpus.

Quadro 46 – SNs ordenados de acordo com o índice *idf*.

SINTAGMAS NOMINAIS	Freq. Coleção	IDF	POSIÇÃO
a isonomia no contexto do Direito Tributário	1	1,477121255	1ª
o contexto do Direito Tributário	1	1,477121255	2ª
o Direito Tributário	1	1,477121255	3ª
o princípio da isonomia	1	1,477121255	4ª
a isonomia	1	1,477121255	5ª
direitos do homem	1	1,477121255	6ª
o homem	1	1,477121255	7ª
as constituições hodiernas	1	1,477121255	8ª

a isonomia tributária	1	1,477121255	9ª
os direitos fundamentais	2	1,176091259	10ª
a capacidade contributiva	2	1,176091259	11ª
a concessão de incentivos fiscais pelos estados	2	1,176091259	12ª
incentivos fiscais pelos estados	2	1,176091259	13ª
os estados	2	1,176091259	14ª
Guerra fiscal	2	1,176091259	15ª
Este trabalho	4	0,875061263	16ª
os princípios	4	0,875061263	17ª
o direito	11	0,43572857	18ª

Fonte: desenvolvido pelo autor.

No quadro 46, pode-se perceber a aplicabilidade e a funcionalidade do índice *idf* na seleção de sintagmas que possam servir como descritores documentais levando em consideração tanto a frequência absoluta quanto a frequência inversa de documento. O sintagma “**os princípios**” ocorre 4 vezes na coleção inteira, ficando, portanto, na 17ª posição de candidatos a descritores. Já o sintagma “**o direito**” é o que tem uma maior frequência na coleção inteira, ficando assim na 18ª posição.

A aplicação do índice *idf* demonstrou viabilidade e utilidade para seleção de sintagmas nominais no *corpus* trabalhado, não excetuando a necessidade de outros critérios que contribuam para uma seleção dos melhores SNs.

No Quadro 47 encontram-se as medidas de revocação e precisão quando da aplicação desse critério a cada documento, individualmente, seguidos de média geral para toda a coleção. Para esses cálculos, tomou-se como ponto de corte os SNs que apareciam em um ou em mais de um documento e que possuíam *idf* maior ou igual a 1, ou seja, se um determinado SN aparecia uma vez no documento e possuíam *idf* acima de um, esse SN seria selecionado, todavia se esse SN, hipoteticamente, aparecesse em mais de um documento e possuísse *idf* abaixo de um (1,0), o mesmo não seria selecionado. Esse limiar foi definido tomando como base os dados demonstrados nos quadros anteriores acerca desse critério, os quais demonstraram viabilidade de seleção de SNs descritores com esse ponto de corte.

Quadro 47 – Médias de revocação e precisão / Critério: Freq. *Idf*.

QUADRO RESUMO DE APLICAÇÃO DO CRITÉRIO: FREQ. IDF.			
DOCUMENTO	QUANTIDADE DE SNs SELECIONADOS	REVOCAÇÃO	PRECISÃO
DOC. 1	32	100,0%	28,1%
DOC. 2	36	100,0%	27,8%
DOC. 3	54	100,0%	22,2%
DOC. 4	50	100,0%	26,0%
DOC. 5	44	80,0%	9,1%
DOC. 6	50	100,0%	56,0%
DOC. 7	69	100,0%	11,6%
DOC. 8	49	100,0%	32,7%
DOC. 9	28	100,0%	71,4%
DOC. 10	54	100,0%	27,8%
DOC. 11	70	100,0%	20,0%
DOC. 12	33	100,0%	18,2%
DOC. 13	56	100,0%	21,4%
DOC. 14	57	100,0%	26,3%
DOC. 15	59	100,0%	16,9%
DOC. 16	69	100,0%	15,9%
DOC. 17	67	96,0%	35,8%
DOC. 18	71	100,0%	36,6%
DOC. 19	45	100,0%	40,0%
DOC. 20	82	100,0%	18,2%
DOC. 21	79	100,0%	26,5%
DOC. 22	67	100,0%	10,5%
DOC. 23	53	100,0%	26,4%
DOC. 24	65	100,0%	23,0%
DOC. 25	53	100,0%	15,1%
DOC. 26	87	100,0%	9,2%
DOC. 27	87	100,0%	20,6%
DOC. 28	57	100,0%	31,5%
DOC. 29	66	100,0%	27,2%
DOC. 30	46	100,0%	17,3%
MEDIA	57,83333333	99,2%	25,6%
DESVIO PADRÃO	15,55875036	3,7%	13,2%
MEDIANA	56,5	100,0%	24,5%

Fonte: desenvolvido pelo autor.

A média de precisão alcançada com a aplicação desse critério se mostra pouco maior que a taxa alcançada com a não aplicação do mesmo, entretanto a revocação, em média, foi bem alta. Comparando-se os sintagmas descritores que tinham frequência na coleção igual ou maior que um (1) e *idf* maior ou igual a um (1) com o total de sintagmas descritores (com qualquer freq. na coleção e qualquer *idf*), verificou-se, em média, uma taxa de 99,2% de revocação. Já a taxa de precisão, em média, foi de 25,6% quando da comparação entre os sintagmas descritores que possuíam frequência na coleção maior ou igual a um (1) e *idf* maior ou igual a um (1) com o total de sintagmas (descritores ou não) que se encontraram dentro do ponto de corte escolhido para este critério.

Durante a aplicação desse índice percebeu-se que esse ponto de corte se mostrou útil, ao passo que eliminou sintagmas que não eram SNs descritores como, por exemplo: “o estado – *idf*:0,698970004”; “o trabalho – *idf*:0,77815125”; entre outros. A frequência inversa de documento se mostrou, conforme foi visto na literatura, um recurso efetivo na seleção de descritores documentais, neste caso específico, de SNs descritores. Com base em todos os dados expostos anteriormente, pode-se dizer que este critério foi bem útil na seleção de NS descritores no contexto deste experimento.

Análise e comparação das médias de revocação e precisão de cada critério aplicado na etapa 8 deste experimento, as quais foram demonstradas anteriormente.

Nesse momento, analisaram-se as médias de revocação e precisão alcançadas por cada critério, buscando realizar comparações no que diz respeito ao desempenho dos mesmos por meio dessas médias.

No Quadro 48, são resumidas as médias de revocação e precisão alcançadas quando da aplicação de cada critério nos documentos. Os critérios estão organizados por categorias, para que sejam passíveis de comparações.

Quadro 48 – Resumo das médias de revocação e precisão de todos os critérios

CRITÉRIOS DE ELIMINAÇÃO	REVOCAÇÃO			PRECISÃO		
	MÉDIA	DESVIO	MEDIANA	MÉDIA	DESVIO	MEDIANA
NUMERAIS	96,1%	8,9%	100,0%	24,8%	13,7%	21,7%
PRONOMES/NÚCLEO	100,0%	0,0%	100,0%	25,1%	12,8%	23,5%
ADVÉRBIOS	99,2%	2,5%	100,0%	24,9%	12,8%	23,2%
STOP WORDS	100,0%	0,0%	100,0%	26,4%	13,9%	24,7%
CRITÉRIOS DE ADIÇÃO	REVOCAÇÃO			PRECISÃO		
	MÉDIA	DESVIO	MEDIANA	MÉDIA	DESVIO	MEDIANA
MÚLTIPLOS ADJETIVOS	31,9%	18,6%	27,5%	26,0%	15,5%	22,9%
ADJETIVOS. CONJUN.	0,9%	3,1%	0,0%	7,2%	25,4%	0,0%
NÍVEL	REVOCAÇÃO			PRECISÃO		
	MÉDIA	DESVIO	MEDIANA	MÉDIA	DESVIO	MEDIANA
	72,0%	13,1%	73,5%	36,6%	17,7%	34,2%
POSIÇÃO/ORDEM	REVOCAÇÃO			PRECISÃO		
	MÉDIA	DESVIO	MEDIANA	MÉDIA	DESVIO	MEDIANA
	34,7%	15,1%	32,7%	45,0%	20,5%	40,0%
CRITÉRIOS DE ORDENAÇÃO	REVOCAÇÃO			PRECISÃO		
	MÉDIA	DESVIO	MEDIANA	MÉDIA	DESVIO	MEDIANA
FREQ. ABSOL. E NORM.	9,6%	8,9%	8,9%	36,7%	35,1%	33,3%
FREQ. <i>IDF</i>	99,2%	3,7%	100%	25,6%	13,2%	24,5%

Fonte: desenvolvido pelo autor

Em relação aos critérios de eliminação, perceberam-se comportamentos distintos, onde uns se mostraram útil e outros não, na separação de SNs descritores de SNs não descritores. Em relação ao primeiro critério de eliminação, o de DESCARTE DE SNs COM NUMERAIS EM SUAS ESTRUTURAS, verificou-se que o mesmo se comportou de maneira não tão satisfatória, pois, apesar de ter alcançado uma média de revocação alta, como pode ser vista no quadro anterior, a precisão foi baixa. Verificou-se que esse critério eliminou uma quantidade SNs descritores semelhante à quantidade de SNs não descritores que foi descartada, demonstrando assim que não conseguiu eliminar uma maior quantidade de SNs não descritores e evitar mais a eliminação de SNs descritores. De todas as eliminações feitas por este critério, 40,5% delas eram SNs descritores, ou seja, quase metade dos descartes.

O segundo critério de eliminação, o de DESCARTE DE SNs QUE POSSUEM PRONOME COMO NÚCLEO, foi aplicado a um numero bastante pequeno de sintagmas, pois poucos foram os SNs que possuíam um pronome como núcleo. Essa baixa frequência de ocorrência de SNs com pronomes se deve, talvez, pelas próprias características do corpus trabalhado, títulos e resumos. Essa baixa aplicação teria sido bem mais frequente se o corpus, por exemplo, possuísse características diferentes, como por exemplo, se tivesse sido utilizados artigos por completo, onde as construções anafóricas e a ocorrência de pronomes seriam mais frequentes. Contudo, mesmo com baixa aplicação, esse critério eliminou SNs que não eram SNs descritores, demonstrando assim sua utilidade, ao passo que eliminou sintagmas que em nada contribuíam para a representação dos documentos aos quais pertenciam. A média da taxa de revocação e precisão desse critério demonstram que o mesmo comportou-se de maneira útil, ficando a revocação em 100% e a precisão em 25,1%, ou seja, um pouco acima da taxa obtida sem aplicação de nenhum critério.

O terceiro critério de eliminação, o de DESCARTE DE SNs QUE INICIAM COM ADVÉRBIOS alcançou uma média 99,2% de revocação, média considerada boa, no entanto a precisão ficou abaixo do percentual sem aplicação de nenhum critério. A precisão alcançada por este critério foi de 24,9%. Apesar, de pelos números, acabar-se por se julgar esse critério como útil, verificou-se que os SNs que foram eliminados por ele, apesar de iniciarem com advérbios, eram SNs descritores, por isso a precisão e a revocação baixaram, ainda que de forma tímida, nesse critério. O que se esperava desse critério era que o mesmo eliminasse SNs não descritores, no entanto os SNs que iniciaram com advérbios, neste corpus específico, eram SNs descritores.

O quarto critério de eliminação, o de ELIMINAÇÃO DE SNs CATEGORIZADOS COMO STOPWORDS, obteve, em média, 100% de revocação e 26,4% de precisão, ou seja, a precisão acima da taxa alcançada sem aplicação de nenhum critério. Com esse critério foram eliminados 86 SNs que em nada contribuíam em termos de representação temática, e por isso foram categorizados como *stopwords*.

O primeiro critério de adição, o de DETECÇÃO DE SNs CONTIDOS EM SNs MAIORES POR MEIO DA REMOÇÃO SUCESSIVA DE ADJETIVOS, se comportou de maneira não tão satisfatória, como pôde ser percebido pelas médias das taxas de revocação e precisão obtidas, respectivamente, 31,9% e 26%. Verificou-se que tanto a revocação quanto a precisão foram baixas, demonstrando que o percentual de SNs descritores que atenderam e

não atenderam a esse critério são próximos, fazendo inferir que esse critério não ajudou a separar SNs descritores de SNs não descritores, não se mostrando útil para a seleção.

O segundo critério de adição, o de IDENTIFICAÇÃO DE SNs POR MEIO DE CONJUNÇÃO ENTRE ADJETIVOS, do mesmo modo que o critério anterior, não se mostrou útil. Este critério ainda alcançou médias de revocação e precisão bem mais baixas, onde a revocação ficou em 0,9% e a precisão em 7,2%. Em suma, esse critério, bem como o anterior de adição, não se mostraram úteis para a seleção de SNs descritores.

Em relação ao critério “NÍVEL”, verificou-se que o mesmo se mostrou pertinente. Apesar de não ter alcançado, em média, uma taxa de revocação tão alta, a precisão foi bem acima da média. Como, convencionalmente, na prática o que se pode ter é um equilíbrio entre a revocação e precisão, acredita-se que esse critério obteve, de forma satisfatória, esse equilíbrio, pois, em média, alcançou uma taxa de revocação de 72,0% e uma precisão de 36,6%. Grande parte dos SNs descritores encontravam-se nos SNs de nível 2 ou mais, o que contribui para o alcance dessas taxas, tendo em vista que o ponto de corte utilizado nesse critério foi justamente o de selecionar os SNs que possuíam nível dois (2) ou mais.

O critério “POSIÇÃO/ORDEM DOS SINTAGMAS” foi um critério que surgiu no decorrer da aplicação dos outros critérios, pois percebeu-se que uma boa parte dos SNs descritores de cada documento encontravam-se entre os primeiros SNs extraídos de cada documento. Nesse contexto, verificou-se os dez primeiros SNs extraídos de cada documento. Identificou-se, em média, uma revocação de 34,7% e uma precisão de 45,0%. Apesar da baixa revocação alcançada por este critério, a precisão foi bem acima da média, demonstrando assim sua utilidade. Isso demonstra a existência de uma relação entre os primeiros sintagmas extraídos com os SNs descritores. Essa relação pode ser facilmente entendida, pois os primeiros sintagmas foram extraídos do título do documento, ou seja, um elemento que, convencionalmente, traz expressões bem representativas do conteúdo do documento ao qual pertence. Conclui-se que esse critério, apesar de necessitar mais estudo, deve ser considerado um elemento subsidiário a seleção de SNs descritores. É pertinente ressaltar, que diferentemente dos outros critérios, esse especificamente não foi estudado em profundidade, nem aplicado em outros contextos, o que demonstra a necessidade de aplicá-lo em outros ambientes e em até outro tipo de *corpus*. Todavia, pela capacidade de ser preciso, esse critério foi considerado útil.

No que se refere aos critérios de ordenação, o primeiro critério, o de FREQUÊNCIA DE OCORRÊNCIA DOS SNs NO TEXTO/DOCUMENTO E FREQUÊNCIA

NORMALIZADA, alcançou, em média, uma baixa revocação, ficando em torno de 9,6%. Isso se deu pelo fato de que a maior parte dos SNs descritores não se repetiram, ou seja, não tiveram uma frequência de ocorrência acima de um (1). Já a precisão se mostrou bem melhor, ao alcançar, em média, uma taxa de 36,7%. A baixa revocação, como já mencionado, pode ser justificada pelas características próprias do corpus utilizado, o qual possuía poucas chances de repetições, uma vez que não é tão convencional repetir expressões dentro de um título e de um resumo, talvez esse comportamento fosse diferente se o corpus tivesse sido bem maior, como, por exemplo, artigos por completo, ou dissertações, etc. Levando em conta mais a precisão, nesse critério específico, conclui-se que esse critério se mostrou útil.

O segundo e último critério de ordenação, o de FREQUÊNCIA DE OCORRÊNCIA DOS SNs NO CONJUNTO DE DOCUMENTOS – FREQUÊNCIA INVERSA DE DOCUMENTO – *idf*, apresentou boas taxas de revocação e precisão, demonstrando assim a sua utilidade na seleção de SNs descritores. A revocação alcançada com este critério ficou, em média, em torno de 99,2% e a precisão em torno de 25,6%.

Em suma, apesar de alguns critérios terem alcançado desempenhos melhores do que outros, a aplicação de todos esses critérios foi relevante, pois permitiu a visualização do comportamento deles quando aplicados ao corpus específico estudado nessa dissertação. Mesmo com a verificação de que alguns não contribuíram efetivamente para a seleção de sintagmas nominais descritores, todo esse experimento se mostrou válido, pois só se pode afirmar que um critério é útil ou não para a seleção, fazendo uso dele em algum ambiente específico, e esse uso foi feito nesse experimento. Em um julgamento geral acerca da utilidade ou não dos critérios, pode-se inferir que os dois critérios de adição não se mostraram úteis para a seleção de SNs descritores no contexto deste experimento. Já em relação aos critérios de eliminação, pode-se dizer que os de eliminação de *stopwords* bem como o de eliminação de SN com pronomes se mostraram úteis. Em contrapartida, os de eliminação de SNs com numerais e SNs com adverbio não se mostraram úteis neste experimento. Já os critérios de ordenação se mostraram viáveis para a seleção de SNs descritores neste experimento, ressaltando que os mesmos apresentariam melhores comportamentos em outros contextos. O critério de nível, além de se mostrar útil, permitiu ratificar o que se encontrou na literatura acerca da relação entre nível e potencial discriminatório dos SNs. Por último, no que se refere ao elemento posição/ordem do SN, pode-se dizer que o critério de posição também obteve bons resultados.

5 CONSIDERAÇÕES FINAIS

As considerações feitas neste momento dizem respeito ao alcance dos objetivos propostos no início deste trabalho, bem como também as contribuições resultantes dessa pesquisa e possíveis estudos futuros.

Retornando aos objetivos elencados no início deste trabalho, tem-se:

- Investigar o processo de indexação automática por sintagmas nominais;
- Verificar quais são as características de um SN como um bom descritor documental;
- Identificar na literatura científica nacional metodologias para seleção de sintagmas nominais em textos em português, bem como os critérios de seleção de cada metodologia;
- Planejar experimento, onde os Sintagmas Nominais extraídos são categorizados quanto ao atendimento ou não de critérios de seleção propostos na literatura quanto ao valor como Descritores, quando semelhantes aos Descritores documentais resultantes da indexação manual;
- Avaliar, os critérios de seleção na indexação automática por meio de Sintagmas Nominais para teses e dissertações da área jurídica.

A seguir, expõem-se o alcance dos objetivos propostos neste trabalho, bem como o percurso utilizado para cada objetivo específico. As considerações que seguem também respondem às indagações que foram feitas na problemática desta pesquisa, as quais estão ligadas aos objetivos específicos supracitados.

Em relação ao primeiro objetivo específico, o mesmo foi alcançado por meio da elaboração do referencial teórico deste trabalho. A aproximação com os autores que estudaram direta e indiretamente a indexação automática por meio de SNs foi essencial para todo o desenrolar desta pesquisa, pois o entendimento da utilização dos SNs em indexação automática foi basilar para o desenvolvimento da pesquisa e o alcance dos outros objetivos. Como forma de complementar o que se alcançou com a literatura para o entendimento dessa nova metodologia de indexação automática, foi feito o experimento, o qual permitiu ratificar resultados de outros estudos acerca dessa temática. Além de todo o trabalho, o referencial teórico desta pesquisa, em especial, trás as informações essenciais para o entendimento da indexação automática por meio de SNs, bem como alguns resultados alcançados por pesquisadores com os estudos desenvolvidos em prol dessa metodologia.

O referencial teórico desta pesquisa, além de expor o que já se fez acerca da indexação que faz uso dos SNs como recursos de acesso à informação, traz uma visão dessa indexação dividida em etapas, igualmente à indexação manual, o que permite fazer analogia dessas indexações, vendo-as como processo realizado por etapas, fases. Essa forma de visualização da indexação automática por meio de SNs, em etapas, permite que se analise cada etapa individualmente, identificando até que ponto uma afeta outra, além de permitir uma análise conjunta de todas as etapas, contribuindo assim para os estudos voltados para essa temática. É importante que se deixe claro que a estruturação em formato de um processo com essas três etapas foi feita com o intuito de proporcionar uma melhor visualização como um processo análogo ao processo de indexação manual, onde é preciso que se dê a atenção necessária para essas três etapas, desenvolvendo mecanismos e ferramentas de identificação e extração de SNs, não deixando de lado a seleção, a qual se mostra crucial para finalização de todo o processo e, também, para o alcance de SNs que possam ser tidos como os sintagmas mais representativos do conteúdo de um determinado documento.

Depois de compreendida a indexação automática por meio de SNs, procedeu-se com a análise que a literatura da área faz acerca das características dos SNs, para que seja possível identificar as características dos SNs descritores documentais ou não, sendo assim o segundo objetivo específico deste trabalho. Assim, verificou-se, por meio da literatura e por meio do experimento realizado que os SNs que atuam como descritores documentais possuem algumas características particulares em contrapartida de outros SNs que não as possuem, como, por exemplo, o nível dos SNs, a frequência de ocorrência dos SNs, a frequência inversa de documento, e a posição dos SNs. Como pôde ser vista na análise dos resultados experimentais desta pesquisa, SNs que possuem níveis específicos possuem, em sua maioria, mais chances de serem utilizados como descritores documentais. Percebeu-se, também, que as frequências de ocorrência (absoluta e inversa) são características de SNs que contribuem efetivamente para a classificação de um SN como relevante ou não como descritor documental. Tanto o referencial teórico deste trabalho, quanto a análise dos resultados experimentais tratam das características dos SNs como descritores documentais.

No tocante ao terceiro objetivo específico desta pesquisa, o qual procurou na literatura nacional metodologias utilizadas para a seleção de SNs, bem como os critérios imbuídos em cada metodologia, pode-se dizer que a análise cuidadosa e exaustiva de materiais informacionais (dissertações, teses e artigos) disponíveis física e virtualmente foi o insumo

básico para o alcance do mesmo, pois foi possível aproximar-se de estudos que se voltaram direta e indiretamente para a questão da metodologia e dos critérios de seleção de SNs.

Tanto as metodologias de seleção de SNs como os critérios utilizados nas mesmas, como, por exemplo, frequência de ocorrência, critérios de eliminação, entre outros, serviram de base para a realização do experimento, onde foi possível perceber a utilidade ou não dos mesmos, permitindo, assim, um julgamento desses critérios utilizados por distintos autores, porém aplicados em um ambiente específico nesta pesquisa, identificando, assim, o comportamento dos mesmos em um ambiente específico, ou seja, no contexto de informações doutrinárias advindas de dissertações e teses da área jurídica. Assim, esta pesquisa permitiu perceber um desenvolvimento considerável de pesquisas voltadas para a utilização de SNs como recursos de acesso a informação.

A utilização dos SNs como termos de indexação vem se mostrando, por meio de diversas pesquisas, uma alternativa positiva e que contribui para buscas menos cansativas, exaustivas e mais precisas. Percebeu-se, também, que varias ferramentas vem sendo desenvolvidas para operacionalizar a extração automática, bem como também a seleção automática de SNs. Um ponto importante percebido é que existe uma contribuição mútua entres diferentes áreas (Computação, Linguística, Ciência da Informação) que possuem um mesmo propósito, que é justamente o de proporcionar avanços na área de organização e recuperação de informações por meio de SNs.

No que se refere ao quarto objetivo específico, voltado para o planejamento e desenvolvimento do experimento realizado, pode-se dizer que o mesmo foi alcançado. Apesar de que, no início, algumas ideias ainda estivessem sendo amadurecidas, todo o desenvolver do experimento foi feito retomando sempre que possível o que se tinha visto na literatura da área, o que foi interessante, pois permitiu fazer ligações entre a teoria e a prática, ou seja, o que se tinha visto, por exemplo, sobre a importância da frequência de um determinado SN em um documento como fator determinante do mesmo como um SN que possui maior poder discriminatório do que outros, foi operacionalizado por meio do experimento. Assim, pôde-se ver na prática o comportamento dos critérios utilizados em pesquisas anteriores na seleção de SNs. Esses critérios foram aplicados a um ambiente específico, o que permitiu tirar conclusões acerca da viabilidade ou não dos mesmos no que se refere à seleção de SNs. Verificou-se que a maioria dos critérios utilizados contribuiu para a seleção de melhores sintagmas nesse ambiente específico. Os resultados obtidos com a comparação de cada critério, dentro de cada grupo, demonstraram alguns critérios com desempenho melhor do que

outros, o que pois permite refletir sobre conclusões e considerações acerca de um ou outro critério de seleção.

Entrando no quinto e último objetivo específico desta dissertação, o qual procurou fazer uma síntese e uma reflexão acerca dos critérios de seleção aplicados no experimento realizado, chegou-se a algumas conclusões acerca de cada critério utilizado para selecionar SNs descritores e evitar SNs não descritores. Assim, percebeu-se que os dois critérios de ordenação se mostraram úteis para a seleção de SNs descritores, ressaltando que o critério de “Frequência inversa de documento - *idf*.” obteve melhor desempenho quando comparado com o de “Frequência normalizada”. No tocante aos critérios de adição, os dois utilizados neste trabalho não serviram e conseqüentemente não contribuíram para a seleção de SNs, se mostrando assim inviáveis neste contexto. Dos quatro critérios de eliminação, dois se mostraram úteis e dois não alcançaram bons desempenhos. O critério de eliminação de *stopword* e o de eliminação de SNs com pronomes mostraram-se de utilidade, já os de eliminação de descarte de SNs com numerais e descarte de SNs que iniciavam com advérbio não alcançaram resultados satisfatórios. No que se refere ao nível do SN, pôde-se perceber que realmente o nível dos SNs tem relação com o seu potencial semântico, como já ressaltado por Souza (2006). Outro critério que se mostrou viável, ainda que pouco estudado até o momento, foi o de posição do SN, o qual evidenciou que a posição do SN tem relação com o seu potencial discriminatório. Este critério se mostrou útil nesta pesquisa.

Todas as considerações acerca dos critérios mencionados anteriormente são feitas para o contexto de um domínio específico (Direito) e com materiais informacionais específicos dentro desse domínio (informações doutrinárias – Dissertações e Teses). Sendo assim, a utilidade ou não de um determinado critério nessa pesquisa não pode ser generalizada para outras pesquisas desenvolvidas em outros contextos, pois cada domínio comporta-se de maneira distinta.

Nesse contexto, espera-se que este trabalho venha a contribuir para os estudos sobre a indexação manual e indexação automática por meio de SNs, principalmente no tocante à fase de seleção dos SNs.

Acredita-se, com base nas considerações feitas acerca da utilidade de cada critério de forma individual, que o julgamento dos critérios analisados nesta pesquisa contribuem diretamente para o desenvolvimento de uma metodologia própria de avaliação e seleção de SNs neste domínio específico, pois pôde-se perceber quais critérios foram úteis e quais não foram, contribuindo para a seleção de SNs relevantes e a eliminação de SNs irrelevantes.

Nesse contexto, é importante que se atente para o domínio específico estudado, bem como também para o gênero documental utilizado neste trabalho. Assim, o julgamento dos critérios mencionados neste trabalho, mais especificamente, no experimento, são direcionados para um domínio específico e também para uma tipologia documental específica. O julgamento do comportamento de um ou de outro critério foi levantado com base em um tipo de documento específico, com superestrutura específica. Domínios diferentes e superestruturas diferentes de documentos podem conduzir a resultados distintos no que se refere ao comportamento dos critérios.

O presente trabalho apresenta as seguintes limitações, fruto do percurso metodológico adotado no desenvolvimento da pesquisa:

- A atribuição de palavras-chaves pelos indexadores que não constam no texto do documento analisado – como a indexação realizada pelos indexadores foi uma indexação livre, verificou-se uma grande incidência de palavras-chave que não constavam nos títulos e nem nos resumos, constituintes do *corpus* deste trabalho.
- O método de seleção de cinco palavras-chaves como descritores das atribuídas pelos indexadores e autor a cada documento – esse quantitativo de cinco palavras-chave foi escolhido, pois era o quantitativo comumente utilizado pelos indexadores participantes deste trabalho, bem como também é o número encontrado convencionalmente de termos que são utilizados na indexação manual.
- A definição de SNs descritores – foi utilizado o critério de semelhança para a definição dos SNs descritores, onde se comparou os SNs e as palavras-chave de cada documento, ou, ainda, os SNs que continham as palavras-chaves em suas estruturas.
- Extração de SNs manualmente – a extração manual dos SNs identificados de forma automática pelo software PALAVRAS foi uma das etapas mais exaustivas desenvolvidas neste experimento, uma vez que foi um trabalho delicado, pois o referido software apresentou algumas falhas tanto na identificação de SNs como na omissão dos mesmos. Como não se tinha à disposição um extrator automático dos SNs, a saída era realizar a extração de todas as estruturas identificadas como SNs manualmente. Todo essa etapa se tornou ainda mais cansativa e exaustiva, devido aos erros encontrados, sendo

necessário que ao passo que cada SN ia sendo extraído, procedia-se com a conferência do mesmo no título e no resumo de cada documento, ou seja, sendo uma atividade de conferência, tendo em vista não omitir SNs que poderiam ou não serem descritores documentais.

- Aplicação de critérios manualmente – o processo que envolveu a aplicação dos critérios de seleção foi bem delicado, uma vez que foi a etapa crucial de todo o experimento, pois merecia uma atenção especial, tendo em vista que a categorização errônea de um SN, fosse num critério de eliminação, ou adição, por exemplo, conduziria a resultados errôneos e conseqüentemente em conclusões erradas acerca da utilidade ou não de cada critério. Toda a aplicação dos critérios foi feita manualmente, o que exigiu do pesquisador um cuidado maior, além de uma revisão ao final da aplicação de cada critério, pois como ser humano, o pesquisador está sujeito a erros, sendo assim, necessário revisões e, em alguns casos, correções.
- Avaliação isolada de cada critério – não foi avaliado os critérios em conjunto ou combinados, ou seja, cada critério foi aplicado e analisado individualmente.
- A análise dos critérios através das métricas de precisão e revocação – todos os resultados quantitativos do experimento realizado foram alcançados por meio das métricas de revocação e precisão, as quais são utilizadas nessa área de pesquisa. Como pôde ser visto a análise dos resultados experimentais, em alguns casos apenas a taxa de revocação, ou precisão, por exemplo, não eram suficientes para o julgamento de um ou de outro critério, necessitando, assim, de análises e interpretações qualitativas, para que se chegasse a um julgamento mais preciso.

Em relação a trabalhos futuros, acredita-se que podem ser desenvolvidos estudos nas seguintes vertentes:

- Continuidade dessa pesquisa com aplicação em um *corpus* maior, identificando o comportamento de cada critério em textos completos e não somente nos títulos e resumos dos documentos.
- Continuidade desta pesquisa, com algumas alterações, especialmente na definição das palavras-chave descritoras, ou seja, buscando realizar a indexação com mais consistência, fazendo uso de uma maior padronização da

etapa inicial da indexação manual. À definição das palavras-chave, bem como o etapa inicial do experimento, a qual consistiu na indexação manual realizada por quatro Bibliotecários, em trabalhos futuros deve ser dada uma maior atenção, tendo em vista que todas as etapas posteriores do experimento seriam refletias por essa primeira fase, sendo assim necessária uma maior padronização na etapa da indexação manual.

- Análise e definição mais precisas dos SNs como SNs descritores e SNs não descritores, buscando outras fontes que sirvam de referência para essa categorização.
- Aplicação dos critérios de seleção estudados nessa pesquisa em outros ambientes (domínios científicos), com vistas a verificar a generalidade das conclusões sobre a utilidade de alguns critérios
- Aprofundamento nos estudos voltados para a seleção de SNs descritores nesse mesmo ambiente, podendo desenvolver metodologia de seleção de SNs para a informação doutrinária, ou seja, dissertações e teses da área de Direito escritas em português. Bem como também, avaliação dos critérios em conjunto.

REFERÊNCIAS

- ARAUJO, Eliany Alvarenga; OLIVEIRA, Marlene de. A produção de conhecimentos e a origem das bibliotecas. In.: OLIVEIRA, Marlene de. (Organizadora). **Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação**. 2. Ed. Belo Horizonte: UFMG, 2011.
- BAXENDALE, P. Machine-made index for technical literature - an experiment. **IBM Journal of Research Development**, v. 2, n.4, p. 354–361, 1958.
- BICK, E. **The Parsing System Palavras**: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. 2000. 505 f. Tese (Doutorado em Linguística), Department Of Linguistics, University Of Aarhus, Aarhus, 2000
- BORKO, H. Information science: what is it? **American Documentation**, Chicago, v.19, n.1, p.3-5, Jan. 1968.
- BORGES, Graciane Silva Bruzina; MACULAN, Benildes Coura Moreira dos.; LIMA, Gercina Ângela Borém de. Indexação Automática e Semântica: estudo da análise do conteúdo de teses e dissertações. **Informação & Sociedade: estudos**, João Pessoa-PB, v. 18, n.2, p. 181-193, mai./ago. 2008.
- BRASIL. SUPERIOR TRIBUNAL DE JUSTIÇA. Sistema de Consulta ao vocabulário jurídico – Tesouro Jurídico. Disponível em: www.stj.jus.br/SCON/thesaurus. Acesso em: 04 fev. 2015.
- BRASIL. TRIBUNAL SUPERIOR ELEITORAL. Tesouro da Justiça Eleitoral. Organização e revisão técnica de Ormar Carmo Arouck. 7.ed. Brasília: Tribunal Superior Eleitoral, Secretaria de Gestão da Informação, 2010. Disponível em: http://www.tse.jus.br/hotSites/CatalogoPublicacoes/pdf/tesouro/Tesouro_web_v2.pdf. Acesso em: 06 fev. 2015.
- BRAZ, Márcia Ivo. **Dispositivos de memória para informação jurídica**: análise de procedimentos de indexação. 2013, 190 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação, Departamento de Ciência da Informação, Universidade Federal de Pernambuco – UFPE, Recife-PE, 2013.
- BRITO, Marcilio de. Sistemas de Informação em linguagem natural: em busca de uma indexação automática. **Ci. Inf.**, Brasília, 21(3): 223-232, set./dez. 1992.
- CÂMARA JÚNIOR, A. T. **Indexação automática de acórdãos por meio de processamento de linguagem natural**. Brasília, DF, 2007. 142 f. Dissertação (Mestrado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação da Universidade de Brasília. Brasília, DF, 2007.
- CHAVES, Rodrigo Soares. **Agrupamento de documentos eletrônicos por meio de Sintagmas Nominais**. 2013. Dissertação (Mestrado Profissional em Sistemas da Informação e Gestão do Conhecimento) – Universidade Fundação Mineira de Educação e Cultura. Belo Horizonte, 2013.

CHAUMIER, J. **Analisis y lenguajes documentales**. Barcelona: Mitre, 1986.

CORNU, Gérard. (1990) **Linguistique juridique**. Paris: Monchrestien, 1990.

CORRÊA, et al. Indexação e recuperação de teses e dissertações por meio de sintagmas nominais. **AtoZ**, Curitiba, v. 1, n. 1, p. 11-22, jan./jun. 2011.

DIAS, Eduardo José Wense. Análise de assunto: percepção do usuário quanto ao conteúdo de documentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 2, p. 146-157, jul./dez. 2004.

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

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

FARROW, J. All in the mind: concept analysis in indexing. *The indexer*, v.19, n.4, Oct. 1995. p.243-247.

FUJITA, Mariângela Spotti Lopes. Sistema de Indexação PRECIS (Continuação). **R. Bras. Bibliotecon. E Doc.**, São Paulo, 21(3/4): 7-33, jul./dez. 1988.

FUJITA, Mariângela Spotti Lopes; RUBI, Milena Polsinelli. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. **DataGramZero**, Rio de Janeiro, v. 7, n. 3, jun. 2006.

FUJITA, Mariândela Spotti Lopes; GIL LEIVA, Isidoro. As linguagens de indexação em bibliotecas nacionais, arquivos nacionais e sistemas de informação na América Latina. In.: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 16., 2010, Rio de Janeiro. **Anais...**Rio de Janeiro: UFRJ, 2010. disponível em: <http://www.leffa.pro.br/textos/abnt.htm>. Acesso em: 05 maio 2015. não paginado.

GIL, Antonio Carlos. **Métodos e Técnicas de Pesquisa Social**. 4.ed. São Paulo: Atlas, 2002.

GIL LEIVA, Isidoro. Aspectos conceituais da indexação. In.: **Política de indexação**. Isidoro Gil Leiva e Mariângela Spotti Lopes Fujita [editores]. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. Cap.2, p. 31- 106.

GIL LEIVA, I. **Lá automatización de lá indizacion, propuesta teórico-metodológica**: aplicación al área de Biblioteconomía y Documentación. 1997. 268 f. Tese – Universidad de Murcia, Murcia, España, 1997. Disponível em <digitum.um.es/xmlui/handle/10201/13286?mode=full>. Acesso em: 21 dez. 2014.

GIL LEIVA, I. Consistencia **en la asignación de materias en bibliotecas públicas del Estado**. *Boletín de la Asociación Andaluza de Bibliotecarios*, n.63, p.69-96, 2001. Disponível em: <[http:// webs.um.es/isgil](http://webs.um.es/isgil)>. Acesso em: 13 jan. 2015..

GIL LEIVA, I.; RODRÍGUEZ MUÑOZ, J. V. **Los orígenes del almacenamiento y**

recuperación de información. Boletín de la Asociación Andaluza de Bibliotecarios, Málaga, n. 42, p. 51-66, 1996. Disponível em: <<http://webs.um.es/isgil/>>. Acesso em: 21 dez. 2014.

GIL LEIVA, Isidor; RUBI, Milena Polsinelli; FUJITA, Mariângela Spotti Lopes. **Consistência na indexação em Bibliotecas Universitárias Brasileiras.** Transinformação, Campinas, v. 20, n. 3, p.233-253 set./dez., 2008. Disponível em: < <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/523>>. Acesso em: 12 jan. 2015.

GOMES, H. E. **Classificação, tesouro e terminologia:** fundamentos comuns. Disponível em: <http://www.conexao.org/bit/tertulia/tertulias.htm>>. Acesso em: 04 fev. 2015.

GONSALVES, Elisa Pereira. Conversas sobre iniciação à pesquisa científica. 4.ed. revisada e ampliada. Campinas, SP: Alínea, 2007.

GUIMARÃES, Jose Augusto Chaves. **Abordagens teóricas de tratamento temático da informação (TTI):** catalogação de assunto, indexação e análise documental. Ibersid. (2009) 105-117. ISSN 1888-0967.

GUIMARÃES, José Augusto Chaves; SALES, Rodrigo de.; GRACIO, Maria Claudia. **DataGramaZero** – Revista de Ciência da Informação, v. 13, n. 06, dez. 2012.

INSTITUTO Brasileiro de Informação em Ciência e Tecnologia – IBICT. Biblioteca Digital Brasileira de Teses e Dissertações – BDTD. Disponível em: <www.bdttd.ibict.br>. Acesso em: 21 maio 2014.

KURAMOTO, Hélio. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. **Ciência da Informação**, v. 25, n. 2, p. 1- 18, 1995.

KURAMOTO, Hélio. Sintagmas Nominais: uma nova proposta para a recuperação de informação. **DataGramaZero** – revista de ciência da informação. Rio de Janeiro, v. 3, n. 1, fev. 2002.

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

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

LAPA, Remi Correa. **A indexação automática no Brasil no âmbito da Ciência da Informação (1973-2012):** o estado da arte. 2014, 151 f. Dissertação (Mestrado em Ciência da Informação) – Centro de Artes e Comunicação. Departamento de Ciência da Informação, Universidade Federal de Pernambuco, Recife-PE, 2014.

LE COADIC, Y. **A ciência da informação.** Tradução de Maria Yêda F. S. de Filgueiras Gomes. 2. ed. Brasília, DF: Briquet de Lemos, 2002.

LE GUERN, Michel. **Unanalyseur morpho-syntaxique pour l'indexation automatique.** Le Français Moderne, juin, 1991

LOPES, E. F. Avaliação de serviços de indexação e resumo: critérios, medidas e metodologia. **R. Esc. Biblioteconomia.**, Belo Horizonte, v.14, n. 2, p. 242-256, set. 1985.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, Brasília, v. 31, n. 1, p. 41-52, jan./abr. 2002.

LOPES, Lucelene. **Extração automática de conceitos a partir de textos em língua portuguesa**. 2012, 156 f. Tese (Doutorado em Ciência da Computação). Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2012.

MAIA, Luiz Cláudio Gomes. **Uso de Sintagmas Nominais na classificação automática de documentos eletrônicos**. 2008, 158 f. Tese (Doutorado em Ciência da Informação). – Escola de Ciência da Informação, Universidade Federal de Minas Gerais Minas Gerais, 2008.

MACIEL, Anna Maria Becker. Terminologia, linguística de corpus e organização da informação legislativa e jurídica. In: Ciclo de conferências UNILEGIS, 2009, Brasília. **Apresentação**. Brasília, 2009. Brasília, 29-30 junho.

MARTINS, Agnaldo Lopes. **O uso do sintagma nominal na recuperação de documentos [manuscrito]**: proposta de um mecanismo automático para classificação temática de textos digitais. 2014, 192 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais Minas Gerais, 2014.

MIORELLI, Sandra Teresinha. **Extração do sintagma nominal em sentenças em português**. 2001. 98 f. Dissertação (Mestrado em Ciência da Computação) – Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2001.

MOREIRA, Alexandra. **Tesaurus e Ontologias**: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético. Belo Horizonte, 2003. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2003.

MORELLATO, Luana Vieira. **SIDSN**: sistema identificador de sintagmas nominais, 2007. 58 f. Monografia (Bacharelado em Ciência da Computação) – Departamento de Informática, Centro Tecnológico, Universidade Federal do Espírito Santo, Vitória, 2007.

NARUKAWA, Cristina Miyuki; GIL LEIVA, Isidoro; FUJITA, Mariângela Spotti Lopes. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia Decs na área de Odontologia. **Inf. & Soc.: Est.**, João Pessoa, v. 19, n 2, p. 99 – 118, maio/ago. 2009.

NAVES, Madalena Martins Lopes; KURAMOTO, Hélio. **Organização da informação**: princípios e tendências. Brasília (DF): Briquet de Lemos, 2006. 142 p.

OGMA. Disponível em: www.luizmaia.com.br/ogma. Acesso em 04 dez. 2014.

OLIVEIRA, Marlene de. (Coord.). **Ciência da informação e biblioteconomia**: novos conteúdos e espaços de atuação. Belo Horizonte: UFMG, 2011.

OTHERO, Gabriel de Ávila. **Grammar Play: um parser sintático em Prolog para a língua portuguesa**. 2004, 265 f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS, Porto Alegre, 2004.

PERINI, M. A. **Gramática descritiva do português**. 3.ed. São Paulo: Ática, 1998.

_____. **Gramática do português brasileiro**. São Paulo: Parábola editorial, 2010. 336 p.

REDIGOLO, Franciele Marques. **O processo de análise de assunto na catalogação de documentos: a perspectiva sociocognitiva do catalogador em contexto de Biblioteca Universitária**. 2010, 176 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, São Paulo, 2010.

ROBREDO, J. **Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas**. 4. ed. rev. e ampl. Brasília: Edição de autor, 2005.

RUBI, Milena Pousinelli; FUJITA, Mariângela Spotti Lopes. Elementos de política de indexação em manuais de indexação de sistemas de informação especializado. **Perspec. Ciênc. Inf.**, Belo Horizonte, v. 8, n. 1, p. 66-77, jan./jun. 2003.

RUBI, Milena Pousinelli. **Política de indexação para construção de catálogos coletivos em bibliotecas universitárias**. 2008. 166 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciência, Universidade Estadual Paulista, São Paulo, 2008.

SANTOS, Cícero Nogueira dos. **Aprendizado de máquina na identificação de sintagmas nominais: o caso do português brasileiro**. 2005. 104 f. Dissertação (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro, 2005.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SOUZA, Marcia Isabel Fugisawa ... [et al.]. Representação descritiva e temática no Sistema Agência de Informação Embrapa: controle de vocabulário. **Transinformação**. Pontifícia Universidade Católica de Campinas, v. 32, n. 1, p. 61-75, jan./abr., 2010.

SOUZA, Renato Rocha. **Uma proposta de metodologia para a escolha automática de descritores utilizando sintagmas nominais**. 2005. 197 f. Tese (Doutorado) – Curso de Doutorado em Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais – UFMG, Belo Horizonte, 2005.

_____. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**. Florianópolis, v. 11, n. esp., p. 42-59, 1º sem. 2006.

SOUZA, Renato Rocha; Alvarenga Neto, Rivadávia Correa Drummond de.; MENDES, Kellen Christina Ignácia. Mapeamento semântico através da análise de ocorrência de descritores sobre gestão do conhecimento. **Transinformação**, Campinas, v. 19, n.1, p. 19-30, jan./abr., 2007.

SOUZA, Renato Rocha; RAGHAVAN, K. S. Extraction of keywords from texts: an exploratory study using Noun Phrases. **Informação & Tecnologia (ITEC)**. Marília/ João Pessoa. v. 1, n. 1. p. 5-16, jan./jun., 2014.

SILVA, M. R.; FUJITA, M. S. L. **A prática de indexação**: análise da evolução de tendências teóricas e metodológicas. *Transinformação*, Campinas, v. 16, n. 2, p. 133-161, maio/ago. 2004

SILVA, Tiago Jose da. **Indexação automática por meio da extração e seleção de sintagmas nominais em textos em língua portuguesa**. 2014, 144 f. Dissertação (Mestrado)–Mestrado em Ciência da Informação, Departamento de Ciência da Informação, Universidade Federal de Pernambuco, Recife-PE, 2014.

UNIVERSIDADE Federal do Paraná – UFPR. Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação – BRAPCI. Disponível em: www.brapci.ufpr. Acesso em: 08 jun. 2014.

VAN SLYPE, G. **Lenguajes de indización**: concepción, construcción y utilización en los sistemas documentales. Trad. Pedro Hípola e Félix de Moya. Madrid: Fundación Germán Sánchez Ruipérez, 1991.

VIEIRA, Simone Bastos. Indexação automática e manual: revisão de literatura. **Cin. Inf.** Brasília, 17 (1): 43-57, jan./jun. 1988.

_____. Análise comparativa entre indexação automática e manual da literatura brasileira da Ciência da Informação. **R. Biblioteconomia**. Brasília, 16 (1): 83-94 jan./jun. 1988.

VIEIRA, Renata et al. Extração de sintagmas nominais para o processamento de coreferência. In: Encontro para o Processamento Computacional do Português Escrito e Falado (PROPOR), 5. **Anais...** São Carlos: ICMC/USP, 2000. p. 165-173.

APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO**

**TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(PARA MAIORES DE 18 ANOS OU EMANCIPADOS - Resolução 466/12)**

Convidamos o (a) Sr. (a) para participar como voluntário (a) da pesquisa “Dos Sintagmas Nominais aos Descritores Documentais: estudo de caso na indexação de Teses e Dissertações da área de Direito”, que está sob a responsabilidade do (a) pesquisador (a) Gustavo Diniz do Nascimento, com endereço Rua 19, Curado IV, Bloco 96, apto. 106, Jaboatão dos Guararapes/PE, – Telefone:(83) 8746-1506 e e-mail : dinizufcg@hotmail.com. Está sob a orientação do: Prof. Dr. Renato Fernandes Corrêa. Telefone para contato: (81) 2126-8000, e-mail: fc_renato@yahoo.com.br

Este Termo de Consentimento pode conter informações que o/a senhor/a não entenda. Caso haja alguma dúvida, pergunte à pessoa que está lhe entrevistando para que o/a senhor/a esteja bem esclarecido (a) sobre sua participação na pesquisa. Após ser esclarecido (a) sobre as informações a seguir, caso aceite em fazer parte do estudo, rubriche as folhas e assine ao final deste documento, que está em duas vias. Uma delas é sua e a outra é do pesquisador responsável. Em caso de recusa o (a) Sr. (a) não será penalizado (a) de forma alguma. Também garantimos que o (a) Senhor (a) tem o direito de retirar o consentimento da sua participação em qualquer fase da pesquisa, sem qualquer penalidade.

INFORMAÇÕES SOBRE A PESQUISA:

Para a execução da pesquisa será solicitado ao participante que o mesmo indexe um total de 30 documentos (dissertações e teses), salientando que o participante só terá em mãos o título e o resumo de cada documento para que assim prossiga com a indexação. Os riscos

são mínimos, ligado ao possível constrangimento de não querer se sentir a vontade para realização da indexação dos documentos. Como forma de minimizar esse possível constrangimento ou qualquer outro que possa surgir, o pesquisador deixará o participante a vontade, uma vez que será dado o prazo de até trinta (30) dias para que cada participante entregue a atividade feita. A atividade de indexação será feita de forma individual de acordo com a experiência e os conhecimentos de cada participante, bem como, também, levando em consideração a política de indexação da instituição, se existente.

Não estão previstos benefícios diretos. Os benefícios indiretos para os participantes da pesquisa serão possibilitar a geração de novos conhecimentos acerca da prática da indexação manual e automática, uma vez que propõe a reflexão sobre a área de estudo de representação temática da informação, com ênfase nas questões relacionadas com a indexação automática por meio de Sintagmas Nominais.

As informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos voluntários, sendo assegurado o sigilo sobre a sua participação. Os dados coletados nesta pesquisa, documentos indexados (Descritores), ficarão armazenados em pastas do computador pessoal, sob a responsabilidade do pesquisador, no endereço acima informado pelo período de no mínimo 5 anos.

O (a) senhor (a) não pagará nada para participar desta pesquisa. Se houver necessidade, as despesas para a sua participação serão assumidos pelos pesquisadores (ressarcimento de transporte e alimentação). Fica também garantida indenização em casos de danos, comprovadamente decorrentes da participação na pesquisa, conforme decisão judicial ou extra-judicial.

Em caso de dúvidas relacionadas aos aspectos éticos deste estudo, você poderá consultar o Comitê de Ética em Pesquisa Envolvendo Seres Humanos da UFPE no endereço: **(Avenida da Engenharia s/n – 1º Andar, sala 4 - Cidade Universitária, Recife-PE, CEP: 50740-600, Tel.: (81) 2126.8588 – e-mail: cepccs@ufpe.br).**

(assinatura do pesquisador)

CONSENTIMENTO DA PARTICIPAÇÃO DA PESSOA COMO VOLUNTÁRIO (A)

Eu, _____, CPF _____, abaixo assinado, após a leitura (ou a escuta da leitura) deste documento e de ter tido a oportunidade de conversar e ter esclarecido as minhas dúvidas com o pesquisador responsável, concordo em participar do estudo “Dos Sintagmas Nominais aos Descritores Documentais: estudo de caso na indexação de Teses e Dissertações da área de Direito”, como voluntário(a). Fui devidamente informado(a) e esclarecido (a) pelo(a) pesquisador (a) Gustavo Diniz do Nascimento sobre a pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Foi-me garantido que posso retirar o meu consentimento a qualquer momento, sem que isto leve a qualquer penalidade (ou interrupção de meu acompanhamento/ assistência/tratamento).

Local e data _____

Assinatura do participante: _____

Presenciamos a solicitação de consentimento, esclarecimentos sobre a pesquisa e o aceite do voluntário em participar.

Nome:

Assinatura:

Nome:

Assinatura:

**APÊNDICE B – CARTA DE ANUÊNCIA DAS DUAS INSTITUIÇÕES AS QUAIS OS
BIBLIOTECÁRIOS PARTICIPANTES ESTÃO LIGADOS**

CARTA DE ANUÊNCIA

Declaramos para os devidos fins, que aceitaremos (o) a pesquisador (a) Gustavo Diniz do Nascimento, a desenvolver o seu projeto de pesquisa, Dos Sintagmas Nominais aos Descritores Documentais: estudo de caso na indexação de Teses e Dissertações da área de Direito, que está sob a coordenação/orientação do(a) Prof. (a) Dr. Renato Fernandes Corrêa cujo objetivo é realizar por meio de dois Bibliotecários/ Indexadores a indexação de um conjunto de trinta documentos, tomando como fonte o título e o resumo dos mesmos. Esta autorização está condicionada ao cumprimento do (a) pesquisador (a) aos requisitos da Resolução 466/12 e suas complementares, comprometendo-se o mesmo a utilizar os dados obtidos com a indexação feita pelos pesquisados, exclusivamente para os fins científicos, mantendo o sigilo e garantindo a não utilização das informações em prejuízo das pessoas e/ou das comunidades. Antes de iniciar a coleta de dados o/a pesquisador/a deverá apresentar a esta Instituição o Parecer Consubstanciado devidamente aprovado, emitido por Comitê de Ética em Pesquisa Envolvendo Seres Humanos, credenciado ao Sistema CEP/CONEP.

Local, em ____/____/____.

APÊNDICE C – LISTA DE SINTAGMAS NOMINAIS QUE COMPUSERAM AS *STOPWORDS*

Neste apêndice são apresentados os Sintagmas Nominais que foram escolhidos dentre os SNs trabalhados, por serem bastante comuns e possuírem pouco valor conceitual. Foram considerados conjuntamente os sintagmas que apresentavam diferença no determinante em maiúsculo ou minúsculo, como, por exemplo, “os recursos” e “Os recursos”, optando-se pela forma minúscula.

a autora	esta dissertação	o presente estudo
a conclusão	esta pesquisa	o presente trabalho
a continuidade	este trabalho	o problema
a diferença	nosso estudo	o tema
a dissertação	o âmbito	o trabalho
a entrega da dissertação	o capítulo final	objeto de estudo
a justificativa desta pesquisa	O capítulo seguinte	os dados coletados
a pesquisa	O debate	os estudos teóricos
a presente dissertação de mestrado	o estudo	os questionamentos
A presente tese	o fim	os questionamentos levantados
a presente tese	o final	seu referencial teórico
a tese	o instituto	sua fundamentação teórica
as demandas	o marco teórico do nosso trabalho	um problema
Autor	o nosso trabalho	Um questionamento
esse conjunto	o objetivo	uma abordagem qualitativa
esse corte metodológico	o objetivo do estudo	Uma breve explanação
esse objetivo	o objeto da pesquisa	Uma dissertação
esse trabalho	o objeto do presente trabalho	uma ideia pertinente

**APÊNDICE D – PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS
(PRODUTO DA INDEXAÇÃO MANUAL)**

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 1	DOC. 2	DOC. 3
	Direitos Fundamentais	Direito Constitucional	Poder Judiciário
	Direito Constitucional	Direitos Fundamentais	Arbitragem
	Direito Tributário	Direitos Humanos	Decisão Judicial
	Contribuição Previdenciária	Relações Internacionais	Análise do Discurso
	Justiça Fiscal	Ordenamento Jurídico	Autoridade Judicial
BIBLIOTECÁRIO 2	DOC. 1	DOC. 2	DOC. 3
	Direito Civil	Direito Internacional	Poder judiciário
	Previdência Social	Direitos Humanos	Credibilidade
	Direito Tributário	Direitos Fundamentais	Imparcialidade
	Direitos Fundamentais	Constituição Federal	Análise de Discurso
	Contribuição Previdenciária	Normas Jurídicas	Retórica Jurídica
BIBLIOTECÁRIO 3	DOC. 1	DOC. 2	DOC. 3
	Previdência Social	Constituição Brasileira	Poder Judiciário
	Estado Moderno	Direitos Humanos	Direito Administrativo
	Contribuinte	Relações Internacionais	Direito Civil
	Impostos	Carta Magna	Constituição
	Direito Tributário	Normas Jurídicas	Reforma Judiciária
BIBLIOTECÁRIO 4	DOC. 1	DOC. 2	DOC. 3
	Receita estatal tributária	Constituição Federal	Poder Judiciário Brasileiro
	Contribuições Previdenciárias	Relações Internacionais	Crise das Funções Públicas
	Aumento nas Contribuições	Direitos Humanos	Decisões do Poder Judiciário
	Contribuições aos Direitos Fundamentais	Realidade Constitucional	Supremo Tribunal Federal
	Mudanças	Normas Jurídicas	Ordem Judicial
PRÓPRIO AUTOR	DOC. 1	DOC. 2	DOC. 3
	Contribuições Previdenciárias	Direitos Humanos - tratados internacionais	Crise de Autoridade
	Capacidade Contributiva	Direito Constitucional	Vontade
	Mínimo existencial		Discurso Judicial
	Direitos Fundamentais do Contribuinte.		

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 4	DOC. 5	DOC. 6
	Direito Civil	Direito	Direito Privado
	Direito	Filosofia	Direito Civil
	Direito Constitucional	Filosofia do Direito	Dignidade
	Entidade Familiar	Ética	Direito Constitucional
	Código Civil	Psicologia Jungniana	Personalidade
BIBLIOTECÁRIO 2	DOC. 4	DOC. 5	DOC. 6
	Direito Civil	Filosofia Ocidental	Direito Civil
	Código Civil	Pensamento Jurídico	Novo Código Civil
	Direito de Família	Realidade Jurídica	Direito Privado
	Entidades Familiares	Ética	Direitos Humanos
	Reconhecimento Parental	Onipotência Jurídica	Ordenamento Jurídico Brasileiro
BIBLIOTECÁRIO 3	DOC. 4	DOC. 5	DOC. 6
	Direito de Família	Princípios Jurídicos	Direito Privado
	Entidades Familiares	Direito	Relações Humanas
	Família Brasileira	Ética no Direito	Dignidade da Pessoa Humana
	Direito Brasileiro	Edgar Morin	Código Civil
	Código Civil	Segurança	Direito da Personalidade
BIBLIOTECÁRIO 4	DOC. 4	DOC. 5	DOC.6
	Direito de Família	Direito	Direito Privado
	Código Civil Brasileiro	Filosofia	Código Civil
	Entidades Familiares	Tradição Filosófica Ocidental	Princípio Constitucional da Dignidade
	Legislação Brasileira	Ética no Direito	Ordenamento Jurídico
	Constituição Federal	Testamento Jurídico	Tutela da personalidade
PRÓPRIO AUTOR	DOC. 4	DOC. 5	DOC. 6
	Dignidade Humana	Filosofia	Cláusulas Gerais
	Afetividade	Hermenêutica Filosófica	Dignidade da Pessoa Humana
	Entidades Familiares	Verdade-Causalidade	Personalidade

INDEXADORES	PALAVRAS-CHAVE ATRIBUIDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 7	DOC. 8	DOC. 9
	Direito do Trabalho	Direito Penal	Mandado de Segurança
	Tutela Jurisdicional	Compensação	Processo Penal
	Acesso à Justiça	Vítimas de Crime	Hermenêutica
	Direito Processual	Pena	Execuções
	Coisa Julgada	Penas Alternativas	Legitimidade
BIBLIOTECÁRIO 2	DOC. 7	DOC. 8	DOC. 9
	Proteção de Emprego	Direito Penal	Mandado de Segurança
	Acesso à Justiça	Persecução Criminal	Atribuição Pública
	Direito Processual Brasileiro	Legislação Penal	Legitimidade Passiva
	Reintegração Judicial	Justiça Criminal	Normas Jurídicas
	Direito do Trabalho	Aplicabilidade Penal	Duplo Grau Obrigatório
BIBLIOTECÁRIO 3	DOC. 7	DOC. 8	DOC. 9
	Direito Trabalhista	Justiça Criminal	Mandado de Segurança
	Direito Humano	Direito Penal	Ação Pública
	Relações Trabalhistas	Direito Criminal	Direito Coletivo
	Emprego	Código Penal	Ação Civil
	Reintegração do Trabalhador	Legislação Penal	Ação Judicial
BIBLIOTECÁRIO 4	DOC. 7	DOC. 8	DOC. 9
	Reintegração Judicial	Justiça Criminal Brasileira	Mandado de segurança
	Emprego	Vítimas de Crime	Interpretação de Normas Jurídicas
	Contrato Empregatício	Legislação Penal	Normas Jurídicas
	Indenização Reparatória	Penas	Legitimidade
	Restituição Integral	Medidas de Conciliação	Duplo Grau Obrigatório
PRÓPRIO AUTOR	DOC. 7	DOC. 8	DOC. 9
	Reintegração no Emprego	Justiça Criminal	Mandado de Segurança
	Tutela Jurisdicional	Satisfação das Vítimas	Duplo Grau Obrigatório
	Efetividade Processual		Postulado da razoabilidade
			Conflito de princípios
			Razão motivadora da regra que prevê o mandado de segurança contra atos de particulares
			Razão motivadora da regra do duplo grau obrigatório.

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 10	DOC. 11	DOC. 12
	Direito Tributário	Direito de Família	Direito Penal
	Princípio da Isonomia	Direito Fundamental	Princípio da Insignificância
	Direitos Humanos	Abandono Afetivo	Administração Pública
	Direito Constitucional	Poder Familiar	Corrupção Ativa
	Direito Natural	Abandono de Incapaz	Corrupção Passiva
BIBLIOTECÁRIO 2	DOC. 10	DOC. 11	DOC. 12
	Direito Tributário	Políticas Públicas	Tutela Penal
	Isonomia	Direito de Família	Direito Penal
	Direito Constitucional	Parto Anônimo	Direito Administrativo
	Direitos Fundamentais	Abandono de Recém-nascido	Crimes contra a Administração Pública
	Princípios Constitucionais	Responsabilidade Parental	Ética
BIBLIOTECÁRIO 3	DOC. 10	DOC. 11	DOC. 12
	Direito Tributário	Afetividade	Direito Penal
	Isonomia	Família	Direito Processual Penal
	Direitos Humanos	Direito da Família	Direitos Cíveis
	Constituição	Dignidade Humana	Administração Pública
	Justiça	Relação Familiar	Sociedade
BIBLIOTECÁRIO 4	DOC. 10	DOC. 11	DOC. 12
	Direito Tributário	Recém-nascido	Direito Penal Brasileiro
	República Federativa	Abandono Clandestino	Administração Pública
	Isonomia	Falta de Afetividade	Crimes
	Instrumentos Privados	Dignidade Humana	Conduta Criminosa
	Incentivos Fiscais	Parto Anônimo Brasileiro	Bens Jurídicos
PRÓPRIO AUTOR	DOC. 10	DOC. 11	DOC. 12
	Isonomia	Abandono	Direito Penal
	Direito Natural	Parto Anônimo	Princípio da Insignificância e Crimes contra a Administração Pública
	Direitos do Homem	Afetividade	Crimes Próprios
	Constituições Hodiernas	Dignidade Humana	Funcionário Público
			Doutrina Brasileira e Estrangeira
			Jurisprudência Nacional.

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 13	DOC. 14	DOC. 15
	Direito Penal	Direito Administrativo	Administração Pública
	Princípio da Efetividade	ANP	Ordenamento Jurídico
	Prescrição Antecipada	Petróleo	Globalização
	Poder Judiciário	Agência Reguladora	Agência Reguladora
	Prescrição Virtual	Entidade Reguladora	Reforma Administrativa
BIBLIOTECÁRIO 2	DOC. 13	DOC. 14	DOC. 15
	Prescrição Virtual	Pré-sal	Responsabilidade Civil
	Tutela Penal	Exploração do Petróleo	Modelo de Regulação
	Aplicação da Prescrição Virtual	Direito Ambiental	Ordenamento Jurídico Nacional
	Violação Constitucional	Meio-ambiente	Agências Reguladoras
	Celeridade da Justiça	Modelo Jurídico-regulador	Competência Normativa
BIBLIOTECÁRIO 3	DOC. 13	DOC. 14	DOC. 15
	Poder Judiciário	Constituição Brasileira	Código Civil
	Prescrição – Direito	Petróleo	Cirurgião Plástico
	Advogado	Crise Internacional	Agências Reguladoras
	Constituição Federal	Pré-Sal	Administração Pública
	Poder Judiciário	ANP	Regulamentações
BIBLIOTECÁRIO 4	DOC. 13	DOC. 14	DOC. 15
	Prescrição Virtual	Petróleo	Cirurgia Plástica
	Operadores Jurídicos Brasileiros	Exploração do Petróleo	Cirurgião Plástico
	Aplicabilidade da Prescrição Virtual	Agência Nacional de Petróleo	Agências Reguladoras
	Princípios Constitucionais	Pré-sal	Reforma do Estado
	Poder Judiciário	Biocombustíveis	Regulamentação
PRÓPRIO AUTOR	DOC. 13	DOC. 14	DOC. 15
	Prescrição Virtual	Direito do Petróleo	Administrativo
	Teorias da Prescrição	Marco Regulador	Agências Reguladoras
	Prescrição Retroativa	Pré-sal	Limites legais
	Garantias		Independência Normativa
	Efetividade Penal		Função reguladora

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 16	DOC. 17	DOC. 18
	Personalidade Jurídica	Filosofia do Direito	Clonagem
	Direito Privado	Hermenêutica	Biossegurança
	Função Social da Empresa	Nilismo	Biodireito
	Direito Constitucional	Direito	Reprodução Humana
	Função Social do Contrato	Norma Jurídica	Direito
	DOC. 16	DOC. 17	DOC. 18
	Desvio de Função Social	Teoria do Direito	Biodireito
BIBLIOTECÁRIO 2	Personalidade Jurídica	Hermenêutica Jurídica	Bioética
	Teoria do Direito	Pensamento Jurídico	Direito Constitucional
	Função Social da Propriedade	Teorias de Nietzsche x Delcuze	Direito à Vida
	Direito Empresarial	Filosofia do Direito	Biossegurança
	DOC. 16	DOC. 17	DOC. 18
BIBLIOTECÁRIO 3	Direito Empresarial	Filosofia Jurídica	Direitos Fundamentais
	Direito Privado	Filosofia Nilismo	Constituição
	Ordem Pública	Processo de Desenvolvimento Histórico	Lei da Biossegurança
	Autonomia Privada	Hermenêutica – Direito	Clonagem Terapêutica
	Direitos Fundamentais	Pensamento Jurídico	Dignidade da Pessoa Humana
	DOC. 16	DOC. 17	DOC. 18
BIBLIOTECÁRIO 4	Direito	Nilismo	Células-tronco
	Direito Privado	Filosofia	Direito Constitucional Brasileiro
	Valores Constitucionais	Jurisprudência Hermenêutica	Biossegurança
	Constituição Federal	Pensamento Teológico	Reprodução Humana Assistida
	Função Social da Empresa	Filosofia Jurídica	Clonagem Terapêutica
	DOC. 16	DOC. 17	DOC.18
PRÓPRIO AUTOR	Função Social da Empresa	Direito	Células-tronco
	Funcionalismo Jurídico	Nilismo	Clonagem Terapêutica
	Desconsideração da Personalidade Jurídica.	Secularização Ordem	Direitos Fundamentais
		Hermenêutica	

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 19	DOC. 20	DOC. 21
	Direito Penal	Direito Constitucional	Direito
	Crime contra a Honra	Audiências Públicas	Ordenamento Jurídico
	Bem Jurídico	Direitos Fundamentais	Internet
	Crime	Decisão Judicial	Coletividade
	Delito	Constituição Federal	Globalização
BIBLIOTECÁRIO 2	DOC. 19	DOC. 20	DOC. 21
	Proteção Jurídica da Honra	Audiências públicas	Sociedade Virtual
	Direito Penal	Supremo Tribunal Federal	Novo Ordenamento Jurídico
	Mudanças de Valores	Ação Direta de Inconstitucionalidade	Modificação do Sujeito de Direito
	Evolução História do Direito	Legitimidade Democrática	Crimes Virtuais
	Intervenção Mínima de Crimes contra à honra	Direitos Fundamentais	Sujeito de Direito
BIBLIOTECÁRIO 3	DOC. 19	DOC. 20	DOC. 21
	Direito Civil	Constituição Federal	Virtualização
	Bens Jurídicos	Legislação	Sociedade Virtual
	Direito Penal	Audiências Públicas	Direito Civil
	Proteção Penal	Jurisdição Constitucional	Internet
	Criminalização	Direitos Fundamentais	Informática
BIBLIOTECÁRIO 4	DOC. 19	DOC. 20	DOC. 21
	Direito Penal Brasileiro	Supremo Tribunal Federal	Práticas Jurídicas
	Honra	Direitos Humanos	Sociedade Virtual
	Órgãos Públicos	Políticas Públicas	Ordenamento Jurídico Autônomo
	Liberdade de Pensamento	Direitos à Saúde	Conceitos Jurídicos
	Abolição da Criminalização	Instituto Processual	Virtualização
PRÓPRIO AUTOR	DOC. 19	DOC. 20	DOC. 21
	Minimalismo Penal	Jurisdição Constitucional	Sujeito de Direito
	Intervenção Mínima	Audiência Pública	Virtualização
	Subsidiariedade	Legitimidade Democrática	SecondLife
	Honra		
	Controle jurisdicional		

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 22	DOC. 23	DOC. 24.
	Agências Reguladoras	Criminalidade	Ilicitude
	Direito Constitucional	Crime	Direito Penal
	Princípio Constitucional	Direito Penal	Crime
	Estado Democrático de Direito	Violência	Ordenamento Jurídico
	Administração Pública	Programas de Televisão	STJ
BIBLIOTECÁRIO 2	DOC. 22	DOC. 23	DOC. 24
	Agências Reguladoras Brasileiras	Mídia Exploratória	Consciência da Ilicitude
	Princípio da Legalidade	Construção Social da Criminalidade	Estrutura da Culpabilidade
	Direito Constitucional	Cultura da Intolerância	Direito Penal
	Estado Regulador	Programa Televisivo	Dogmática Penal
	Fundamento Jurídico-constitucional	Programa Bronca Pesada	Norma Jurídica
BIBLIOTECÁRIO 3	DOC. 22	DOC. 23	DOC. 24
	Direito Penal	Mídia Social	Ilicitude
	Princípio da Legalidade	Criminalidade	Direito Penal
	Princípios Constitucionais Penais	Pernambuco	Justiça Penal
	Ato Administrativo	Programa Policial	Criminalidade
	Legalidade	Bronca Pesada	Processo Penal
BIBLIOTECÁRIO 4	DOC. 22	DOC. 23	DOC. 24
	Agências Reguladoras Brasileiras	Mídia	Superior Tribunal de Justiça
	Função normativa das Agências Reguladoras	Programa Televisivo	Consciência da Ilicitude
	Fiscalização	Bronca Pesada	Culpabilidade
	Estado Democrático de Direito	Criminalidade	Crimes
	Fundamento Jurídico	Crimes	Consequências
PRÓPRIO AUTOR	DOC. 22	DOC. 23	DOC. 24
	Agência Reguladora	Enfoque do Etiquetamento	Consciência
	Função Normativa	Crime	Antijuridicidade
	Princípio da Legalidade	Mídia	Culpabilidade

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 25	DOC. 26	DOC. 27
	Direito Civil	Sonegação Fiscal	Direito Penal
	Direito de Família	Direito Tributário	Direito Administrativo
	União Estável	Crime Fiscal	Criminologia
	Entidade Familiar	Constituição Federal	Patente
	Homoafetivo	Estado Democrático de Direito	Marca
BIBLIOTECÁRIO 2	DOC. 25	DOC. 26	DOC. 27
	União Estável	Direito Tributário	Direito Penal
	Tutela do Direito à Diferença	Sonegação Fiscal	Tutela Penal
	União Homoafetiva	Crime Patrimonial	Falta de Legitimidade
	Reconhecimento da União Homoafetiva	Pagamentos de Tributos	Bens Culturais
	Direito de Família	Administração Tributária	Novos Bens Culturais
BIBLIOTECÁRIO 3	DOC. 25	DOC. 26	DOC. 27
	União Homoafetiva	Sonegação Fiscal	Direito Penal
	Entidade Familiar	Constituição Federal	Legitimidade
	Direito de Família	Crime Tributário	Bens Jurídicos
	Homossexualismo	Direito Penal	Criminalidade
	Direito Civil	Ação Penal Tributária	Direito Administrativo
BIBLIOTECÁRIO 4	DOC. 25	DOC. 26	DOC. 27
	União Homoafetiva	Sonegação Fiscal	Direito Penal Econômico
	Entidade Familiar	Carga Tributária	Ilegitimidade
	Proteção do Direito de Família	Desequilíbrios Econômicos	Patentes
	Direitos	Regras Tributárias	Sistemas Paralelos
	Argumentos Jurídicos	Patrimônio Estatal	Intervenção Penal
PRÓPRIO AUTOR	DOC. 25	DOC. 26	DOC. 27
	União Homoafetiva	Tributo	Direito Penal Econômico
	Constitucionalização	Sonegação	Bens Culturais
	Direito Civil	Patrimônio	Ilegitimidade
		Dever	Direito Sancionador
		Solidariedade	

INDEXADORES	PALAVRAS-CHAVE ATRIBUÍDAS AOS DOCUMENTOS		
BIBLIOTECÁRIO 1	DOC. 28	DOC. 29	DOC. 30
	Direito Civil	Incentivos Fiscais	Mandado de Segurança
	Direito Privado	Direito Tributário	Decisão Judicial
	Comércio Internacional	ONU	Dialogismo
	Contrato	CEPAL	Direito Processual
	Acordo Internacional	ICMS	Norma Jurídica
BIBLIOTECÁRIO 2	DOC. 28	DOC. 29	DOC. 30
	Direito Internacional	Direito Tributário	Análise do Discurso
	Direito Civil	Direito Internacional	Pensamento de Mikhail Bakhtin
	Direito Privado	Incentivos Fiscais	Discurso Enunciativo
	Autonomia da Vontade	Guerra Fiscal	Processo Enunciativo
	Reformulação no Direito privado	CEPAL	Análise do Mandado de Segurança
BIBLIOTECÁRIO 3	DOC. 28	DOC. 29	DOC. 30
	Direito Privado	Direito Tributário	Mandado de Segurança
	Direito Civil	Incentivos Fiscais	Ações Constitucionais
	Direito Econômico	Constituição Brasileira	Ciências Humanas
	Legislação	Impostos	Direito
	Direito Internacional Privado	Direito ao Desenvolvimento	Poder Judiciário
BIBLIOTECÁRIO 4	DOC. 28	DOC. 29	DOC. 30
	Direito Privado	Incentivos Fiscais	Mandado de Segurança
	Direito Civil	Constituição Brasileira	Decisão Judicial
	Privatização	Industrialização	Mikhail Bakhtin
	Relações Internacionais	Impostos	Enunciado Dialógico
	Relações Contratuais	Normas Tributárias	Produção Judicial
PRÓPRIO AUTOR	DOC. 28	DOC. 29	DOC. 30
	Dirigismo Contratual	Desenvolvimento	Enunciado Dialógico de Mikhail Bakhtin
	Autonomia da Vontade	Intervenção no Domínio Econômico	Cognição Judicial
	Contratos Internacionais	Incentivos Fiscais	Decisão Jurídica
	Constitucionalização	Guerra Fiscal	Alternância na Decisão Jurídica
	Direito Internacional Privado	ICMS	Mandado de Segurança

APÊNDICE E – PALAVRAS-CHAVE E SNS DESCRITORES DE CADA DOCUMENTO

PALAVRAS-CHAVE E SNS DESCRITORES	DOCUMENTOS		
	DOC. 1	DOC. 2	DOC. 3
PALAVRAS-CHAVE	Contribuição Previdenciária	Direitos Humanos	Poder Judiciário
	Direito Tributário	Normas Jurídicas	Análise de Discurso
	Direitos Fundamentais	Relações Internacionais	Crise das Funções Públicas
	Previdência Social	Constituição Federal	Supremo Tribunal Federal
	capacidade Contributiva	Direitos Fundamentais	Ordem Judicial
SNS DESCRITORES	capacidade contributiva nas contribuições à Previdência Social	Direitos fundamentais	a crise das funções públicas estatais
	as contribuições à Previdência Social	Os tratados internacionais sobre direitos humanos	o Poder Judiciário brasileiro
	a Previdência Social	direitos humanos	As decisões oriundas do poder judiciário
	direitos fundamentais do cidadão-contribuinte e justiça fiscal	a inteligência do art. 5º, § 2º da Constituição Federal	o poder judiciário
	os direitos fundamentais do contribuinte	art. 5º, § 2º da Constituição Federal	a técnica conhecida como análise de discurso
	Esses direitos fundamentais	a Constituição Federal	análise de discurso
a aplicação do princípio da capacidade contributiva	o futuro das relações internacionais diante de o atual quadro constitucional	a decisão do Supremo Tribunal Federal	
o princípio da capacidade contributiva	as relações internacionais diante de o atual quadro constitucional	o Supremo Tribunal Federal	
a capacidade contributiva	um plexo concatenado de normas jurídicas	pedidos de intervenção federal por descumprimento de ordem judicial	
	normas jurídicas	intervenção federal por descumprimento de ordem judicial	
		descumprimento de ordem judicial	
		ordem judicial	

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
PALAVRAS-CHAVE	DOC. 4	DOC. 5	DOC. 6
	Entidades Familiares	Direito	Direito Privado
	Direito de Família	Filosofia	Dignidade da Pessoa Humana
	Código Civil	Ética	Código Civil
	Direito Civil	Ética no Direito	Personalidade
	Dignidade Humana	Psicologia Junguiana	Direito Civil
SNs DESCRITORES	DOC. 4	DOC. 5	DOC. 6
	as entidades familiares afetivas	ética	a tutela dos direitos da personalidade por meio de a aplicabilidade direta do princípio da dignidade da pessoa humana nas relações de direito privado
	matéria de Direito de Família	complexidade no direito	os direitos da personalidade por meio de a aplicabilidade direta do princípio da dignidade da pessoa humana nas relações de direito privado
	Direito de Família	o direito	a personalidade por meio de a aplicabilidade direta do princípio da dignidade da pessoa humana nas relações de direito privado
	apenas três modos de entidades familiares	os campos da psicanálise, psicologia junguiana	a aplicabilidade direta do princípio da dignidade da pessoa humana nas relações de direito privado
	diversos outros tipos de entidades familiares	psicologia junguiana	o princípio da dignidade da pessoa humana nas relações de direito privado
	o código civil		a dignidade da pessoa humana nas relações de direito privado
	relevantes e inúmeras alterações trazidas por a Constituição Federal(cf), através do processo de constitucionalização do Direito Civil		a pessoa humana nas relações de direito privado
	a Constituição Federal(cf), através do processo de constitucionalização do Direito Civil		as relações de direito privado
	o processo de constitucionalização do Direito Civil		um estudo do sistema de proteção dos direitos da personalidade no ordenamento jurídico brasileiro
	constitucionalização do Direito Civil		o sistema de proteção dos direitos da personalidade no ordenamento jurídico brasileiro
	o Direito Civil		proteção dos direitos da personalidade no ordenamento jurídico brasileiro
	a dignidade humana de seus membros		os direitos da personalidade no ordenamento jurídico brasileiro
	o Novo Código Civil		a personalidade no ordenamento jurídico brasileiro
			a possibilidade de aplicação direta do princípio constitucional da dignidade da pessoa humana nas relações de direito privado
			aplicação direta do princípio constitucional da dignidade da pessoa humana nas relações de direito privado
			o princípio constitucional da dignidade da pessoa humana nas relações de direito privado
			a constatação da mudança paradigmática da codificação experimentada com a promulgação do Código Civil

			a mudança paradigmática da codificação experimentada com a promulgação do Código Civil
			a codificação experimentada com a promulgação do Código Civil
			a promulgação do Código Civil
			o Código Civil
			O sistema normativo de direito privado
			a personalidade humana
			A tutela da personalidade
			a personalidade
			o direito privado
			a aplicação direta do princípio da dignidade da pessoa humana nas relações de direito privado
			direito privado

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
	DOC. 7	DOC. 8	DOC. 9
PALAVRAS-CHAVE	Acesso à Justiça	Direito Penal	Mandado de Segurança
	Direito do Trabalho	Justiça Criminal	Duplo Grau Obrigatório
	Emprego	Legislação Penal	Legitimidade
	Reintegração Judicial	Vítimas de Crime	Normas Jurídicas
	Tutela Jurisdicional	Penas	Atribuição Pública
SNs DESCRITORES	DOC. 7	DOC. 8	DOC. 9
	nova dinâmica da reintegração judicial no emprego	os reais interesses das vítimas de crimes	duplo grau obrigatório do procedimento do mandado de segurança contra ato
	a reintegração judicial no emprego	as vítimas de crimes	o procedimento do mandado de segurança contra ato
	o emprego	a Justiça criminal	o exercício de atribuição pública
	o modelo brasileiro de proteção à relação de emprego	uma pesquisa de campo, baseada em entrevistas realizadas com vítimas de crimes	atribuição pública
	proteção à relação de emprego	baseada em entrevistas realizadas com vítimas de crimes	O Duplo Grau Obrigatório no procedimento Do Mandado De Segurança contra Ato de particulares
	a relação de emprego	entrevistas realizadas com vítimas de crimes	o procedimento Do Mandado De Segurança contra Ato de particulares
	emprego	vítimas de crimes	interpretação das normas jurídicas suas espécies e seus postulados
	a tutela jurisdicional reintegratória	o grau de satisfação da vítima com a justiça criminal	as normas jurídicas suas espécies
		satisfação da vítima com a justiça criminal	a análise do histórico do surgimento do mandado de segurança
		a vítima com a justiça criminal	o histórico do surgimento do mandado de segurança
		a legislação penal, tais como medidas de conciliação e a aplicação de outras penas que	o surgimento do mandado de segurança

		não a privativa de liberdade	
		medidas de conciliação e a aplicação de outras penas que não a privativa de liberdade	o mandado de segurança
		a aplicação de outras penas que não a privativa de liberdade	a legitimidade passiva da ação de mandado de segurança
		outras penas que não a privativa de liberdade	a ação de mandado de segurança
		a necessidade de sua participação na justiça criminal	mandado de segurança
		sua participação na justiça criminal	o instituto do duplo grau obrigatório
			o duplo grau obrigatório
			a interpretação possível do duplo grau obrigatório no procedimento de mandado de segurança
			o duplo grau obrigatório no procedimento de mandado de segurança
			o procedimento de mandado de segurança

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
	DOC. 10	DOC. 11	DOC. 12
	Direito Tributário	Dignidade Humana	Direito Penal
	Isonomia	Direito de Família	Administração Pública
	Direito Natural	Afetividade	Crimes Próprios
	Direitos Humanos	Parto Anônimo	Bens Jurídicos
	Direito Constitucional	Abandono de Recém-nascido	Princípio da Insignificância
	DOC. 10	DOC. 11	DOC. 12
	o princípio da isonomia no contexto do Direito Tributário	parto anônimo	sua aplicação nos crimes próprios contra a administração pública
	a isonomia no contexto do Direito Tributário	a afetividade	os crimes próprios contra a administração pública
	o contexto do Direito Tributário	o abandono de recém-nascidos	a administração pública
	o Direito Tributário	um contexto de consagração da afetividade como guia das relações familiares	o direito penal e a problemática de sua aplicação
	o princípio da isonomia	consagração da afetividade como guia das relações familiares	a administração pública nos seus aspectos patrimonial e moral
		a afetividade como guia das relações familiares	a administração pública como sujeito passivo
	a isonomia		
	o direito natural assimilado por as declarações de direitos do homem	a análise da viabilidade de implementação do parto anônimo	
	numa introdução à compreensão do princípio da isonomia	a viabilidade de implementação do parto anônimo	
	a compreensão do princípio da isonomia	implementação do parto anônimo	
	a relação entre o princípio constitucional da isonomia e as normas infraconstitucionais	o parto anônimo	
	o princípio constitucional da isonomia e as normas infraconstitucionais	o direito de família atual	

	a discussão sobre a concretização da isonomia	o princípio da dignidade humana	
	a concretização da isonomia	a dignidade humana	
	a mais flagrante violação da isonomia tributária.	a possível implementação do parto anônimo no país	
	a isonomia tributária		

PALAVRAS-CHAVE e SNs DESCRITORES	DOCUMENTOS		
	DOC. 13	DOC. 14	DOC. 15
PALAVRAS-CHAVE	Poder Judiciário	Pré-sal	Agências Reguladoras
	Prescrição Virtual	Petróleo	Administração Pública
	Operadores Jurídicos Brasileiros	ANP	Cirurgião Plástico
	Prescrição Retroativa	Exploração do Petróleo	Regulamentação
	Tutela Penal	Biocombustíveis	Responsabilidade civil
SNs DESCRITORES	DOC. 13	DOC. 14	DOC. 15
	prescrição virtual	petróleo no Brasil	responsabilidade civil do cirurgião plástico
	a prescrição virtual	a exploração das reservas de petróleo	o cirurgião plástico
	nova disciplina dada pelos operadores jurídicos brasileiros à ancestral causa de extinção de punibilidade	as reservas de petróleo	as agências reguladoras
	os operadores jurídicos brasileiros à ancestral causa de extinção de punibilidade	as áreas do pré-sal	mais variados setores da administração pública brasileira nos últimos seis anos
	a prescrição retroativa	a criação da pré-sal petróleo S.A. e da redefinição de algumas atribuições dos entes públicos envolvidos	a administração pública brasileira nos últimos seis anos
	a aplicabilidade da prescrição virtual	a pré-sal petróleo S.A.	principal paradigma das agências reguladoras nacionais
os benefícios decorrentes da adoção da nova técnica prescricional na solução dos problemas enfrentados pelo poder judiciário e os meios propostos pela doutrina para que o desiderato de desobstrução dos tribunais	o desenvolvimento da atividade, com especial destaque para a Agência Nacional de Petróleo, Gás Natural e Biocombustíveis ANP	as agências reguladoras nacionais	
a adoção da nova técnica prescricional na solução dos problemas enfrentados pelo poder judiciário e os meios propostos pela doutrina para que o desiderato de desobstrução dos tribunais	a atividade com especial destaque para a Agência Nacional de Petróleo, Gás Natural e Biocombustíveis ANP	um exercício concorrente de tal competência entre as agências reguladoras e o presidente	
a nova técnica prescricional na solução dos problemas enfrentados pelo poder judiciário e os meios propostos pela doutrina para que o desiderato de desobstrução dos tribunais	especial destaque para a Agência Nacional de Petróleo, Gás Natural e Biocombustíveis ANP	competência entre as agências reguladoras	
a solução dos problemas enfrentados pelo poder judiciário e os meios propostos pela doutrina para que o desiderato de desobstrução dos tribunais	a Agência Nacional de Petróleo, Gás Natural e Biocombustíveis ANP	as agências reguladoras brasileiras	
os problemas enfrentados pelo	a atividade de exploração e produção		

	poder judiciário e os meios propostos pela doutrina para que o desiderato de desobstrução dos tribunais	de petróleo no país	
	o poder judiciário e os meios propostos pela doutrina para que o desiderato de desobstrução dos tribunais	exploração e produção de petróleo no país	
		produção de petróleo no país	
		o advento do pré-sal	
		o pré-sal	

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
	DOC. 16	DOC. 17	DOC. 18
PALAVRAS-CHAVE	Função Social da Empresa	Niilismo	Clonagem Terapêutica
	Direito Privado	Direito	Biossegurança
	Direito Empresarial	Filosofia do Direito	Biodireito
	Personalidade Jurídica	Hermenêutica	Células-tronco
	Constituição Federal	Pensamento Jurídico	Direitos Fundamentais
SNs DESCRITORES	o desvio da função social da empresa	niilismo	a nova lei da biossegurança
	desconsideração da personalidade jurídica	hermenêutica como filosofia jurídica ativa	a biossegurança
	a desconsideração da personalidade jurídica	o fenômeno do niilismo	células-tronco embrionárias para fins terapêuticos
	as relações de Direito privado	o niilismo	a questão das sobras embrionárias e da clonagem terapêutica em confronto
	a desconsideração da personalidade jurídica a partir das principais teorias	seus reflexos na teorização do Direito	as sobras embrionárias e da clonagem terapêutica em confronto
	a personalidade jurídica a partir das principais teorias	a teorização do Direito	a clonagem terapêutica em confronto
	uma teoria funcional para a desconsideração da personalidade jurídica	o niilismo no âmbito	o tratamento ético e jurídico da questão das células-tronco embrionárias
	a personalidade jurídica	os tipos de niilismo	a questão das células-tronco embrionárias
	decorrência do descumprimento da função social da empresa	o pensamento jurídico	as células-tronco embrionárias
	o descumprimento da função social da empresa	a superação do niilismo no pensamento	embriões produzidos pelo processo de clonagem terapêutica
a função social da empresa	o niilismo no pensamento	o processo de clonagem terapêutica	
	a tentativa de recepção dos postulados da hermenêutica filosófica de Hans-Georg Gadamer para a compreensão do processo decisório de aplicação das normas jurídicas	clonagem terapêutica	
	recepção dos postulados da hermenêutica filosófica de	uma perspectiva dos Direitos fundamentais	

		Hans-Georg Gadamer para a compreensão do processo decisório de aplicação das normas jurídicas	
		os postulados da hermenêutica filosófica de Hans-Georg Gadamer para a compreensão do processo decisório de aplicação das normas jurídicas	os direitos fundamentais
		a corrente de pensamento intitulada jurisprudência hermenêutica	a interface da Bioética e do Biodireito com o Direito Constitucional
		contraste com essa corrente	a Bioética e do Biodireito com o Direito Constitucional
		a jurisprudência hermenêutica	o Biodireito com o Direito Constitucional
		a saída do ciclo vicioso do niilismo	a possibilidade de obtenção de células-tronco
		o ciclo vicioso no niilismo	obtenção de células-tronco
		o Direito como ordem e hermenêutica	direta repercussão no tratamento da clonagem terapêutica
		ordem e hermenêutica	o tratamento da clonagem terapêutica
		hermenêutica	a clonagem terapêutica
		uma filosofia ativa do Direito	os processos de obtenção de células-tronco embrionárias em confronto
		o Direito	obtenção de células-tronco embrionárias em confronto
		o niilismo no pensamento jurídico	células-tronco embrionárias em confronto
			a clonagem terapêutica utilizada de forma ponderada e com finalidade estritamente terapêutica

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
	DOC. 19	DOC. 20	DOC. 21
	Direito Penal	Audiência Pública	Sociedade Virtual
	Honra	Direitos Fundamentais	Virtualização
	Bens Jurídicos	Legitimidade Democrática	Sujeito de Direito
PALAVRAS-CHAVE	Intervenção Mínima	Jurisdição Constitucional	Internet
	Criminalização	Supremo Tribunal Federal	Ordenamento Jurídico
	DOC. 19	DOC. 20	DOC. 21
	a proteção jurídica da honra	a experiência das audiências públicas no Supremo Tribunal Federal como instrumento de ampliação do círculo de intérpretes da constituição	impactos da virtualização da sociedade no mundo jurídico
	a honra	as audiências públicas no Supremo Tribunal Federal como instrumento de ampliação do círculo de intérpretes da constituição	a virtualização da sociedade no mundo jurídico
	o direito penal mínimo	o Supremo Tribunal Federal como instrumento de ampliação do círculo de intérpretes da constituição	modificações no conceito de sujeito de Direito
SNs DESCRITORES	compatibilidade dos princípios da proteção de bens jurídicos e da intervenção mínima com	o instituto processual da audiência pública introduzido no direito brasileiro pelas leis nº 9.868/99 e 9.882/99	o conceito de sujeito de Direito

	os crimes contra a honra		
	os princípios da proteção de bens jurídicos e da intervenção mínima com os crimes contra a honra	a audiência pública introduzido no direito brasileiro pelas leis nº 9.868/99 e 9.882/99	sujeito de Direito
	a proteção de bens jurídicos e da intervenção mínima com os crimes contra a honra	a experiência do Supremo Tribunal Federal	a internet
	bens jurídicos e da intervenção mínima com os crimes contra a honra	o Supremo Tribunal Federal	o fenômeno da virtualização
	a intervenção mínima com os crimes contra a honra	a audiência pública da saúde	a virtualização
	os crimes contra a honra	tema relacionado à jurisdição constitucional	a sociedade virtual
	uma apreciação da intervenção mínima da sua natureza subsidiária do seu caráter fragmentário e do princípio da eficácia	a jurisdição constitucional	a sociedade virtual composta por sujeitos
	a intervenção mínima da sua natureza subsidiária do seu caráter fragmentário e do princípio da eficácia	a utilização das audiências públicas pelo Supremo Tribunal Federal	o processo da virtualização
	a proteção penal da honra	as audiências públicas pelo Supremo Tribunal Federal	um ordenamento jurídico autônomo
	imposição constitucional de sua criminalização	o conteúdo normativo dos direitos fundamentais e da definição de políticas públicas	o fenômeno da virtualização da sociedade
	a criminalização	os direitos fundamentais e da definição de políticas públicas	a virtualização da sociedade
	a inconstitucionalidade da criminalização no julgamento das demandas	legitimidade democrática ao Supremo Tribunal Federal	o conceito lógico-jurídico de sujeito de direito pela doutrina civilística brasileira
	a criminalização no julgamento das demandas		sujeito de direito pela doutrina civilística brasileira
	a abolição da criminalização		o sujeito de direito
	a criminalização		ordenamento jurídico nacional
			os direitos e deveres reconhecidos pelo ordenamento jurídico
			deveres reconhecidos pelo ordenamento jurídico
			o ordenamento jurídico

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
	DOC. 22	DOC. 23	DOC. 24
PALAVRAS-CHAVE	Princípio da Legalidade	Criminalidade	Direito Penal
	Estado Democrático de Direito	Crime	Consciência da Ilicitude
	Direito Constitucional	Bronca Pesada	Crime
	Agências Reguladoras Brasileiras	Mídia	Culpabilidade
	Agência Reguladora	Programa Televisivo	Ilicitude
SNs DESCRITORES	o princípio da legalidade	a mídia	a consciência da ilicitude
	a produção normativa das agências reguladoras brasileiras	um estudo sobre o programa televisivo bronca pesada	a ilicitude
	as agências reguladoras brasileiras	o programa televisivo bronca pesada	sua localização na estrutura da culpabilidade
	a mutação do princípio da legalidade	bronca pesada	a estrutura da culpabilidade
	as regras e os princípios previstos no estado democrático de direito	o estudo dos relatos veiculados pelo programa de televisão Bronca Pesada transmitido pela TV jornal em relação à criminalidade em Pernambuco	a culpabilidade
os princípios previstos no estado democrático de direito	os relatos veiculados pelo programa de televisão Bronca Pesada transmitido pela TV jornal em relação à criminalidade em Pernambuco	o crime	
o estado democrático de direito	o programa de televisão Bronca Pesada transmitido pela TV jornal em relação à criminalidade em Pernambuco	os temas mais complexos da teoria jurídica do crime	
	Bronca Pesada transmitido pela TV jornal em relação à criminalidade em Pernambuco	a teoria jurídica do crime	
	a TV jornal em relação à criminalidade em Pernambuco	a evolução da culpabilidade no âmbito da dogmática penal	
	a criminalidade em Pernambuco	a culpabilidade no âmbito da dogmática penal	
	o crime	a aferição da culpabilidade do sujeito	
	a condição da mídia	a culpabilidade do sujeito	
	o bronca pesada	o conceito da consciência da ilicitude	
	a problemática do crime	potencial consciência da ilicitude	
		o tema da consciência da ilicitude	

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
	DOC. 25	DOC. 26	DOC. 27
PALAVRAS-CHAVE	União Homoafetiva	Sonegação Fiscal	Direito Penal
	Direito Civil	Direito Tributário	Bens Culturais
	Direito de Família	Constituição Federal	Direito Administrativo
	Entidade Familiar	Crime Patrimonial	Ilegitimidade
	União Estável	Patrimônio Estatal	Patentes
SNs DESCRITORES	DOC. 25	DOC. 26	DOC. 27
	união estável	crime de sonegação fiscal	ilegitimidade da tutela penal da ordem econômica
	a tutela da união homoafetiva	sonegação fiscal	o direito penal econômico
	a união homoafetiva	o crime de sonegação fiscal em conformidade com o art. 1º da lei nº 8.137/1990	o aparecimento do direito penal econômico a partir das convenções de Paris
	entidade familiar	sonegação fiscal em conformidade com o art. 1º da lei nº 8.137/1990	o direito penal econômico a partir das convenções de Paris
	a influência do direito civil constitucional	crime patrimonial, material e de dano	marcas e patentes
	a proteção do direito de família	a sonegação fiscal	estes bens culturais
	o direito de família	o patrimônio estatal	o método usado no direito penal
	a união homoafetiva como entidade familiar	a constituição federal	o direito penal
			os bens culturais
			a mutabilidade dos bens culturais
			as garantias do direito penal
			o direito de punir para o direito penal econômico
			meio termo entre o direito penal e o direito administrativo
			o direito penal e o direito administrativo
		o direito administrativo	
		o direito penal e do direito administrativo	
		o direito penal de duas velocidades	
		sua ilegitimidade	

PALAVRAS-CHAVE E SNs DESCRITORES	DOCUMENTOS		
PALAVRAS-CHAVE	DOC. 28	DOC. 29	DOC. 30
	Direito Privado	Incentivos Fiscais	Mandado de Segurança
	Direito Civil	Direito Tributário	Decisão Judicial
	Autonomia da Vontade	Constituição Brasileira	Mikhail Bakhtin
	Direito Internacional Privado	Guerra Fiscal	Enunciado Dialógico
	Relações Contratuais	ICMS	Produção Judicial
SNs DESCRITORES	DOC. 28	DOC. 29	DOC. 30
	a autonomia da vontade	a legitimidade dos incentivos fiscais num contexto de guerra fiscal	a cognição no mandado de segurança sob o prisma dialógico de Mikhail Bakhtin
	a reformulação da atuação do estado nas relações contratuais privadas e sua repercussão no plano internacional	os incentivos fiscais num contexto de guerra fiscal	o mandado de segurança sob o prisma dialógico de Mikhail Bakhtin
	a atuação do estado nas relações contratuais privadas e sua repercussão no plano internacional	um contexto de guerra fiscal	a teoria do enunciado dialógico desenvolvido por Mikhail Bakhtin
	o estado nas relações contratuais privadas e sua repercussão no plano internacional	guerra fiscal	o enunciado dialógico desenvolvido por Mikhail Bakhtin
	as relações contratuais privadas e sua repercussão no plano internacional	a constituição brasileira de 1988	a produção judicial do direito
	o direito civil	a concessão de incentivos fiscais relativos ao imposto sobre operações relativas à circulação de mercadorias e prestações de Serviço de Transporte Interestadual e municipal e de comunicação (ICMs)	mandado de segurança
	um necessário compromisso internacional na análise de determinados institutos do direito privado	incentivos fiscais relativos ao imposto sobre operações relativas à circulação de mercadorias e prestações de serviço de transporte interestadual e municipal e de comunicação (ICMs)	a decisão judicial
	a análise de determinados institutos do direito privado	o imposto sobre operações relativas à circulação de mercadorias e prestações de serviço de transporte interestadual e municipal e de comunicação (ICMs)	um enunciado dialógico
	determinados institutos do direito privado	operações relativas à circulação de mercadorias e prestações de serviço de transporte interestadual e municipal e de comunicação (ICMs)	
	o direito privado	a circulação de mercadorias e prestações de serviço de transporte interestadual e municipal e de comunicação (ICMs)	
	autonomia da vontade	mercadorias e prestações de serviço de transporte interestadual e municipal e de comunicação (ICMs)	
	o processo de constitucionalização do direito privado	prestações de serviço de transporte interestadual e municipal e de comunicação (ICMs)	
	constitucionalização do direito privado	serviço de transporte interestadual e municipal e de comunicação (ICMs)	
	consequência direta a ampliação da atuação do estado nas relações contratuais privadas	transporte interestadual e municipal e de comunicação (ICMs)	
	a ampliação da atuação do estado nas relações contratuais privadas	as características básicas do ICMS	

	a atuação do estado nas relações contratuais privadas	o ICMS	
	o estado nas relações contratuais privadas	a concessão de incentivos fiscais pelos estados	
	as relações contratuais privadas	incentivos fiscais pelos estados	

APÊNDICE F – TÍTULOS E RESUMOS CONSTITUINTES QUE COMPUSERAM O CORPUS DESTA PESQUISA

<p>Capacidade contributiva nas contribuições à Previdência Social: direitos fundamentais do cidadão-contribuinte e justiça fiscal.</p> <p>O Estado, desde quando passou a ser denominado Estado Moderno, vivenciou inúmeras transformações. A fiscalidade, fenômeno essencial à existência do Estado, também tem sofrido mudanças, dentre as quais se destaca o grande aumento de contribuições percentualmente em relação à receita estatal tributária, que, devido a inúmeras razões, passam a substituir gradativamente os impostos diretos. Diante dessa nova perspectiva do fenômeno tributário, torna-se fundamental adequar as contribuições aos direitos fundamentais do contribuinte. O direito à existência, materializado pelo mínimo existencial, é um desses direitos fundamentais que deve ser inexoravelmente respeitado e preenchido, mas em seu aspecto ampliado que é justamente a condição humana, tal qual descreve Hannah Arendt. Surge, assim, a necessidade de harmonizar as contribuições previdenciárias do cidadão ao mínimo existencial, através da aplicação do princípio da capacidade contributiva, mediado não pelos seus subprincípios clássicos, como a progressividade, mas pela proporcionalidade.</p>
<p>Direitos fundamentais : constituição e tratados internacionais (análise do ordenamento jurídico brasileiro) .</p> <p>A presente tese tem por objetivo discutir em que nível hierárquico os tratados internacionais sobre direitos humanos ingressam na ordem interna, tendo em vista a polêmica hermenêutica envolvendo a inteligência do art.5º, § 2º da Constituição Federal. Toda o debate terá por substrato o ordenamento positivo brasileiro. Evidente que questões laterais serão levantadas, mas tendo sempre por centro gravitacional a Carta Magna em vigor. As críticas formuladas buscam despertar o poder reformador e os operados do direito para um questionamento sobre a realidade constitucional e o futuro das relações internacionais diante do atual quadro constitucional. Todo esforço empreendido tem por suporte uma hermenêutica sistêmica do texto magno na demonstração de que a obra do constituinte é um plexo concatenado de normas jurídicas, não sendo possível qualquer interpretação alheia a esse conjunto por mais altruísta que sejam seus propósitos..</p>
<p>Liberdade e decisão judicial : a autoridade da vontade como instrumento legitimador das razões jurídicas.</p> <p>Esta dissertação tem como pano de fundo a crise das funções públicas estatais, notadamente a crise de credibilidade que atravessa o Poder Judiciário brasileiro, constada em meio ao crescimento de caminhos alternativos para solução de conflitos, como o pluralismo de jurisdição, a mediação e a arbitragem. As decisões oriundas do poder judiciário vêm sendo apresentadas como estratégias de manutenção de poder, fazendo jus às convicções pessoais do julgador. O direito, não aquele idealizado pelas escolas dogmáticas, perdeu seu referencial teórico, realizando-se casuisticamente através de sentenças que julgam instintivamente, a pretexto de uma justiça ou moral, guiada pelo subjetivismo. Com efeito, o discurso oficial encontra base em elementos não dogmáticos que conduzem a tomada de decisão, muito embora o julgador apresente uma homilia ornada, fruto da utilização de ferramentas retóricas, sonogando as premissas de onde partiu. A crítica que se faz ao formalismo é devida à característica que possui de conseguir, através de fórmulas engendradas, esconder as vicissitudes de uma decisão. Utilizando a técnica conhecida como análise de discurso, a dissertação decompôs a decisão do Supremo Tribunal Federal, que denegou pedidos de intervenção federal por descumprimento de ordem judicial, tendo sido verificados inúmeros artifícios retóricos que servem à manipulação das razões jurídicas e aos exercícios argumentativos, criando a ilusão de certeza em meio a insegurança e incerteza.</p>
<p>O reconhecimento jurídico das entidades familiares afetivas: uma análise baseada no princípio fundamental da dignidade da pessoa humana.</p> <p>Objetivou esta dissertação, a princípio, demonstrar o quanto a lei brasileira, em matéria de Direito de Família, encontra-se defasada. Isto porque, a Carta Magna em vigor previu explicitamente apenas três modos de entidades familiares legítimas: o casamento, a união estável e a comunidade monoparental. Mas a realidade social aponta diversos outros tipos de entidades familiares, os quais não foram legislados, sendo todos baseados num elemento comum: o afeto. Para que isto restasse comprovado, analisou-se a família ocidental desde os seus primórdios até a influência desta na maneira de viver das famílias brasileiras. A partir de então, fez-se uma análise legislativa do assunto no Direito brasileiro, iniciando-se com o Código Civil (CC) do século anterior até o atual, passando por relevantes e inúmeras alterações trazidas pela Constituição Federal (CF), através do processo de constitucionalização do Direito Civil. Para reforçar a idéia defendida de que existem vários tipos de famílias, além daqueles previstos pelo constituinte, analisou-se o afeto como sendo o elemento indispensável e caracterizador de determinada relação entre pessoas, ser ou não uma família, para que só assim esteja assegurada a dignidade humana de seus membros. Trouxe-se ainda uma análise crítica do Novo Código Civil (NCC), o qual não disciplinou devidamente a matéria. Tudo isso para comprovar que há entidades familiares que não foram previstas explicitamente pelo legislador, mas, sim, de forma implícita, aquidenominadas de entidades afetivas.</p>
<p>Subjetividade, ética e complexidade no direito : a segurança que vem da admissão da insegurança: uma crítica à pressuposição de onipotência que subjaz às razões jurídicas.</p> <p>Buscaremos nesse trabalho ampliar canais comunicativos, conectar, sobretudo, filosofia e direito. Investigaremos novos paradigmas que nascem do pensamento absolutamente original de Edgar Morin, autor que fornece um sentido radicalmente novo à idéia de complexidade, rompendo com os conceitos tradicionais de verdade e causalidade. A partir dessa nova perspectiva, faremos uma crítica à tradição filosófica ocidental e mostraremos como pensamento jurídico encontra-se impregnado por uma razão ?mitológica?, que tenta a todo custo manter viva a ilusão de segurança e certeza do conhecimento. Vislumbraremos, com Morin, a possibilidade de emergência de uma nova forma de pensar e se relacionar com a realidade jurídica; referimo-nos a um modo de estar que aceita a insegurança, que acolhe a falta e daí faz nascer uma nova concepção de segurança. Para corroborar nossos argumentos faremos incursões nos campos da psicanálise, psicologia junguiana bem como da hermenêutica filosófica..</p>
<p>A tutela dos direitos da personalidade por meio da aplicabilidade direta do princípio da dignidade da pessoa humana nas relações de direito privado.</p> <p>O presente trabalho faz um estudo do sistema de proteção dos direitos da personalidade no ordenamento jurídico brasileiro com vistas a demonstrar a possibilidade de aplicação direta do princípio constitucional da dignidade da pessoa humana nas relações de direito privado. Esta possibilidade de apresenta a partir da constatação da mudança paradigmática da codificação experimentada com a promulgação do Código Civil de 2002 que adotou como uma das suas orientações metodológicas a técnica das cláusulas gerais. Por meio da efetividade das chamadas ?cláusulas gerais?, o sistema normativo de direito privado deixa de ser um sistema fechado e passa a estabelecer conectividade com todo o ordenamento. A personalidade humana não se restringe a um requisito formal do sujeito de direito. Ela é mais do que isso, é um valor juridicamente tutelado. A tutela da personalidade confere unidade ao ordenamento jurídico, e torna evidente a superação da tradicional dicotomia entre o direito público e o direito privado. A aplicação direta do princípio da dignidade da pessoa humana nas relações de direito privado reflete e superação dessa dicotomia e estabelece um verdadeiro sistema de proteção da pessoa humana no ordenamento jurídico brasileiro.</p>

Nova dinâmica da reintegração judicial no emprego : mito, realidade e utopia na concretização do direito de acesso à justiça à luz da efetividade do modelo processual brasileiro .

O modelo brasileiro de proteção à relação de emprego, em consonância com o princípio da continuidade, é formado por institutos que se destinam a proporcionar o prosseguimento do contrato empregatício, a ponto de estipular a nulidade do ato de despojamento praticado em arripio às suas medidas de restrição ao exercício do direito de despedir. Ocorre que o controle abstrato proporcionado pelo modelo normativo não é capaz de, isoladamente, assegurar a concretização do direito do hipossuficiente. Caso as normas sejam descumpridas, o empregado terá que utilizar o modelo processual para materializar o direito à reintegração. Como este consiste em restituere in integrum, restituindo integralmente o estado anterior ao ato invalidado, a tutela jurisdicional reintegratória consiste em uma tutela reparatória específica envolvendo três prestações a ser cumpridas pelo empregador: pagar uma indenização reparatória; promover o retorno físico do empregado e respectivo registro documental. À luz do atual modelo processual, existem três vias processuais aptas a proporcionar a concretização da tutela reintegratória: 1) efetivação de provimento de antecipação de tutela; 2) efetivação imediata dos capítulos mandamentais da sentença reintegratória; e, em casos excepcionais, 3) execução forçada de título executivo. Para proporcionar a inserção do obreiro dentro de uma realidade jurídica na qual o seu direito à reintegração não será limitado à previsão abstrata do legislador, será essencial assegurar a plena efetividade do modelo processual enquanto instrumento assecuratório de real acesso à justiça.

Em busca da satisfação dos interesses da vítima penal : uma tentativa de demonstrar os reais interesses das vítimas de crimes, desde a participação na persecução criminal à reparação do dano, e de se propor alternativas para a melhor satisfação desta.

O presente trabalho consiste numa dissertação acerca da participação da vítima na justiça criminal, resgatando o histórico do seu desenvolvimento desde os primórdios até o momento atual. Foi desenvolvida uma pesquisa de campo, baseada em entrevistas realizadas com vítimas de crimes, por meio das quais nos propusemos a constatar o grau de satisfação da vítima com a justiça criminal, dando ênfase aos questionamentos quanto ao seu interesse em participar da persecução criminal e sua satisfação acerca da pena a ser possivelmente aplicada pelo Estado. Alterações recentes na legislação penal, tais como medidas de conciliação e a aplicação de outras penas que não a privativa de liberdade, são também apresentadas, demonstrando sua contribuição para a melhor satisfação das vítimas. Finalmente, sugerimos outras medidas que poderão contribuir para a satisfação dos interesses da vítima penal, ao tempo que concluímos pela necessidade de sua participação na justiça criminal, enquanto sujeito de direitos e parte do drama criminal.

Duplo grau obrigatório do procedimento do mandado de segurança contra ato de particulares no exercício de atribuição pública.

O trabalho ora apresentado tem como título O DUPLO GRAU OBRIGATÓRIO NO PROCEDIMENTO DO MANDADO DE SEGURANÇA CONTRA ATO DE PARTICULARES NO EXERCÍCIO DE ATRIBUIÇÃO PÚBLICA. Inicia-se com uma breve explanação sobre interpretação das normas jurídicas, suas espécies e seus postulados. Segue com a análise do histórico do surgimento do mandado de segurança. Adiante, verifica-se a legitimidade passiva da ação de mandado de segurança. No capítulo seguinte, estuda-se o instituto do duplo grau obrigatório e, no capítulo final, analisa-se a interpretação possível do duplo grau obrigatório no procedimento de mandado de segurança quando o ato é praticado por particulares no exercício de atribuição pública.

O princípio da isonomia no contexto do Direito Tributário.

Neste trabalho, o princípio da isonomia é contextualizado como uma idéia pertinente ao direito natural, assimilado pelas declarações de direitos do homem, além de positivado nas constituições hodiernas, a exemplo da Constituição da República Federativa do Brasil, de 5 de outubro de 1988. Aborda-se a função ordenadora dos princípios, como que numa introdução à compreensão do princípio da isonomia como princípio fundante da vigente ordem constitucional pátria, assim como se evidencia a relação entre o princípio constitucional da isonomia e as normas infraconstitucionais. Considerando a positivação dos direitos fundamentais, a dicotomia jusnaturalismo versus juspositivismo ficou operacionalmente enfraquecida, uma vez que a consagração dos direitos fundamentais mitigou a tensão entre o real e o ideal, que sempre inspirou as concepções jusnaturalistas. Os tipos ideais ? jusnaturalismo e juspositivismo ? são tratados como generalizações necessárias ao aparato cognitivo do ser humano, ou seja, nem-se unidades reais em conceitos ideais aproximativos com o propósito de ultrapassar o abismo cognoscitivo do direito. É levada a efeito a discussão sobre a concretização da isonomia por meio dos subprincípios vinculados à capacidade contributiva, assim como a falácia da atração de investimentos privados por meio da concessão de incentivos fiscais pelos Estados. A ?guerra fiscal? é apresentada como a mais flagrante violação da isonomia tributária, verdadeira apropriação privada de recursos públicos.

Parto anônimo e princípio da afetividade

No Brasil, o abandono de recém-nascidos e sua exposição a condições subumanas têm-se tornado uma realidade. Por razões diversas, os genitores rejeitam seus filhos desde a gestação, o que desemboca na falta de desenvolvimento adequado aos bebês, abandono clandestino e possível condenação desses seres indefesos à morte. Embora responsáveis parentais pela prole que colocam no mundo, não se pode cobrar afeto dos genitores, porque afeto não se exige, sente-se. Num contexto de consagração da afetividade como guia das relações familiares, a pesquisa dedica-se à análise da viabilidade de implementação do parto anônimo como política pública voltada à defesa da vida e à valorização da dignidade dos bebês abandonados. Conhecido, antigamente, como roda dos expostos, o instituto volta à cena brasileira em decorrência da oferta legislativa, no Congresso Nacional, de projetos de lei que intentam a sua regulamentação, os quais são analisados no presente estudo. A dissertação realiza uma análise histórica do instituto, assim como aprecia a condição da mulher-mãe e o significado do abandono ao longo dos anos. O trabalho examina também o arcabouço principiológico que fundamenta o direito de família atual, com especial atenção ao princípio da dignidade humana. Investiga, ainda, o parto anônimo face ao direito comparado, enfrentando, no capítulo final, os aspectos favoráveis e contrários ao instituto, com o uso do método crítico nos estudos teóricos. Por fim, a conclusão revela o posicionamento da autora sobre os questionamentos levantados, notadamente no que concerne à possível implementação do parto anônimo no país.

Princípio da insignificância e tutela penal dos bens jurídicos: sua aplicação nos crimes próprios contra a administração pública

A presente dissertação de mestrado tem por objeto o estudo do princípio da insignificância no Direito Penal e a problemática de sua aplicação nas condutas em tese caracterizadas como crime, tendo a Administração Pública nos seus aspectos patrimonial e moral - como bem jurídico tutelado e seu funcionário como agente. O princípio da insignificância tem sido amplamente utilizado pelos julgadores contemporaneamente, seguindo as máximas minima non curat praetor, ou minimis non curat praetor ou, ainda, minimis praetor non curat, constituindo uma política criminal garantista, o que leva, em regra, ao afastamento da tipicidade penal para parte majoritária da doutrina e jurisprudência pátria e estrangeira. Contudo, a posição majoritária não apresenta o mesmo entendimento quando a conduta criminosa tem a Administração Pública como sujeito passivo e é praticada por seus funcionários ou pessoas a eles equiparadas.

Prescrição virtual: análise de sua aplicabilidade à luz dos princípios e garantias penais

O trabalho analisa a prescrição virtual, nova disciplina dada pelos operadores jurídicos brasileiros à ancestral causa de extinção de punibilidade. Após debater a natureza jurídica da prescrição e desenvolver um resumo histórico do surgimento da prescrição em sua modalidade clássica e de sua evolução legislativa, a dissertação avança sobre as teorias que fundamentam a prescrição na intenção de pesquisar eventual lastro conferida por elas à inovação pátria sem, contudo, furtar-se ao questionamento sobre a seriedade de sua

fundamentação teórica. Antes de adentrar no tema, o estudo faz ainda necessária digressão acerca da prescrição retroativa, pressuposto do objeto da pesquisa. O objetivo do estudo é avaliar a aplicabilidade da prescrição virtual bem como sua resistência diante dos argumentos que questionam a violação a princípios constitucionais e legais e, principalmente, a possibilidade de realização de prognóstico nos moldes idealizados pelos seus entusiastas. A dissertação procura realizar ponderação entre os benefícios decorrentes da adoção da nova técnica prescricional na solução dos problemas enfrentados pelo Poder Judiciário e os meios propostos pela doutrina para que o desiderato de desobstrução dos tribunais seja alcançado. Questiona-se até que ponto a celeridade alcançada representa avanço ou comprometimento das garantias conferidas aos cidadãos e da efetividade da tutela penal.

Exploração e produção de petróleo no Brasil: enfoques de viabilidade do modelo jurídico-regulador instituído para o Pré-Sal

A partir da nova roupagem legal atribuída pelas leis federais nº 12.351/10, 12.304/10 e 12.276/10 à exploração da atividade petrolífera no país, esboçou-se um aparato normativo propício à ampliação dos ganhos econômicos e materiais em favor do Estado durante a exploração das reservas de petróleo, bem como à maior segurança do mercado brasileiro em tempos de crise internacional. O objeto do presente trabalho perpassa justamente a análise do novo marco regulador, através do contrato de partilha, que substituirá, nas áreas do Pré-Sal ou em outras áreas estratégicas, o contrato de concessão, estabelecido na lei 9.478/97, da criação da Pré-Sal Petróleo S.A. e da redefinição de algumas atribuições dos entes públicos envolvidos diretamente no desenvolvimento da atividade, com especial destaque para a Agência Nacional de Petróleo, Gás Natural e Biocombustíveis ANP. Busca-se introduzir, de maneira panorâmica, as principais características do novo marco regulador e apresentar argumentos de natureza jurídica, econômica e social para compreender a viabilidade desse arcabouço jurídico recém criado. Para tanto utilizaremos a doutrina brasileira e estrangeira mais abalizadas, como Alexandre Santos Aragão, Marilda de Sá Ribeiro, Carlos Ari Sundfeld, e outros, com intuito de contribuir com os debates sobre o tema e auxiliar na compreensão das razões para a edição de um novo modelo legal para regular a atividade de exploração e produção de petróleo no país a partir do advento do Pré-Sal

Responsabilidade civil do cirurgião plástico

As agências reguladoras, entes autárquicos com autonomia estendida, passaram a fazer parte de nosso dia-dia de forma crescente, tamanha a proliferação que têm experimentado, ocupando espaço nos mais variados setores da Administração Pública brasileira nos últimos seis anos. Tal circunstância traz uma questão de relevo, que diz respeito à independência desses entes reguladores, ao grau de autonomia necessário ao desempenho de suas atribuições. O presente estudo enfoca aspectos ligados de forma mediata ao instituto, como o fenômeno da globalização e a reforma do Estado, tendo por alvo o núcleo central do problema, que reside, de um lado, no fato de os modelos de regulação realizados no estrangeiro - com destaque ao norte-americano, principal paradigma das agências reguladoras nacionais - pressuporem o exercício da função reguladora com independência normativa e, de outro, nas limitações impostas pelo atual ordenamento jurídico brasileiro à normalização pela via administrativa. Para equacionar o problema, procura-se realçar os limites à competência normativa do Chefe Poder do Executivo, confrontada com a atribuída aos entes reguladores. Justifica-se em face de a atividade de produção normativa de cunho administrativo não ser realizada de forma exclusiva pelas agências, como no estrangeiro, havendo um exercício concorrente de tal competência entre as agências reguladoras e o Presidente. Conclui-se o trabalho descrevendo o perfil que se desenha para as agências reguladoras brasileiras, em razão das peculiaridades impostas pelo ordenamento jurídico nacional

O desvio da função social da empresa como pressuposto para descon sideração da personalidade jurídica

A partir de uma perspectiva funcional do Direito este trabalho investiga como a função social da empresa é o fundamento da regra de limitação de responsabilidade dos sócios e, ao mesmo tempo, da descon sideração da personalidade jurídica. Para isso, apresentamos a mudança de paradigma da ciência do direito, partindo do estruturalismo preconizado por Hans Kelsen em sua Teoria Pura do Direito para o funcionalismo descrito por Norberto Bobbio. A partir da concepção de que ao direito cabe uma função promocional na sociedade, analisamos os fundamentos teóricos que justificam a aplicação dos princípios, normas e valores constitucionais às relações de direito privado. Analisamos então a empresa utilizando a teoria dos perfis da empresa, do jurista italiano Alberto Asquini, e a teoria jurídica da empresa, de Waldírio Bulgarelli. A partir dessas duas teorias conseguimos identificar aquilo que caracteriza a empresa sob a perspectiva do direito (a tríade estabelecimento, empresário e atividade), o que nos permite utilizar os conceitos de função social da propriedade e de função social do contrato para identificar os contornos do que seria a função social da empresa. Para isso, buscamos na história como surgiu o conceito de função social e analisamos os princípios e normas aplicáveis contidos na Constituição Federal e na legislação infraconstitucional. Definida a função social da empresa, passamos a analisar a descon sideração da personalidade jurídica a partir das principais teorias a seu respeito no Brasil e dos textos legais em que se baseia sua aplicação. Finalmente, propomos uma teoria funcional para a descon sideração da personalidade jurídica, tendo como premissa a compreensão de que as normas de limitação de responsabilidade dos sócios são normas de incentivo para a atividade empresarial realizar a função social da empresa e que a descon sideração da personalidade jurídica deve ser aplicada tendo como objetivo garantir o cumprimento daquela função, ou seja, como perda episódica da sanção positiva da limitação de responsabilidade dos sócios em decorrência do descumprimento da função social da empresa.

Nihilismo e direito: Ordem e hermenêutica como filosofia jurídica ativa

O presente trabalho objetiva relacionar o fenômeno do nihilismo e seus reflexos na teorização do direito. O nihilismo é analisado na específica acepção que se lhe dá o filósofo alemão Friedrich Wilhelm Nietzsche. Já que a obra de Nietzsche se apresenta em caráter fragmentário, foi escolhida a interpretação que Gilles Deleuze faz do nihilismo no âmbito desta, uma vez que o filósofo francês faz uso de conceitos precisos, explanando, inclusive, uma tipologia peculiar do fenômeno. Os tipos de nihilismo, correlatos a determinadas épocas históricas, são relacionados com as teorias que refletiram sobre o direito para que se possam perceber os reflexos daqueles no pensamento jurídico. Toda essa reflexão leva a um questionamento fundamental sobre a contemporaneidade, a sua relação com o saber jurídico e sua capacidade de superar o nihilismo. Segundo Deleuze, a superação do nihilismo no pensamento se daria através de uma filosofia ativa, cujo conceito procura-se compreender. Um dos efeitos no saber jurídico da vigência de um interpretativismo geral na sociedade contemporânea Ocidental é a tentativa de recepção dos postulados da hermenêutica filosófica de Hans-Georg Gadamer para a compreensão do processo decisório de aplicação das normas jurídicas, realizada pela corrente de pensamento intitulada jurisprudência hermenêutica. Tenta-se, a partir daí, demonstrar que a jurisprudência hermenêutica ainda guarda parentescos com o modo teológico de pensamento, o que impossibilitaria a saída do ciclo vicioso do nihilismo. Em contraste com essa corrente é apresentada a filosofia jurídica de Nelson Saldanha, a qual concebe o direito como ordem e hermenêutica, e que, pelas características a ela atinentes, revela-se uma filosofia ativa do direito, por isso mesmo capaz de superar o nihilismo no pensamento jurídico.

A nova lei da biossegurança e a utilização de células-tronco embrionárias para fins terapêuticos: a questão das sobras embrionárias e da clonagem terapêutica em confronto

O trabalho tem por objeto a análise da diferença no tratamento ético e jurídico da questão das células-tronco embrionárias quando obtidas, de um lado, dos embriões excedentes da técnica de reprodução humana assistida e, de outro lado, de embriões produzidos pelo processo de clonagem terapêutica. A problemática é analisada numa perspectiva dos direitos fundamentais, na interface da Bioética e do Biodireito com o Direito Constitucional. A nova Lei da Biossegurança regulamentou a questão, permitindo a utilização, para fins de pesquisa e terapia, de embriões resultantes da reprodução assistida e não utilizados no procedimento; todavia, vetou a possibilidade de obtenção de células-tronco

pela clonagem terapêutica. Busca-se averiguar, assim, as semelhanças e diferenças dessas situações jurídicas, abordando a possibilidade de nova legislação vir a igualar tais procedimentos ora tratados de maneira distinta. A decisão do STF, em sede da ADI 3510, é também analisada, sob a ótica de que as considerações ali lançadas têm direta repercussão no tratamento da clonagem terapêutica. O exame da problemática perpassa a questão do tratamento do embrião in vitro, cuja destruição constitui traço comum entre os dois processos de obtenção de células-tronco embrionárias em confronto, levantando o debate acerca do início da proteção jurídica da vida. O trabalho propõe que também a clonagem terapêutica utilizada de forma ponderada e com finalidade estritamente terapêutica, assim como a utilização das sobras embrionárias, não representa uma violação ao direito à vida e à dignidade da pessoa humana.

A proteção jurídica da honra e o direito penal mínimo

Nosso estudo analisa se há compatibilidade dos princípios da proteção de bens jurídicos e da intervenção mínima com os crimes contra a honra. Apresenta a evolução histórica daqueles princípios, suas teorias modernas, suas funções, materialização e compreensão como forma de limitar, nunca de aumentar, o direito de punir do Estado. Estuda também a honra, expondo a alteração do seu significado no tempo, além de aspectos filosóficos, éticos e dogmáticos que esclarecem a sua situação atual, externando a sua colisão com a liberdade de pensamento e destacando que a última é essencial para democracia. Efetua uma apreciação da intervenção mínima, da sua natureza subsidiária, do seu caráter fragmentário e do princípio da eficácia, demonstrando que a proteção penal da honra os viola, não havendo imposição constitucional de sua criminalização. Demonstra também que a proteção penal da honra, além de incapaz para a punição efetiva das suas ofensas, sobrecarrega os órgãos públicos e majora os seus gastos. De qualquer modo, conclui que o Judiciário não pode decidir pela inconstitucionalidade da criminalização no julgamento das demandas, servindo o estudo, principalmente, como forma de trazer elementos para o debate do tema, especialmente entre os legisladores, na esperança de incentivar a abolição da criminalização.

A experiência das audiências públicas no Supremo Tribunal Federal como instrumento de ampliação do círculo de intérpretes da Constituição

Nesta dissertação, analisamos o instituto processual da audiência pública, introduzido no direito brasileiro pelas Leis nºs 9.868/99 e 9.882/99, que disciplinam os processos de controle concentrado de constitucionalidade. Tal análise é feita a partir da experiência do Supremo Tribunal Federal na utilização deste instrumento. Contamos, até a entrega da dissertação, com 5 (cinco) audiências públicas realizadas, sendo que em 3 (três) delas foram proferidos julgamentos pelo Tribunal. Como pretendemos apreender se os argumentos desenvolvidos pelos participantes da audiência foram incorporados aos votos dos Ministros, fizemos esse corte metodológico e analisamos apenas as audiências realizadas nos autos da Ação Direta de Inconstitucionalidade (ADI) 3510 e da Arguição de Descumprimento de Preceito Fundamental (ADPF) 101, bem como a audiência pública da saúde. Trata-se de tema relacionado à jurisdição constitucional, o que torna imprescindível a análise prévia dos conflitos entre esta e a democracia e o princípio da separação dos poderes. A teoria da sociedade aberta dos intérpretes da Constituição, de Peter Häberle, constitui o marco teórico do nosso trabalho, diante da ampliação dos instrumentos de informação dos juízes constitucionais proporcionada pelo mecanismo das audiências públicas. A utilização das audiências públicas pelo Supremo Tribunal Federal coincide com o movimento de protagonismo e ativismo judicial do referido Tribunal, que nos últimos anos vem julgando casos envolvendo discussões acerca do conteúdo normativo dos direitos fundamentais e da definição de políticas públicas, que antes eram confiados aos poderes ditos políticos. Nos casos específicos das audiências públicas analisadas, ressaltamos que discussões acerca dos direitos à dignidade da pessoa humana, à vida e à saúde estão sempre presentes. Destacamos, ao final, a função das audiências públicas, consistente em subsidiar os Ministros de informações necessárias ao julgamento e propiciar a construção de alternativas de interpretação constitucional, bem como conferir legitimidade democrática ao Supremo Tribunal Federal.

Impactos da virtualização da sociedade no mundo jurídico: modificações no conceito de sujeito de direito

A Internet extrapolou sua destinação original, transformando-se de um meio de comunicação ubíqua para uma forma de convivência sociocultural. Nesse contexto, identificamos a ocorrência de um fenômeno, ao qual chamamos de virtualização. Como reflexo da globalização, o fenômeno da virtualização caracteriza-se por ser um processo gradual trifásico. Entendemos a Sociedade Virtual como todo e qualquer sistema digital conectado que possibilite a interação, dotada de sentido, entre duas ou mais personalidades, dentro de um único contexto cultural, favorecendo, assim, o surgimento de um fenômeno sociocultural. A Sociedade Virtual, composta por sujeitos que passaram pelo processo da virtualização, se organiza através de um sistema de regras e normas próprio que pode ser considerado um ordenamento jurídico autônomo. Mas será que os sujeitos que participam da Sociedade Virtual podem ser considerados sujeitos de direito? Com base em uma concepção contemporânea de ciência, que inclui o homem como observador e a reflexão como método, estudamos o fenômeno da virtualização da sociedade com o foco nos seus elementos subjetivos, buscando encontrar respostas, ou ao menos suscitar dúvidas, através de análises não reducionistas do sistema de conceitos jurídicos fundamentais e pautadas em um critério de validação compartilhado pela comunidade acadêmico-jurídica. Percebemos que a reprodução crítica do conceito lógico-jurídico de sujeito de direito pela doutrina civilística brasileira ignorou as recentes transformações sociais pelas quais passou a humanidade, o que trouxe até aos dias atuais consequências desastrosas, como a confusão entre os conceitos jurídicos fundamentais do sujeito de direito, da pessoa e da capacidade. Porém, nada impede que, a exemplo de outros entes desprovidos de personalidade, os avatares (sujeitos virtuais) sejam reconhecidos pelo ordenamento jurídico nacional como sujeitos de direito. Caso isso ocorra, a teoria da dupla imputação de Hans Kelsen pode ser adaptada para atribuir aos avatares a titularidade dos direitos e deveres reconhecidos pelo ordenamento jurídico.

O princípio da legalidade e os limites da produção normativa das agências reguladoras brasileiras

A ascensão de um Estado regulador, que teve como objetivo o repasse da prestação de serviços públicos e o exercício de atividades econômicas à iniciativa privada, sem eximir-se, contudo, de promover a organização e a fiscalização do desempenho dessas funções, fez surgir no sistema brasileiro as agências reguladoras, às quais foram repassadas as funções de normatizar, disciplinar e fiscalizar a prestação daquelas atividades administrativas pelos particulares. Questiona-se qual o fundamento jurídico-constitucional para o exercício dessa função normativa, bem como qual o real limite dessa atuação, quando se depara na ordem jurídica de um lado com os princípios da separação dos poderes e da legalidade e do outro com uma ampla margem de atuação transferida àquelas autarquias pelas leis instituidoras. Para pesquisar o tema, tomou-se como base a análise doutrinária, especialmente, os discursos sobre a deslegalização e a mutação do princípio da legalidade, como supostas formas de fundamentação e limites à atuação normativa das agências reguladoras. As observações resultantes seriam adequadas a uma modificação constitucional para preservação da reserva legal, bem como a imposição de uma padronização mínima de atuação, estritamente para implementação de normas técnicas. Ademais, identificou-se não ser possível que uma lei dita deslegalizadora permita a revogação de outras leis e o disciplinamento das matérias através de atos normativos, por ofensa às regras e os princípios previstos no Estado democrático de direito. Por fim, corrobora-se com o fundamento da atuação normativa no princípio da eficiência, todavia, rebatase a limitação da função normativa tão somente com base nesse princípio, defendendo uma delimitação pautada em lei, de modo que as agências possam exercer dito papel com aspecto meramente complementar.

A mídia e a construção social da criminalidade em Pernambuco: um estudo sobre o programa televisivo Bronca Pesada

Este trabalho tem por objeto o estudo dos relatos veiculados pelo programa de televisão Bronca Pesada, transmitido pela TV Jornal, em relação à criminalidade em Pernambuco. O objetivo é analisar as notícias veiculadas pelo programa como parte da reação social ao delito, que auxilia na construção social do que é o crime, suas dinâmicas e quem são os criminosos. Parte-se do enfoque do etiquetamento,

abordagem que considera a reação social como elemento determinante para a configuração do delito. Analisam-se algumas características do processo produtivo de notícias, assim como a condição da mídia enquanto ator social no mundo contemporâneo, com o fim de compreender melhor os relatos produzidos pelo programa estudado. Utilizou-se no presente trabalho de uma abordagem qualitativa, com uma análise aprofundada de um pequeno número de casos. A partir dos dados coletados, observa-se que a utilização de um discurso popular, mais atrativo para o público, não é em si um problema, mas sim o uso de construções preconceituosas para alcançar esse objetivo. Aproveitando a necessidade que a população, especialmente a mais pobre, tem de segurança, um direito humano legítimo, o Bronca Pesada, das diversas maneiras que aborda a problemática do crime, não apenas constrói uma imagem de Pernambuco como um Estado violento, mas também estimula uma cultura da intolerância cada vez maior, do medo e da reação rigorosa e exagerada a ele.

A consciência da ilicitude: sua função na dogmática penal a partir da teoria finalista

A dissertação tematiza a consciência da ilicitude, sua localização na estrutura da culpabilidade e, por consequência, do crime, bem como as repercussões advindas de sua ausência. Trata-se de um dos temas mais complexos da teoria jurídica do crime. Inicialmente, tratou-se da evolução da culpabilidade no âmbito da dogmática penal, desde a sua feição psicológica até a normativa. Para esta, a aferição da culpabilidade do sujeito, entendida como reprovabilidade daquele que podia agir conforme o direito, mas não o fez, pressupõe a possibilidade de entender o caráter ilícito de seu agir. Em seguida, passou-se a analisar o conceito da consciência da ilicitude, seus pressupostos e objeto. Conhecer pressupõe apreender pelos sentidos e introjetar o conteúdo do conhecimento, o qual, no tema ora debatido, não se refere à norma jurídica como entidade formal tipo legal, mas à contrariedade do atuar em relação ao ordenamento jurídico. Ademais, não se trata de conhecimento efetivo da antijuridicidade, mas de potencial consciência da ilicitude. Ao lado desta, imputabilidade e exigibilidade de conduta diversa compõem a culpabilidade, notadamente normativa, cujos contornos foram traçados pelo finalismo, liberta dos elementos anímicos dolo e culpa, os quais passam à conduta e, por consequência, ao tipo. Fincado nesta estrutura, analisou-se a seguir o erro de proibição que não se confunde com o erro de direito ou a ignorância da lei. Quando inevitável dito erro, exclui a culpabilidade, mas quando evitável importa em atenuação da reprimenda. Por fim, foram analisadas as decisões do Superior Tribunal de Justiça sobre o tema da consciência da ilicitude, destacando as posições adotadas, suas incoerências e aspectos mais relevantes.

União estável e matrimônio entre pessoas de um mesmo sexo: a tutela do direito à diferença

O presente trabalho tem por objetivo estudar a tutela da união homoafetiva enquanto entidade familiar. Para tal fim, a influência do direito civil constitucional foi adotada como parâmetro de pesquisa, com vistas a examinar em que medida houve a inclusão de situações, antes à margem do direito, como ocorre com as uniões entre pessoas de um mesmo sexo. Assim, partindo-se da constatação de que todo ser humano tem a liberdade de escolher a forma familiar que seja adequada a sua necessidade pessoal, deve-se investigar o tratamento dispensado às uniões homoafetivas, abordando os argumentos jurídicos, utilizados pela doutrina e jurisprudência para que lhes fossem conferidas a proteção do direito de família. É identificada, portanto, a necessidade de mudança de paradigma para acomodar o direito aos fatos sociais. A justificativa desta pesquisa está na indispensabilidade de se verificar as respostas jurídicas, as quais vêm sendo apresentadas para as questões relevantes que surgem das uniões homoafetivas ante a ausência de lei sobre o assunto. É abordado o recente julgamento realizado pelo Supremo Tribunal Federal, o qual reconheceu a união homoafetiva como entidade familiar em sede de ação direta de inconstitucionalidade. Ao final, conclui-se que a decisão proferida pela Corte Suprema teve o mérito de compatibilizar o direito com a realidade presente no seio social, oferecendo proteção jurídica com fundamento nos princípios e valores constitucionais que formam a base do sistema jurídico brasileiro.

Crime de sonegação fiscal. A tutela penal do dever fundamental e solidário de pagar tributos não vinculados

O legislador nacional, na esteira da doutrina alemã, tipificou o crime de sonegação fiscal em conformidade com o art. 1º da Lei nº 8.137/1990, caracterizando-o como crime patrimonial, material e de dano, tendo por objeto de tutela o erário. Ao apreciar a conduta caracterizadora da sonegação fiscal constataram-se os inúmeros efeitos prejudiciais dela decorrentes, seja por reduzirem o patrimônio estatal, seja por perturbarem a atuação da administração tributária, seja por causarem desequilíbrios econômicos e sociais. Por acreditar que tais efeitos, bem como os meios pelos quais eles são produzidos - pelo descumprimento das normas jurídicas que estabelecem regras tributárias e econômicas - não são os verdadeiros bens ou valores merecedores da máxima proteção que é a penal, procurou-se demonstrar que o interesse constitucional legitimador da escolha de bens jurídicos recai sobre o dever fundamental e solidário de pagar tributos não vinculados. Buscou-se na Constituição Federal o fundamento de legitimidade da eleição do bem jurídico, por se acreditar que o desideratum do direito penal é a compatibilização entre os substratos normativo e substancial dos bens e valores merecedores de sua proteção. Partiu-se do pressuposto de um Estado social democrático de direito em que o desenho de distribuição da carga tributária é conformado à teoria do sacrifício, de forma solidária e adequado à capacidade tributária de cada um e que, ademais, os recursos assim obtidos são empregados na realização do bem comum. A conduta típica, de lege ferenda, é assim descrita: Fazer declaração falsa ou omitir declaração verdadeira sobre rendas, bens ou fatos, ou empregar qualquer outra fraude com o fim de eximir-se, total ou parcialmente, do pagamento de tributos não vinculados. Na forma como proposto, o delito possui as seguintes características: objeto de tutela: dever fundamental e solidário de pagar tributos não vinculados; objeto material da ação: tributos não vinculados; sujeito passivo: sociedade; sujeito ativo: crime comum; classificação quanto à tipicidade da conduta objetivamente considerada: crime formal; classificação quanto à tipicidade da conduta subjetivamente considerada: crime doloso; classificação quanto à antijuridicidade da conduta: delito de perigo abstrato. Enfim, os indivíduos, para viver em sociedade, devem pagar o preço da solidariedade, contribuindo financeiramente conforme suas capacidades econômicas para a realização do bem comum, de modo que aquele que não o faz pratica conduta merecedora de reprovação penal.

Da (i)legitimidade da tutela penal da ordem econômica: simbolismo, ineficiência e desnecessidade do direito penal econômico

O presente trabalho versa sobre a falta de legitimidade da intervenção penal na ordem econômica. Para tanto, identifica o aparecimento do direito penal econômico a partir das convenções de Paris e do Sherman Act, preocupados, pois, com marcas e patentes e com a tutela anti-truste. Além disso, mostra que a criação destes novos bens jurídicos, ligados à economia, estão indissociavelmente vinculados à noção de cultura. Nestes bens culturais, aparece, ainda, o método usado no direito penal. Os bens jurídicos tradicionais são marcas da civilização, são bens corpóreos, tangíveis. Os bens culturais são incorpóreos, intangíveis. Estes bens são derivação de novas necessidades, criadas pela sociedade em que novos riscos são incrementados. Para tratar destes novos bens, a tipicidade penal, tradicionalmente marcada pela exigência da taxatividade, teria de ceder espaço para a criação de tipos abertos, repletos de elementos normativos, de normas penais em branco, haja vista que a mutabilidade dos bens culturais é incompatível com os rigores da tipicidade. Para além disso, como nestes bens culturais existe uma preocupação muito grande em evitar lesões, há um uso freqüente de crimes de perigo abstrato, antecipando a tutela penal. Assim, as garantias do direito penal restam desfiguradas, tendo as garantias da dogmática penal cedido espaço para as exigências da vida prática. Demonstra-se, adiante, a falta de legitimidade da intervenção penal na ordem econômica, expondo-se, para isso, o conceito de legitimidade, diferenciando-o de legitimação e de legitimidade. Expõe-se, assim, que a intervenção penal somente é legítima quando é necessária, justa e adequada. Apreciaram-se, ademais, as críticas da criminologia em relação ao direito penal econômico, examinando todas as teorias da pena, a evidenciar que a política criminal não tem como justificar e nem como fundamentar o direito de punir para o direito penal econômico. Adiante, cuida-se de apresentar alternativa, pautada no direito sancionador, meio termo entre o direito penal e o direito administrativo, apresentando, pois, todos os princípios inerentes, diferenciando-o do direito penal e do direito administrativo. Apresentaram-se as características do direito sancionador, do direito da intervenção, do direito penal de duas velocidades e do direito das contra ordenações.

Critica-se a manifestação contrária ao direito sancionador, por meio do posicionamento dos chamados sistemas penais paralelos. Defende-se, pois, a aplicação de um direito punitivo enérgico, com diversas garantias (flexibilizando-se apenas a taxatividade e a jurisdicionalidade). Este direito, com tais garantias, apresenta sanções, que não podem conduzir a pena de prisão. Conclui-se a tese, demonstrando que a tutela penal deve reconhecer sua ilegitimidade para tutelar relações advindas da economia, destes bens culturais, cedendo espaço para o direito sancionador

Manifesto por um direito comprometido internacionalmente: um estudo crítico acerca da autonomia da vontade, da reformulação da atuação do Estado nas relações contratuais privadas e sua repercussão no plano internacional

A presente Tese tem como objetivo conscientizar a doutrina privatista brasileira, em especial a do Direito Civil, de que existe um necessário compromisso internacional na análise de determinados institutos do Direito Privado, cujo maior expoente é o relacionado ao princípio da vontade, conhecido pela expressão autonomia da vontade. Para atingir esse objetivo, a tese parte da premissa que o processo de constitucionalização do direito privado tem como consequência direta a ampliação da atuação do Estado nas relações contratuais privadas e resulta na manutenção do ideário nacionalista no Direito contratual, ignorando sua faceta internacional. Desta constatação, a tese adota uma metodologia historicista-internacionalista para oferecer uma análise crítica sobre esse fenômeno de constitucionalização bem como oferecer uma interpretação diferente dos principais institutos envolvidos na questão, na tentativa de substituir o ideário nacionalista dominante por um necessário comprometimento internacional, afastando a desconfiança do jurista para com os elementos extranacionais, cada vez mais presentes nas relações jurídicas particulares. Ao fim, a presente Tese deve ser considerada como um manifesto, um estudo jurídico consciente do compromisso internacional que determinados assuntos apresentam e exigem em sua análise contemporânea, cujo destaque maior é a seara dos negócios jurídicos e a autonomia da vontade

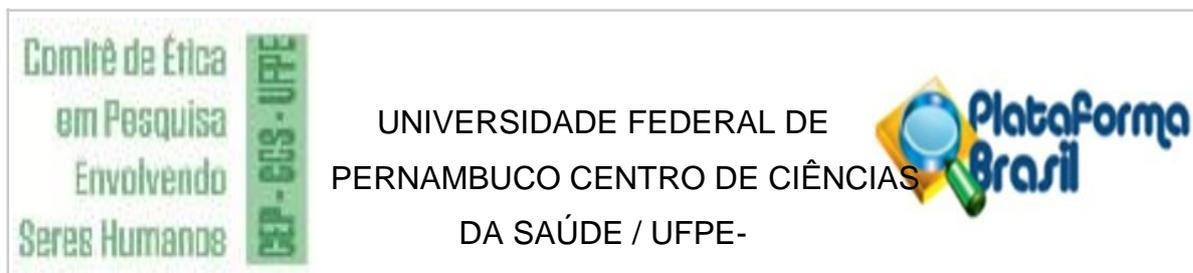
A legitimidade dos incentivos fiscais num contexto de guerra fiscal

Este trabalho tem por objeto de estudo a intervenção do Estado sobre o domínio econômico, por meio das normas tributárias indutoras. Para esse propósito, analisamos a busca pela superação da dependência nas relações centro-periferia, caracterizada pelo subdesenvolvimento, sob a perspectiva da Comissão Econômica para a América Latina e o Caribe (CEPAL). Estudamos o direito ao desenvolvimento, quer no âmbito internacional, notadamente as resoluções da Organização das Nações Unidas (ONU) sobre o desenvolvimento, quer na Constituição brasileira de 1988, haja vista que esse direito, juntamente com a concepção de desenvolvimento por meio da industrialização, constituem, na nossa opinião, o discurso legitimador das políticas estaduais de desenvolvimento, por meio da concessão de incentivos fiscais relativos ao Imposto sobre Operações relativas à Circulação de Mercadorias e Prestações de Serviço de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS). Analisamos a intervenção estatal no e sobre o domínio econômico, assim como as falhas de mercado e a correção dessas falhas pela intervenção estatal, para melhor entender a intervenção sobre o domínio econômico mediante o emprego de normas tributárias indutoras. Explicitamos as características básicas do ICMS, com a finalidade de respaldar as propostas de alteração que formulamos para este imposto, como forma de torná-lo menos regressivo, promover uma melhor distribuição dos recursos decorrentes de sua arrecadação e viabilizar a concessão de incentivos fiscais pelos Estados, numa situação em que o ônus dessas concessões seja suportado pelos Estados concedentes

A cognição no mandado de segurança sob o prisma dialógico de Mikhail Bakhtin

A teoria do enunciado dialógico desenvolvido por Mikhail Bakhtin se apresenta como modelo teórico apto a justificar a produção judicial do direito, já que esta se coloca em ambiente discursivo, em que o outro se apresenta como integrante do processo de comunicação, atuando de forma responsiva. No mandado de segurança, a decisão judicial é posta com base na cognição feita pelo magistrado. O modelo de cognição acolhido na doutrina pátria, apesar de explicar a mudança entre a liminar (cognição sumária) e a sentença (cognição exauriente), não explica a mudança quando esta ocorre entre a sentença (cognição exauriente) e o acórdão (cognição exauriente). Isso se dá porque o modelo clássico parte da premissa equivocada de utilizar no direito as categorias próprias das ciências naturais. Sob o prisma dialógico tal problema não ocorre, já que o ato decisório, seja ele qual for, por ser um enunciado dialógico, goza do atributo da conclusividade e da exauribilidade, que o torna completo, mas aberto a resposta na cadeia discursiva. Assim a decisão que julga a liminar é um enunciado completo, pleno de sentido, que se coloca para o outro, tal qual a sentença e o acórdão. O modelo de análise de Mikhail Bakhtin compreende a mudança não como um problema, mas como algo inerente ao processo enunciativo. Mesmo quando a decisão é confirmada, a confirmação é uma resposta elaborada pelo outro na cadeia do enunciado.

ANEXO A – COMPROVANTE DE ENVIO DE PROJETO AO COMITÊ DE ÉTICA – UFPE



COMPROVANTE DE ENVIO DO PROJETO

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Dos sintagmas nominais aos descritores documentais: estudo de caso na indexação de teses e dissertações da área de Direito.

Pesquisador: Gustavo Diniz

Versão: 1

CAAE: 43024515.3.0000.5208

Instituição Proponente: Centro de Artes e Comunicação

DADOS DO COMPROVANTE

Número do Comprovante: 020044/2015

Patrocinador Principal: Financiamento Próprio