

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO

FLÁVIO PACHECO DA SILVA

**ANÁLISE ENTRE LINGUAGEM NATURAL NO TWITTER E LINGUAGEM
DOCUMENTÁRIA EM TESAUROS: DO #NÃOMEREÇOSERESTUPRADA,
PASSANDO PELO TERMO JURÍDICO AO CONCEITO EDUCACIONAL**

Rio de Janeiro

2014

FLÁVIO PACHECO DA SILVA

**ANÁLISE ENTRE LINGUAGEM NATURAL NO TWITTER E LINGUAGEM
DOCUMENTÁRIA EM TESAUROS: DO #NÃO MEREÇOSERESTUPRADA,
PASSANDO PELO TERMO JURÍDICO AO CONCEITO EDUCACIONAL**

Trabalho de conclusão de curso apresentado à
Escola de Biblioteconomia, da Universidade
Federal do Estado do Rio de Janeiro, como
requisito à obtenção do grau de Bacharel em
Biblioteconomia.

Orientadora: Professora Mestre Brisa Pozzi de Sousa.

Rio de Janeiro

2014

S381a Silva, Flávio Pacheco da, 1981-
Análise entre linguagem natural no twitter e linguagem
documentária em tesouros: do #nãomereçoserestuprada, passando
pelo termo jurídico ao conceito educacional. / Flávio Pacheco da
Silva. — 2014.

74 f. : il. ; 30 cm.

Orientadora: Professora Mestre Brisa Pozzi de Sousa.
Trabalho de Conclusão de Curso (Bacharelado em
Biblioteconomia) — Escola de Biblioteconomia, Universidade
Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2014.
Inclui bibliografia.

1. Linguagem natural. 2. Linguagem documentária. 3.
Representação temática da informação. 4. Twitter. 5. Estupro
I. Brisa Pozzi de Sousa. II. Título.

CDD 22 – 025

FLÁVIO PACHECO DA SILVA

**ANÁLISE ENTRE LINGUAGEM NATURAL NO TWITTER E LINGUAGEM
DOCUMENTÁRIA EM TESAUROS: DO #NÃOMEREÇOSERESTUPRADA,
PASSANDO PELO TERMO JURÍDICO AO CONCEITO EDUCACIONAL**

Trabalho de Conclusão de Curso apresentado à
Escola de Biblioteconomia, da Universidade
Federal do Estado do Rio de Janeiro, como
requisito à obtenção do grau de Bacharel em
Biblioteconomia.

Aprovado em 03 de dezembro de 2014.

Banca Examinadora

Professora Mestre Brisa Pozzi de Sousa – Orientadora
Universidade Federal do Estado do Rio de Janeiro

Professor Doutor Marcos Luiz Cavalcanti de Miranda
Universidade Federal do Estado do Rio de Janeiro

Professora Doutora Miriam Gontijo de Moraes
Universidade Federal do Estado do Rio de Janeiro

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, fonte da vida e de toda inspiração, sem o qual nada seria possível.

Agradeço aos meus pais por todo amor, dedicação e esforço que fizeram para serem os melhores pais que eu poderia ter.

A minha esposa Thais por todo incentivo e apoio ao longo desses 14 anos juntos, por acreditar em mim mesmo quando eu não acreditava. Sem você não seria o homem que sou e não teria chegado até aqui. TE AMO!

Ao meu pequeno Davi, o maior presente que Deus me concedeu nessa vida, e a todos que de alguma forma colaboraram para que eu pudesse concluir esta etapa da minha vida.

A todos os meus professores que muito me ensinaram ao longo desses anos, e a minha orientadora professora Brisa Pozzi de Sousa pela dedicação, incentivo e paciência por me acompanhar e ajudar em cada passo deste trabalho.

*“Eu vim de longe para encontrar o meu caminho
Tinha um sorriso e o sorriso ainda valia
Achei difícil a viagem até aqui
Mas eu cheguei, mas eu cheguei”.*

(A viagem – Agnus Dei - D.R. Paulinas Comep)

RESUMO

Analisa a linguagem natural no Twitter e a linguagem documentária nos tesouros a partir da hashtag #NãoMereçoSerEstuprada. A pesquisa é descritiva e bibliográfica, de cunho qualitativo embasada na investigação teórica da representação do termo estupro do Tesouro Brasileiro da Educação (Brased), do Tesouro Jurídico do Superior Tribunal de Justiça (STJ) e comparando-se com a linguagem natural empregada na rede social Twitter enfoca vantagens e desvantagens no uso dessas linguagens. Aponta que em linguagem natural, qualquer palavra, ou até mesmo conjunto de palavras podem ser utilizadas como hashtag, e estas palavras-chave facilitam a recuperação de informações em sistemas de busca. No entanto observa-se que o controle do vocabulário é um elemento essencial para a organização da informação, pois é uma linguagem artificial produzida com o objetivo de sistematizar o vocabulário do usuário para um vocabulário padronizado, controlando a polissemia. Os resultados demonstram que a linguagem natural não padroniza a informação e este pode ser considerado fator dispersivo da mesma, ao contrário das linguagens documentárias que apresentam um vocabulário que se caracteriza pela maior precisão de seus termos.

Palavras-chave: Linguagem natural. Linguagem documentária. Twitter. Representação temática da informação. Estupro. Tesouro Brased. Tesouro STJ.

ABSTRACT

It Analyzes the natural language on Twitter and the indexing language on thesaurus from the hashtag #NãoMereçoSerEstuprada. The research is descriptive and bibliographic, qualitative based on the theoretical investigation of the representation of the term rape Tesouro Brasileiro da Educação (Brased), of Tesouro Jurídico do Superior Tribunal de Justiça (STJ) and comparing it with the natural language used on Twitter and it focuses advantages and disadvantages of these languages use. It points out that in natural language, any word, or even group of words that can be used as a hashtag, and these keywords and facilitate in the information retrieval in the search systems, however it is noted that the vocabulary control is an essential element for the organization of information, because it is an artificial language produced with the aim of systematizing the user vocabulary for a standardized vocabulary, controlling the polysemy. The results demonstrate that natural language does not standardize the information and this can be considered the same factor dispersive, unlike the documentary languages showing a vocabulary that is characterized by increased accuracy of its terms.

Keywords: Natural language. Documentary language. Twitter. Thematic representation of information. Rape. Brased thesaurus. STJ thesaurus.

LISTA DE IMAGENS

| | |
|--|----|
| Imagem 1 – #NãoMereçoSerEstuprada | 21 |
| Imagem 2 – Twitter da Jornalista Britânica Miriam Wells | 22 |
| Imagem 3 – Twitter em apoio a campanha da Presidente Dilma Rousseff (a) | 23 |
| Imagem 4 – Twitter em apoio a campanha da Presidente Dilma Rousseff (b) | 23 |
| Imagem 5 – Twitter em apoio a campanha da Presidente Dilma Rousseff (c) | 23 |
| Imagem 6 – Twitter em apoio a campanha da Presidente Dilma Rousseff (d) | 23 |
| Imagem 7 – Twitter em apoio a campanha da Presidente Dilma Rousseff (e) | 23 |
| Imagem 8 – Testando limite de palavras em “hashtag”s no Twitter | 34 |
| Imagem 9 – Operadores de buscas na plataforma do Twitter | 35 |
| Imagem 10 – Busca avançada na plataforma do Twitter | 36 |
| Imagem 11 – Testando limite de palavras em “hashtag” no Twitter..... | 37 |
| Imagem 12 – Internautas fazem piada com as focas ‘estupradoras’ de pinguins..... | 63 |

LISTA DE GRÁFICOS

- Gráfico 1 – Mulheres que usam roupas que mostram o corpo merecem ser atacadas (%).....19
- Gráfico 2 – Mulher que é agredida e continua com o parceiro gosta de apanhar (%).....19
- Gráfico 3 – Mulher que é agredida e continua com o parceiro gosta de apanhar (%).....24
- Gráfico 4 – Mulheres que usam roupas que mostram o corpo merecem ser atacadas (%).....24

LISTA DE QUADROS

Quadro 1 – Características da população entrevistada (3.810 pessoas)..... 18

Quadro 2 – Perspectivas de usuários na classificação social ou folksonomia social.....32

LISTA DE ABREVIATURAS E SIGLAS

IPEA – Instituto de Pesquisa Econômica Aplicada

ONU MULHERES – Entidade das Nações Unidas para a Igualdade de Gênero e o Empoderamento das Mulheres

CFEMEA – Centro Feminino de Estudos e Assessoria

LN – Linguagem Natural

LD – Linguagem Documentária

LDs – Linguagens Documentárias

LC – Linguagem Controlada

STJ – Superior Tribunal de Justiça

Brased – Tesouro Brasileiro da Educação

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

Cibec – Centro de Informação e Biblioteca em Educação

CPI – Comissão Parlamentar de Inquérito

SUMÁRIO

| | | |
|-----------|--|----|
| 1 | INTRODUÇÃO | 12 |
| 2 | METODOLOGIA | 15 |
| 3 | CONTEXTUAÇÃO DA ANÁLISE: A PESQUISA DO IPEA, TOLERÂNCIA SOCIAL A VIOLÊNCIA CONTRA AS MULHERES | 17 |
| 3.1 | A REPERCURSÃO: MULHER NENHUMA MERECE SER ESTUPRADA | 21 |
| 3.2 | A ERRATA DA PESQUISA..... | 24 |
| 4 | REFERENCIAL TEÓRICO | 26 |
| 4.1 | CONCEITUANDO A LINGUAGEM | 26 |
| 4.2 | A LINGUAGEM NA BIBLIOTECONOMIA..... | 27 |
| 4.3 | LINGUAGEM NATURAL | 29 |
| 4.3.1 | A Linguagem Natural no Contexto da Web 2.0..... | 29 |
| 4.3.2 | Folksonomia e as Tags..... | 31 |
| 4.3.3 | Rede Social Twitter | 33 |
| 4.3.3.1 | <i>As #hashtags e o Twitter</i> | 33 |
| 4.4 | LINGUAGEM DOCUMENTÁRIA | 38 |
| 4.4.1 | Os Tesouros..... | 40 |
| 4.4.1.2 | <i>O tesouro jurídico do Superior Tribunal de Justiça (STJ)</i> | 44 |
| 4.4.1.2.1 | <u>O termo Estupro no tesouro jurídico do STJ</u> | 47 |
| 4.4.1.3 | <i>O Thesaurus Brased – Thesaurus Brasileiro de Educação</i> | 49 |
| 4.4.1.3.1 | <u>O termo Estupro no Thesaurus Brased</u> | 50 |
| 5 | ANÁLISE DOS DADOS: O TERMO ESTUPRO NA LINGUAGEM DOCUMENTARIA x LINGUAGEM NATURAL | 59 |
| 5.1 | LINGUAGEM NATURAL: #NÃO MEREÇOSERESTUPRADA | 59 |
| 5.2 | LINGUAGEM DOCUMENTÁRIA: O TERMO ESTUPRO NOS TESAUROS DO STJ E BRASED | 60 |
| 6 | CONCLUSÃO | 64 |
| | REFERÊNCIAS | 66 |

1 INTRODUÇÃO

A Representação do Conhecimento, mais precisamente o tratamento temático da informação tem se voltado cada vez para a dimensão social da área e seus impactos para os usuários. Nesse contexto, o bibliotecário tem grandes desafios pela frente, principalmente porque é necessário levar em conta que tanto o produto quanto o processo de representação devem ser pensado para o usuário da informação. Neste trabalho serão apresentados os conceitos de linguagem natural e seus desdobramentos nas Redes Sociais através das marcações denominadas hashtag. Serão apresentados conceitos da linguagem documentária e será analisado a apresentação hierárquica do termo estupro em dois diferentes contextos documentários: o Tesauro Brasileiro de educação (Brased) e o tesauro do Superior Tribunal de Justiça (STJ).

A motivação partiu do resultado da pesquisa do IPEA sobre a tolerância social à violência contra as mulheres, onde, de acordo com os dados divulgados, foi constatado que a maior parte dos brasileiros acredita que as mulheres são responsáveis por sofrerem abusos sexuais. Tal resultado gerou grande repercussão na sociedade, sobretudo nas Redes Sociais com o surgimento da hashtag #NãoMereçoSerEstuprada.

É com esse cenário que se estabelece o tema deste trabalho, ou seja, a motivação para esta pesquisa partiu da instigação de compreender como o processo de representação temática da informação pode ser estruturado em diferentes ambientes informacionais.

A partir do controverso resultado inicial da pesquisa do IPEA, que posteriormente se mostrou errôneo, constatou-se uma rápida reação nas redes sociais, onde se iniciou uma campanha de repúdio ao resultado da pesquisa com a hashtag #NãoMereçoSerEstuprada. A campanha teve grande adesão e em pouco tempo já estava entre os Trending Topics¹ do Twitter. Toda essa mobilização em tão pouco tempo acabou sendo o fator estimulante desse trabalho, que não tem como objetivo a discussão do resultado da pesquisa ou qualquer

¹ Trending Topics são gerados automaticamente por um algoritmo que tenta identificar os tópicos que estão sendo mais comentados. [...] A lista Trends capta os melhores temas emergentes, e não apenas o que é mais popular. TWITTER. **Blog**. Disponível em: <<https://blog.Twitter.com/2010/trend-or-not-trend>>. Acesso em: 03 maio 2014.

Os Trending Topics livremente traduzidos “Tópicos em Tendência” e, popularmente abreviados TT, consistem basicamente dos termos ou frases mais citados ou repetidos dentro do Twitter num dado momento. Em outras palavras, é um indicador de popularidade. Sua seleção e classificação ocorre automaticamente através de algoritmos internos do sistema do Twitter, que encontram (a partir de uma série de regras pré-estabelecidas) as palavras mencionadas com maior frequência. TUDO SOBRE REDES SOCIAIS. **Trending topics**. Disponível em: <<http://www.tudosobreredesociais.com.br/2011/08/03/o-que-sao-trending-topics/>>. Acesso em: 03 maio 2014.

questão de gênero, mais sim, fazer uma abordagem sobre representação temática da informação de termos relativos ao estupro, tanto em linguagem natural, mais precisamente nas Redes Sociais, quanto na linguagem documentária, com uso de dois tesouros.

Nesse contexto, o **problema** de pesquisa consiste na análise da representação da informação em linguagem natural e com uso de instrumentos de controle de vocabulário. Partindo do pressuposto que as mídias sociais possuem importância ao destacar determinado assunto, instiga-nos compreender quais fatores podem aproximar ou distanciar os tipos de linguagem documentária e natural.

O **objetivo geral** baseia-se na investigação da representação do conceito estupro e seu resultado com uso de dois instrumentos de controle de vocabulário, o Tesouro Brasileiro de Educação (BRASED) e o Tesouro do Superior Tribunal de Justiça (STJ) e suas possíveis relações com a linguagem natural empregada no Twitter através das hashtags. Partindo do amplo uso que é feito das redes sociais na atualidade, essa pesquisa busca contribuir com discussões que envolvam a representação da informação nesses ambientes.

Conforme se anuncia no título deste trabalho, o objeto desta pesquisa é uma análise dos termos relativos ao estupro utilizados nas Redes Sociais e a hierarquização do referido termo nos vocabulários controlados, tomando como *corpus* dessa análise dois tesouros. Em consonância com as questões descritas, pretendemos demonstrar as relações que aproximam ou distanciam esses dois tipos de linguagem - linguagem natural (LN) e linguagem documentária (LD) - especificamente no que tange os termos relativos ao estupro na representação temática da informação.

Como **objetivos específicos** à pesquisa almeja:

- a) Discorrer sob a perspectiva da representação da informação, em linguagem natural, de como acontece o processo de tagueamento na rede social Twitter, e qual a interligação desse processo com a linguagem documentária;
- b) Apresentar na perspectiva da linguagem documentária como um mesmo termo pode se apresentar de formas hierárquicas, em contextos diferentes, em dois tipos de vocabulários controlados, neste caso, um da área jurídica e outro da educação;
- c) Analisar semelhanças e diferenças entre linguagem natural e linguagem documentaria.

O trabalho segue estruturado em seis seções, sendo a primeira a parte introdutória. Na segunda seção seguem os processos metodológicos, e na terceira é apresentada a contextualização da pesquisa do IPEA sobre a tolerância social à violência contra as mulheres. Em seguida, na quarta seção apresenta-se o referencial teórico, e, posteriormente, na quinta seção consta a análise dos dados coletados entre as linguagens documentárias através de dois tesouros selecionados, e a linguagem natural utilizada na rede social Twitter. Por fim, segue a conclusão.

2 METODOLOGIA

A pesquisa realizada neste trabalho é classificada como descritiva, pois visa descrever as características das linguagens natural e documentária e estabelecer relações entre ambas através de uma observação sistemática.

De acordo com Danton (2002, p. 10), na pesquisa descritiva o pesquisador “observa, registra e analisa os fenômenos, sem manipulá-los. Procura descobrir a frequência com que o fenômeno ocorre, sua natureza, suas características, sua relação com outros fenômenos”.

Ela também se caracteriza uma pesquisa bibliográfica e qualitativa. Bibliográfica porque é elaborada a partir da revisão material já publicado, constituído de livros, artigos de periódicos, teses, dissertações e material disponibilizado na internet.

Qualitativa porque é obtida através do contato direto do pesquisador com o fenômeno observado. Segundo Kauark, Manhães e Medeiros (2010, p. 24), “a pesquisa qualitativa lida com fenômenos, aquilo que se mostra, que se manifesta”, ou seja, esta pesquisa é tratada como qualitativa pela observação feita nas linguagens natural e documentária, observação esta que permitiu a interpretação dos fenômenos e a atribuição de significados dessas linguagens. Portanto,

[...] Na Pesquisa Qualitativa considera-se que há uma relação dinâmica entre o mundo real e o sujeito, isto é, um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito que não pode ser traduzido em números. [...] O ambiente natural é a fonte direta para coleta de dados e o pesquisador é o instrumento-chave. (KAUARK; MANHÃES; MEDEIROS, 2010, p. 24).

Nesta pesquisa realiza-se a investigação teórica da representação do termo estupro, embasando a análise em dois tesouros - o Tesouro Brasileiro da Educação (Brased) e o Tesouro Jurídico do STJ - e fundamenta-se comparação com a linguagem natural empregada na Rede Social Twitter através de hashtags.

A hashtag utilizada para análise foi a da repórter Nana Queiroz, divulgada no Twitter e representada por: #NãoMereçoSerEstuprada.

A constituição da etapa de revisão de literatura se deu através de referenciais teóricos da área de Ciência da Informação/Biblioteconomia e da Informática, basicamente constituído por livros, artigos, 2 trabalhos apresentados em eventos, 2 teses, 1 dissertação e 14 *sites*. Utilizou-se como fonte de pesquisa a base Scielo e a Brapci, sem, contudo definir um recorte

temporal para recuperação do material, ou seja, documentos que abordassem sobre tesouro, linguagem natural e linguagem documentária foram recuperados independentes do ano de publicação.

Os termos utilizados para recuperação dos artigos nas bases foram: linguagem natural, linguagem documentária, vocabulário controlado, Twitter, redes sociais, representação da informação, indexação, folksonomia e hashtags.

Quanto à questão da delimitação do termo estupro, o fator motivador da escolha foi a pesquisa divulgada no início do ano pelo Instituto de Pesquisa Econômica Aplicada (IPEA), na data de 27 de março de 2014, sobre a tolerância social à violência contra as mulheres (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a). A divulgação do resultado levantou acalorada discussão em território nacional, inclusive nas Redes Sociais onde houve rápida reação.

Há que se ressaltar que foram inseridas, neste trabalho, apenas o recorte de um termo dos referidos tesouros, o estupro, não ampliando a pesquisa para outros termos associados.

A pesquisa possibilitou um levantamento geral de aspectos pertencentes às Redes Sociais e foi possível visualizar como um mesmo termo pode apresentar formas diferente de representação, dependendo do contexto ao qual esta inserido.

A seção seguinte pautará a contextualização que motivou a escrita do trabalho.

3 CONTEXTUALIZAÇÃO DA ANÁLISE: A PESQUISA DO IPEA TOLERÂNCIA SOCIAL À VIOLÊNCIA CONTRA AS MULHERES

Pesquisa divulgada no início do ano pelo Instituto de Pesquisa Econômica Aplicada (IPEA), na data de 27 de março de 2014, sobre a tolerância social à violência contra as mulheres levantou acalorada discussão em território nacional, inclusive nas Redes Sociais onde houve rápida reação ao resultado. Segundo os dados, foi constatado que a maior parte dos brasileiros acredita serem as mulheres responsáveis por sofrerem abusos sexuais (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a).

Na Colômbia, no ano de 2009 ocorreu pesquisa em âmbito nacional com o intuito de investigar os hábitos, atitudes, percepções e práticas individuais, sociais e institucionais no que diz respeito à violência de gênero e esta foi a inspiração para a pesquisa realizada no Brasil (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a).

Sendo assim, um grupo de trabalho no Brasil foi formado em 2012 entre a Entidade das Nações Unidas para a Igualdade de Gênero e o Empoderamento das Mulheres (ONU Mulheres), o Centro Feminino de Estudos e Assessoria (CFEMEA) e o IPEA, com objetivo de adaptar o questionário utilizado na pesquisa colombiana a nossa realidade nacional. Aqui, a pesquisa ganhou um caráter diferenciado, tanto no marco conceitual que embasou a seleção das questões, quanto no formato do questionário e no desenho das amostras. No caso da pesquisa realizada pelo IPEA foi feito um levantamento de opiniões e percepções sobre algumas questões relacionadas ao modelo patriarcal² de família, racismo, sexismo³ e violência contra as mulheres (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a).

De acordo com os dados da pesquisa, um percentual de 91% em relação aos 3.810 entrevistados concordam total ou parcialmente com a afirmativa que homem que bate na esposa tem que ir pra cadeia. Essa resposta apresenta tendência da maioria (quase em sua totalidade) em concordar com a punição para violência doméstica contra a mulher. Esse dado transcende as fronteiras sociais e apresenta pouca variação entre regiões, sexo, raça, religião, idade, renda e educação, ou seja, quase não há tolerância entre os entrevistados para o homem

² O patriarcado remete à origem do termo família, oriundo do vocábulo latino *famulus*, que significa "escravo doméstico". Esse novo organismo social – a família – consolidou-se enquanto instituição na Roma Antiga. A família romana era centrada no homem, sendo as mulheres, no geral, meras coadjuvantes. O patriarca tinha sob seu poder a mulher, os filhos, os escravos e os vassalos, além do direito de vida e de morte sobre todos eles (ENGELS apud NARVAZ; KOLLER, 2006).

³ Teoria que defende a superioridade de um sexo, geralmente o sexo masculino, sobre o outro. (Dicionário Priberam da Língua Portuguesa, 2008-2013). Disponível em: <<http://www.priberam.pt/dlpo/sexismo>>. Acesso em: 01 maio 2014.

que age com violência contra sua esposa (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a). Abaixo é possível observar as características da população entrevistada.

Quadro1 - Características da população entrevistada (3.810 pessoas)

| | |
|--|------------|
| A) Residentes no Sul ou Sudeste (sse): | 56,7% |
| B) Residentes em áreas metropolitanas (metro): | 29,1% |
| C) Pessoas jovens, 16 a 29 anos (jovem): | 28,5% |
| D) Pessoas adultas, 30 a 59 anos: | 52,4% |
| E) Pessoas idosas, 60 ou mais anos (idoso): | 19,1% |
| F) Mulheres (fem): | 66,5% |
| G) Brancos (branco): | 38,7% |
| H) Católicos (cato): | 65,7% |
| I) Evangélicos (evan): | 24,7% |
| J) Demais religiões, ateus e sem religião: | 9,6% |
| K) Menos que o ensino fundamental: | 41,5% |
| L) Ensino fundamental (edufunda): | 22,3% |
| M) Ensino médio (edumedia): | 30,8% |
| N) Ensino superior (edusuper): | 5,4% |
| O) Renda domiciliar <i>per capita</i> média: | R\$ 531,26 |

Fonte: Instituto de Pesquisa Econômica Aplicada (2014a, p. 31)

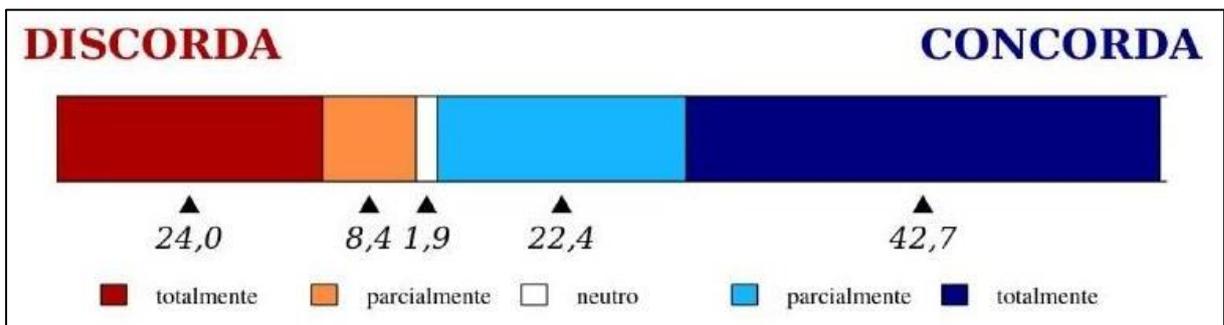
Cerca de 78% dos entrevistados concordaram totalmente com a prisão para maridos que batem em suas esposas e além disso, 89% discordaram da afirmação que o homem pode xingar e gritar com sua própria mulher (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a).

De acordo com a pesquisa, e com base nos resultados seria prematuro concluir pela baixa tolerância à violência contra a mulher na sociedade brasileira, pois os resultados, por outro lado, apontaram evidências contrárias:

Quase três quintos dos entrevistados, 58%, concordaram, total ou parcialmente, que 'se as mulheres soubessem se comportar haveria menos estupros'. E 63% concordaram, total ou parcialmente, que 'casos de violência dentro de casa devem ser discutidos somente entre os membros da família'. Também, 89% dos entrevistados tenderam a concordar que 'a roupa suja deve ser lavada em casa'; e 82% que 'em briga de marido e mulher não se mete a colher'. (INSTITUTO DE PESQUISA ECONÔMICA APLICADA 2014a, p. 3).

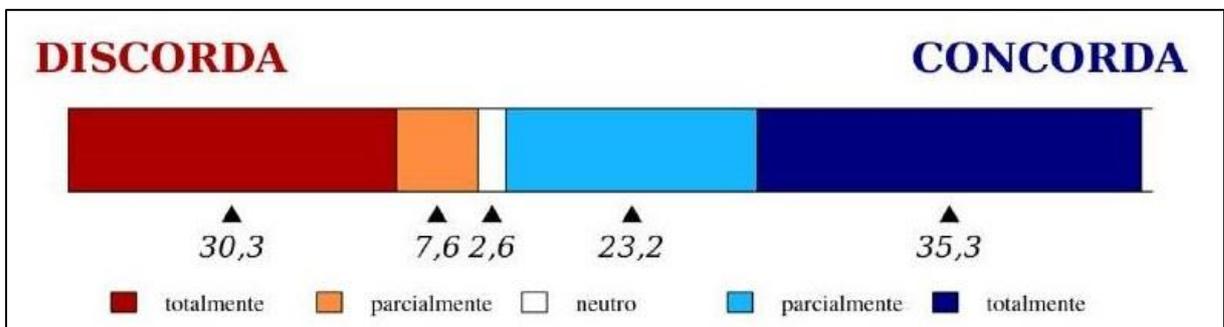
Constata-se a partir dos dados acima que para os entrevistados não é aceitável a violência contra as mulheres, entretanto sugerem um contraponto, pois diante das perguntas: as mulheres que usam roupas que mostram o corpo merecem ser atacadas e, se as mulheres soubessem como se comportar, haveria menos estupros, as respostas surpreendentemente se mostraram contrárias aos demais resultados obtidos, como é possível observar nas figuras abaixo:

Gráfico 1 – Mulheres que usam roupas que mostram o corpo merecem ser atacadas (%)



Fonte: Instituto de Pesquisa Econômica Aplicada (2014a, p. 23)

Gráfico 2 – Se as mulheres soubessem como se comportar, haveria menos estupros (%)



Fonte: Instituto de Pesquisa Econômica Aplicada (2014a, p. 23)

No entanto, segundo o relatório da pesquisa, essa contradição se desfaz pelo fato da população ainda considerar como modelo de família, o modelo patriarcal, pois, embora o homem ainda seja percebido como o chefe da família, seus direitos sobre a mulher não são mais irrestritos, e excluem as formas mais abertas e extremas de violência (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a). Sendo assim,

[...] Um homem deve tratar bem sua esposa, e, enquanto o fizer, rugas 'menores' devem ser resolvidas no espaço privado. A esposa, por sua vez,

deve 'se dar ao respeito', se comportar segundo o papel prescrito pelo modelo. Mas, se os conflitos se tornarem violentos, o casal deve se separar, a mulher não deve tolerar violência pelos filhos, e, se o marido bater, é caso para intervenção do público na esfera privada. (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a, p. 3-4).

O coordenador da pesquisa declarou em entrevista ao programa Fantástico da Rede Globo de televisão, que a principal conclusão diante os resultados é que a sociedade brasileira está impregnada por uma cultura machista, e:

a primeira coisa que temos que fazer, acredito, é trazer à tona esse problema que muitas vezes está escondido embaixo do tapete, está encerrado entre quatro paredes e falar para a mulher o seguinte: que ela não é culpada, ela é sempre a vítima. E por que isso é importante? Porque centenas de vítimas simplesmente não vão prestar queixa à polícia, porque elas vão achar que elas que na verdade fizeram alguma coisa, que facilitaram e vão ser mal vista na sociedade. (CERQUEIRA, 2013, on-line).

Segundo Vilhen e Zamora (2004) o estupro é justificado de diferentes formas em diferentes culturas. Com muita frequência, o argumento do "consentimento" das mulheres violadas é utilizado, e esse consentimento do ataque se dá porque foi pedido pela mulher ao usarem roupas curtas, coladas, perfume e maquiagem chamativos. Contudo, diante desse argumento, é ignorado, por exemplo, que também há violência contra as mulheres de hábito de freira ou de burca. Para as autoras, a ideia de que a "mulher na verdade queria" permite trivializar o estupro, o qual costuma ser reduzido ao privado, pois para muitos, a denúncia amplia a vergonha da vítima e da família devendo, portanto, ser evitada, pois, além do constrangimento, muitas vezes a vítima é julgada por sua conduta e até mesmo "condenada" a merecer um ataque.

Diante a percepção do resultado que considera o estupro como uma consequência da roupa utilizada, muitas manifestações ocorreram em diversos meios de comunicação, como jornais, *sites*, mídias sociais, entre outros.

Essa repercussão motivou o interesse em pesquisar a representação do termo estupro em Linguagem Natural, sobretudo nas Redes Sociais e em linguagem documentária como os Tesouros. Na próxima subseção será traçado um panorama de como se deu essa grande repercussão.

3.1 A REPERCURSÃO: MULHER NENHUMA MERECE SER ESTUPRADA

A partir do polêmico resultado e se sentindo indignada, a jornalista Nana Queiroz começou uma campanha de repúdio nas Redes Sociais com a hashtag⁴ (representada pelo símbolo #, de nome cerquilha) seguido do assunto ou palavra-chave #NãoMereçoSerEstuprada.

Declaro oficialmente abertas às postagens. Aqui vai minha foto, tirada em frente ao Congresso Nacional. Na pulseira e no colar, uma homenagem às minhas ancestrais indígenas, mulheres guerreiras oprimidas por nossa sociedade. No batom vermelho, às prostitutas, talvez as mulheres mais estupradas de nosso país. Elas também não merecem, Nem eu, Nem você. #NãoMereçoSerEstuprada #NinguémMerece (QUEIROZ, 2014, on-line).

Deste modo deu-se início a campanha organizada pela jornalista que publicou em uma Rede Social uma fotografia sem blusa e com os dizeres “Não mereço” no braço esquerdo levantado na altura da testa e, “Ser Estuprada” no braço direito cobrindo os seios.

Imagem 1 – #NãoMereçoSerEstuprada



Fonte: Bresser (2014)

A partir desse fato houve adesão gigantesca da sociedade como um todo, inclusive nas Redes Sociais e a hashtag #NãoMereçoSerEstuprada ganhou força, e em pouco tempo já

4 O símbolo de # (chamado cerquilha), já existia, no entanto passou a ser utilizado como marcador de hashtag, é utilizado para marcar palavras-chave ou tópicos em um Tweet. Foi criado fundamentalmente por usuários do Twitter com o intuito de categorizar mensagens. Esse símbolo é chamado de marcador e foi adicionado especificamente nos Tweets para marcá-los como se estivessem relacionados a um tópico. Assim, as pessoas podem seguir a conversa na pesquisa. (TWITTER, 2014).

estava entre os Trending Topics (TTs) do Twitter. Segundo Recuero (2009), o Twitter foi lançado em 2006 a partir de um projeto submetido por Jack Dorsey, Biz Stone e Evan Williams à empresa Odeo. Assim,

O Twitter nasceu como um serviço para celular. Ele foi criado para caber no limite de caracteres de uma mensagem de texto, e o Twitter ainda funciona em qualquer telefone compatível com SMS. O texto curto mantém o Twitter rápido e relevante, incentivando as pessoas a Tweetarem no momento e a se concentrarem na ideia essencial que estão tentando comunicar. [...] Dentro de um Tweet, você verá fotos e vídeos de pessoas que você conhece ou momentos de bastidores das maiores celebridades. Você pode incluir links para notícias, blogs, Websites e aplicativos. (DISCOVER TWITTER, 2014, on-line).

A campanha ganhou adesão de artistas e personalidades e foi divulgada em todos os meios de comunicação no Brasil e até mesmo no exterior, onde o assunto foi noticiado na TV britânica BBC, no jornal Washington Post, entre outros.

A jornalista Miriam Wells, freelance britânica na América Latina, que escreve para Sunday Times Foreign, The Sunday Times e InSight Crime, publicou em seu Twitter (@missmbc) apoio a campanha.

Imagem 2 – Twitter da Jornalista Britânica Miriam Wells



Fonte: Wells (2014)

A Presidente Dilma Rousseff também aderiu à campanha e manifestou solidariedade as mulheres através do seu Twitter (@dilmabr) e criou uma nova hashtag #Respeiteasmulheres, como é possível observar abaixo.

Imagem 3 – Twitter em apoio à campanha da Presidente Dilma Rousseff (a)



Fonte: Rousseff (2014a)

Imagem 4 – Twitter em apoio à campanha da Presidente Dilma Rousseff (b)



Fonte: Rousseff (2014b)

Imagem 5 – Twitter em apoio à campanha da Presidente Dilma Rousseff (c)



Fonte: Rousseff (2014c)

Imagem 6 – Twitter em apoio à campanha da Presidente Dilma Rousseff (d)



Fonte: Rousseff (2014d)

Imagem 7 – Twitter em apoio à campanha da Presidente Dilma Rousseff (e)



Fonte: Rousseff (2014e)

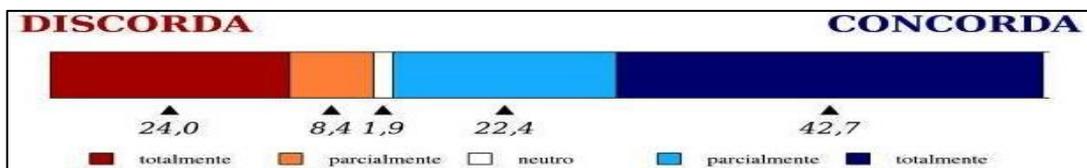
Após esta grande repercussão e mobilização em torno dos resultados da pesquisa, o IPEA divulgou uma errata dos dados alegando que uma troca entre dois gráficos gerou o erro, como veremos a seguir.

3.2 A ERRATA DA PESQUISA

No dia 4 de abril de 2014, uma semana após a divulgação do resultado da polêmica pesquisa sobre Tolerância Social à Violência Contra as Mulheres (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a) ser o estopim da campanha #NãoMereçoSerEstuprada nas Redes Sociais, o IPEA informou haver erro no resultado. “Vimos a público pedir desculpas e corrigir dois erros nos resultados de nossa pesquisa Tolerância social à violência contra as mulheres, divulgada em 27/03/2014”. (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014. Errata).

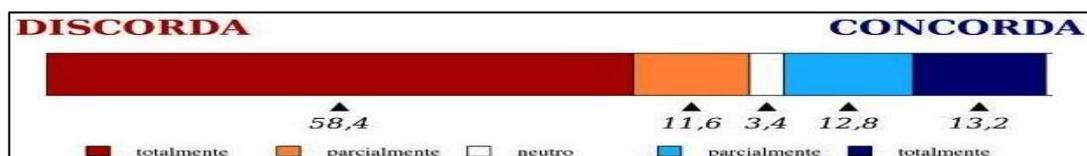
Segundo o IPEA o erro foi causado pela troca dos gráficos relativos aos percentuais das respostas às frases: “Mulher que é agredida e continua com o parceiro gosta de apanhar” e “Mulheres que usam roupas que mostram o corpo merecem ser atacadas”. Com a troca dos gráficos, ao invés dos 65% que concordavam com a frase “mulheres que usam roupas que mostram o corpo merecem ser atacadas”, o número correto é 26%, sendo 70% não concordam e 3,4% são neutros, como é possível observar abaixo (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014. Errata).

Gráfico 3 – Mulher que é agredida e continua com o parceiro gosta de apanhar (%)



Fonte: Instituto de Pesquisa Econômica Aplicada (2014b, p. 22)

Gráfico 4 – Mulheres que usam roupas que mostram o corpo merecem ser atacadas (%)



Fonte: Instituto de Pesquisa Econômica Aplicada (2014b, p. 23)

A correção da inversão dos números entre duas das 41 questões da pesquisa reduz a dimensão do problema anteriormente diagnosticado no item que mais despertou a atenção da opinião pública, no entanto, apesar da mudança de percentual ser muito grande, de 65% para 26%, os números ainda são bem altos. Apesar do grave erro, principalmente por se tratar de um instituto renomado e com alto grau de confiabilidade, não podemos amenizar o cenário apresentado pela análise em questão.

Alguns jornalistas afirmam que a pesquisa foi divulgada com erro para manobrar o povo brasileiro sobre a formação da Comissão Parlamentar de Inquérito (CPI) da Petrobras. Segundo Patu⁵ (2014) erros na apresentação de dados são infelicidades até para os melhores pesquisadores, no entanto há mais em jogo. De acordo com o jornalista, o caso realça ainda mais a guinada sofrida pelo IPEA, a partir do segundo governo do presidente Lula⁶, quando foi transferido do Ministério do Planejamento para o Palácio do Planalto e passou a tratar de temas bem diferentes dos que estão previstos em sua sigla, assumindo tom mais militante em suas pesquisas.

Porém, como o foco desse trabalho não tem por objetivo investigar o IPEA, os resultados da pesquisa apresentados aqui tiveram a função de despertar o interesse na investigação da Representação da Informação. De um lado temos a repercussão nas mídias sociais, como o Twitter, que faz uso da Linguagem Natural, e de outro, a Representação da Informação em ambientes informacionais, como nas Bibliotecas, que utilizam os Tesouros e outras LDs.

A representação temática do assunto de um assunto visa a recuperação eficiente da informação por parte do usuário, e sendo assim, na próxima seção trataremos de fato no que consiste este trabalho, apresentando tipos de linguagens existentes no campo da Biblioteconomia.

⁵ Gustavo Patu é jornalista da Folha de São Paulo.

⁶ LULA - Luiz Inácio Lula da Silva foi Presidente da República Federativa do Brasil entre os anos de 2003 e 2010. Disponível em: <<http://www.biblioteca.presidencia.gov.br/ex-presidentes/luiz-inacio-lula-da-silva/biografia-periodo-presidencial>>. Acesso em: 18 maio 2014.

4 REFERENCIAL TEÓRICO

O homem através da necessidade de se comunicar encontrou na linguagem seu principal mecanismo de comunicação com seus iguais e com o universo a sua volta (PETTER, 2005). De acordo com Dahlberg (1978) a linguagem constituiu-se na capacidade do homem de dar nome aos objetos que estavam ao seu redor e comunicar-se com os iguais, de modo que, permitiu-lhe relacionar-se com esses objetos e elaborar enunciados sobre eles.

Através das linguagens e seus mecanismos de comunicação, as experiências vividas são passadas através das gerações, de modo a garantir nosso modo de vida e a perpetuação da experiência humana. Dentro do contexto das linguagens documentárias não é diferente, no entanto, o mesmo processo acontece de forma organizada e hierarquizada, de modo a garantir a preservação do conhecimento técnico e científico (DAHLBERG, 1978).

Sendo assim, nas próximas seções será realizada a conceituação da linguagem e suas origens, assim como a aplicação das linguagens dentro do campo de estudo da Biblioteconomia. Serão apresentadas as definições de linguagem natural (LN), linguagem documentária (LD) e o desdobramento das mesmas.

4.1 CONCEITUANDO A LINGUAGEM

Para Moreira (2005) a linguagem pode ser entendida como a manifestação empregada pelos seres vivos em suas relações de convivência ou como o instrumento pelo qual eles se comunicam. É constituída por um conjunto de símbolos que se combinam de forma sistemática para o armazenamento e a troca de informações.

Nessa direção, Saussure (1969 apud PETTER, 2005), apresenta a língua como sendo uma parte essencial da linguagem como veremos a seguir:

[...] a Língua é um produto social da faculdade da linguagem e um conjunto de convenções necessárias, adotadas pelo corpo social para permitir o exercício dessa faculdade nos indivíduos e se apresenta na forma de um sistema de signos que se relacionam de forma organizada de dentro de um todo. (SAUSSURE, 1969 apud PETTER, 2005, p. 14).

De acordo com Cintra (1983), embora os problemas de tradução do termo “language” do inglês, e mesmo “langue” do francês dificultem a distinção entre linguagem e língua, em língua portuguesa temos significados bastante distintos para esses termos. Enquanto a

linguagem nomeia uma faculdade natural, a língua se refere a um caso particular de linguagem. Ainda segundo Cintra (1983) a língua é um sistema de signos e regras combinatórias que de fato, não se realiza completamente na fala de um sujeito. Ela só existe completamente no conjunto de uma sociedade.

“A prática da linguagem está norteada por uma tendência natural do homem em compreender, governar e modificar o mundo.” (CINTRA et al., 2002, p. 27). A linguagem é uma representação simbólica que expressa uma função psicossocial complexa e corresponde a uma manifestação intelectual e multiforme dos seres, que pode se manifestar de varias formas: linguagem verbal (oral e escrita), a pictórica, a musical, a cinética, a mímica, a documentária, etc. (CINTRA, 1983).

4.2 A LINGUAGEM NA BIBLIOTECONOMIA

O tratamento temático do documento visa estabelecer uma representação de seu conteúdo, com intuito de uma recuperação eficiente da informação por parte do usuário em um ambiente informacional eletrônico, como bases de dados online ou em catálogos manuais, estes sendo compreendidos, efetivamente, como sistemas de informação (BORBA; LAAN; CHINI, 2012).

Para a recuperação da informação em pesquisa específica é de suma importância que os documentos sejam previamente indexados, ou seja, é imprescindível que cada documento seja analisado de forma que nele sejam identificados os assuntos tratados. Esses assuntos serão representados por termos, que deverão ser utilizados sempre que esses assuntos ocorram em outros documentos que tratam do mesmo assunto.

De acordo com Lancaster (1993), a indexação implica na preparação de uma representação do conteúdo temático dos documentos. Segundo o autor, “a indexação descreve o conteúdo de um documento empregando um ou mais termos de indexação, comumente selecionados de algum tipo de linguagem documentária”. (LANCASTER, 1993, p. 2).

O processo de indexação é dividido em etapas, no entanto não há consenso na área sobre quantas etapas esse processo está dividido. De acordo com Cesarino e Pinto (1978 apud SOUSA; FUJITA, 2014), são três as etapas que envolvem o processo de indexação, sendo estas: analítico ou interpretativa, de sumarização e de tradução. A primeira compreende a identificação de tópicos na obra pelo indexador, a segunda estabelece a importância dos tópicos da primeira etapa levando em consideração a necessidade do usuário e os objetivos do

documento e a terceira etapa, a tradução, trata-se da transposição dos termos selecionados para a linguagem de indexação utilizada.

Para Lancaster (1993) a indexação compreende duas etapas: análise conceitual e tradução, sendo a análise conceitual a responsável em definir sobre o que trata o documento, e a tradução é responsável pela conversão da análise conceitual em um conjunto de termos para indexação.

Já para Chaumier (1988 apud SOUSA; FUJITA, 2014), a indexação está dividida em quatro etapas, sendo elas: conhecimento do conteúdo do documento, escolha dos conceitos que serão representados, tradução dos conceitos e a incorporação dos elementos sintáticos.

A Associação Brasileira de Normas Técnicas, sob a NBR 12676/19921, que é a responsável em subsidiar os processos de indexação, definem três etapas para o processo de indexação de um documento, sendo estes: a análise do documento e o estabelecimento do assunto contido no mesmo, a identificação dos conceitos e a tradução desses conceitos em termos para a indexação (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992 apud SOUSA; FUJITA, 2014).

A linguagem documentária (LD) é um dos instrumentos de padronização da indexação que visa garantir que indexadores usem os mesmos conceitos para representar documentos com assuntos semelhantes. Também é um instrumento de comunicação que tem por objetivo proporcionar que indexadores e usuários partilhem um mesmo vocabulário.

Na representação da informação, a ação de transferência da informação depende de linguagens formalizadas que facilitem a comunicação entre produtores, organizadores e usuários da informação.

[...] As linguagens documentárias são instrumentos intermediários, ou instrumentos de comutação, através dos quais se realiza a 'tradução' da síntese dos textos e das perguntas dos usuários. A formalização das perguntas dos usuários é feita em linguagem do próprio sistema. É por esta razão que as LD podem ser concebidas como instrumentos de comunicação documentária. (CINTRA, 1994, p. 23).

Cintra (1983) define a indexação como a tradução de um documento em termos documentários, ou seja, em descritores, cabeçalhos de assunto e palavra-chave, e esses termos documentários tem como objetivo expressar conteúdo do documento.

Na biblioteconomia, vários são os instrumentos utilizados para representar o conhecimento de uma determinada área do saber, como por exemplo, o tesauro e os esquemas

de classificação, e todos esses são denominados LD. Nos próximos capítulos será abordada os tipos de linguagens que envolvem o processo de indexação.

4.3 LINGUAGEM NATURAL

É possível afirmar que a LN é a comumente usada para comunicação pelos seres humanos em suas relações cotidianas. Segundo Lopes (2002, p. 48), a LN pode ser entendida e definida como “um vocabulário livre, sinônimo do discurso comum”, ou seja, é a linguagem utilizada habitualmente por uma comunidade em sua fala e escrita.

Para Cintra (et al., 2002) a LN é o modo pelo qual nos expressamos à experiência segundo padrões da cultura da qual participamos. Ainda de acordo com Cintra (1983, p. 7) “existem também outras formas de linguagem, como a linguagem musical, a cinética, a mímica, a documentária, etc.”. Todas são formas de linguagem, e como toda linguagem tem por objetivo comunicar alguma coisa a um receptor.

Nos documentos, os termos geralmente atribuídos para a busca do usuário em LN normalmente são termos que estão contidos no título e no resumo dos documentos, e nesse contexto cada palavra é automaticamente candidata a ser pesquisada. Já no contexto da Web, sobretudo nas Redes Sociais, toda e qualquer palavra pode ser um termo de busca, e para isso basta adicionar o símbolo cerquilha (#) seguido da palavra desejada, para que dessa forma se tenha uma hashtag, ou palavra-chave.

4.3.1 A Linguagem Natural no Contexto da Web 2.0

Com a crescente expansão tecnológica e das redes sociais os serviços de informação e comunicação na Web fazem hoje da Internet um dos principais meios de comunicação em nossa sociedade. Vemos a cada dia um aumento crescente do acesso e uso de computadores e até mesmo de smartphones⁷ pela população de um modo geral, e possibilita melhoria de acesso aos ambientes tecnológicos, ocasionando também aumento da participação em redes sociais.

⁷ O smartphone é um celular com tecnologias avançadas, o que inclui programas executados um sistema operacional, equivalente aos computadores. Disponível em: <<http://www.significados.com.br/smartphone/>>. Acesso em: 17 maio 2014.

A nova geração da internet vem transformando comportamentos e ideias com ferramentas interativas para os usuários. Esta nova e promissora fase da internet chama-se Web 2.0⁸ e é baseada na construção coletiva do conhecimento.

[...] a personalização tecnológica, proporcionada pela “Web” 2.0, abriu novas possibilidades para o desenvolvimento de produtos e serviços agregados e ampliaram a presença do usuário comum na coordenação de ações de produção, organização e difusão de informações voltadas a públicos segmentados pela natureza das redes sociais aos quais pertencem. A segmentação alcançada decorre das possibilidades de modelagem da linguagem natural para fins de representação e recuperação da informação em ambientes virtuais de trocas e culturas informacionais. (MOURA, 2009 p. 28-29).

As redes sociais incorporam, de acordo com Marteleto (2005 apud MOURA, 2009), três dimensões fundamentais: a dimensão sócio-comunicacional, a linguístico-discursiva e a produção de sentidos. Assim,

A primeira envolve os elos, as motivações e as interações entre os atores sociais, a segunda incorpora os aspectos cognitivos e informacionais envolvidos no compartilhamento social e a terceira a dimensão da produção de sentidos explicita o fluxo e a dinâmica da ação colaborativa partilhada. (MARTELETO, 2005 apud MOURA, 2009, p. 27).

Qin (2008 apud MOURA, 2009) sinaliza que esses ambientes podem ser considerados espaços sociais semânticos e identifica a existência de comunidades que se organizam a partir do compartilhamento de informações. Ainda de acordo com Qin (2008 apud MOURA, 2009), esses espaços são criados por usuários autônomos que compartilham diferentes tipos de informação e distintas visões do mundo que os cerca.

As tags ou hashtags (etiquetas) são à base dos espaços sociais semânticos e se faz necessário um estudo interdisciplinar envolvendo as áreas de Linguística para o estudo das palavras, da Ciência da Computação para o estudo dos símbolos pragmáticos, da Semiótica e Psicologia para o entendimento das significações e da Biblioteconomia e Ciência da Informação para os metadados. (QIN, 2008 apud MOURA, 2009 p. 30).

No contexto da Web 2.0 e com o uso cada vez mais frequente das tags ou hashtags surge a folksonomia, e esta tem objetivo à representação livre da informação em LN sem adotar regras ou políticas de indexação, e dessa forma, a folksonomia é o resultado da

⁸ O termo que faz um trocadilho com o tipo de notação em informática que indica a versão de um “software”, foi popularizado pela O’Reilly Media e pela Media Live International como denominação de uma série de conferências que tiveram início em outubro de 2004 (O’REILLY, 2005 apud PRIMO, 2006, p. 1).

etiquetagem dos recursos da Web, em um ambiente social, pelos próprios usuários com o objetivo de uma futura recuperação, conforme será abordado a seguir.

4.3.2 Folksonomia e as Tags

Devido a grande demanda de informação no ambiente da Web 2.0 surgem ferramentas propícias para favorecer a comunicação nesse ambiente. Esta conexão atualmente é possibilitada pela marcação por tags. O uso dessa marcação permite que pessoas sem nenhum conhecimento técnico possam ser indexadoras de informação em LN através das tags reforçando suas ideias e identidades em seus ciclos sociais.

[...] Pessoas marcam com etiquetas (tags) e ou com palavras-chaves (metadados) conteúdos a ser evidenciados [...] em contextos sociais participativos. A indexação emerge de maneira bottom-up e forma-se um “sistema de classificação de conteúdos “Web” gerado pelos usuários que permite aos mesmos atribuir tags aos seus recursos digitais favoritos através de palavras ou frases selecionadas de uma linguagem natural” (NORUZI, 2006 apud PEREIRA; CRUZ, 2010, on-line).

De acordo com Catarino e Batista (2006 p. 49), o termo folksonomia foi usado pela primeira vez em 2004 e se deu a partir da junção de dois conceitos existentes, Folk, que significa povo e Taxonomy que significa Taxonomia. Dessa forma, a folksonomia tem por objetivo o uso da LN, a linguagem do povo, em arquiteturas da informação, pelas quais redes sociais alimentam bancos de dados.

As marcações sociais abordadas pela folksonomia, essa “linguagem do povo”, utilizada nas redes sociais foram analisadas e segundo Javier Canadá (2006 apud MOURA, 2009, p. 36) e essas “marcações sociais” acontecem por diversas motivações, às quais podem ser denominadas egocêntrica, amigável, altruísta e populista, como é possível observar no quadro abaixo.

Quadro 2 – Perspectivas de usuários na classificação social ou folksonomia

| |
|---|
| A perspectiva egocêntrica |
| Orienta-se pela descrição pessoal da informação com vista a sua recuperação futura, sem, contudo, haver uma preocupação com o fato de essa informação poder ser útil a outros sujeitos em distintos contextos sociais. Nesse caso, a indexação realizada funciona como lembretes mnemônicos que revelam o universo semântico de seu criador. |
| A perspectiva amigável |
| Introduz um primeiro nível de colaboração que se manifesta em um acordo de linguagem que é adotado na descrição dos objetos informacionais em contextos digitais. Nesses casos, os parceiros em cooperação pactuam uma linguagem de referência que tem a validade circunscrita ao “círculo íntimo” de compartilhamento. |
| A perspectiva altruísta |
| Também se manifesta no nível da linguagem e incorpora um segundo nível de cooperação. Nessa abordagem os agentes tentam estabelecer alguma regularidade e padronização na adoção das tags, de modo que a recuperação da informação seja mais efetiva e garanta o acesso de uma gama mais ampla de usuários. |
| A perspectiva populista |
| Por outro lado, busca incorporar o aprendizado da dinâmica colaborativa em rede para ofertar conteúdos comerciais embalados em etiquetas de apelo popular. Nesse sentido, essa perspectiva viabiliza comercialmente o monitoramento das práticas colaborativas na Web ao vincularem o universo semântico de determinadas etiquetas à oferta de bens e serviços. |

Fonte: Moura (2009, p. 36)

É possível observar claramente no quadro que são diversos os motivos que levam usuários a fazer a marcação em conteúdo disponibilizado em suas redes, dessa forma, preocupada com o impacto social da atividade de indexação social na descrição e recuperação da informação, Moura (2009) acredita que a ação direta do usuário disponibilizando conteúdos e identificando-os por meio das tags ou hashtags nas redes sociais, fazem com que

a Ciência da Informação tenha a preocupação de constituir um plano de formação humana voltado a esse contexto.

Portanto, a seguir será abordado como se dão o uso da LN utilizada nessas marcações em redes sociais, mais especificamente na Rede Social Twitter.

4.3.3 Rede Social Twitter

Criado como um serviço para celular, o Twitter nasceu para caber no limite de caracteres de uma mensagem de texto, no entanto, essa característica do texto curto é o que mantém o serviço do Twitter rápido e relevante, incentivando as pessoas a tweetarem⁹ de modo fácil e somente com a essência da ideia que estão tentando comunicar.

As Redes Sociais, de um modo geral representam um espaço de compartilhamento de opiniões, experiências e conteúdo. Dentro das Redes, cada usuário se torna um produtor de conteúdo de informação para todas as pessoas presentes em sua rede. Dentro de um Tweet, é possível ver fotos, vídeos e incluir links para notícias, blogs, Websites e aplicativos (DISCOVER TWITTER, 2014).

No Twitter, os usuários usam o símbolo de marcador # antes de uma palavra-chave ou frase relevante (sem espaços) nos Tweets que postam, e dessa forma essa marcação tem a finalidade de categorizar esses Tweets e facilitar a sua localização em uma busca no próprio Twitter. Clicar na palavra com o símbolo # em uma mensagem mostra todos os outros Tweets marcados com essa palavra-chave. Os marcadores (#) podem ocorrer em qualquer parte de um Tweet – no início, meio ou fim. Palavras marcadas que passam a ser muito populares são, muitas vezes, “Assuntos do Momento” como veremos a seguir.

4.3.3.1 As #hashtags e o Twitter

As palavras-chave são uma unidade tanto de representação como de recuperação da informação, e diante dessa afirmação, o que se pretende discutir é a capacidade de representação da informação e sua recuperação, cuja demanda de informação seja compatível com o assunto representado pela palavra-chave ou hashtag.

⁹ Ato de enviar mensagens pela plataforma do Twitter.

De acordo com Primo (2010, on-line) a “tag ou hashtag é um recurso que foi criado dentro do micro blog Twitter de forma espontânea pelos seus usuários”. De acordo com o autor, a hashtag, que é um tipo de palavra-chave, é determinada pelo sinal # (cerquilha) seguido da palavra ou conjunto de palavras escritas juntas e sucessivas sem nenhum outro tipo de pontuação. Foi cunhado pela primeira vez de modo a facilitar a organização de mensagens sobre um mesmo tema. Segundo Primo (2010), o primeiro usuário a utilizar este recurso foi Nate Ritter em 22 de outubro de 2007 utilizando as hashtag #sandiegofire nos relatos que vinha fazendo sobre os incêndios que estavam acontecendo em San Diego.

No Brasil, as hashtags vêm também sendo muito usadas, como é o caso recente da campanha pela não violência contra mulher, #NãoMereçoSerEstuprada, por exemplo. Ainda de acordo com Primo (2010, on-line), “Tagging, que significa etiquetar, é um processo de associação de metadados, que são dados sobre dados, a textos, imagens, bookmarks, etc.”

Em LN, qualquer palavra, ou até mesmo conjunto de palavras podem ser utilizadas como hashtag, e estas palavras-chave facilitam a recuperação de informações em sistemas de busca como o Twitter Search. Este serviço de busca dentro da plataforma do Twitter oferece além de uma busca simplificada, outros elementos como operadores e buscas avançadas, como será possível observar nas figuras abaixo.

Imagem 8 – Busca simples na plataforma do Twitter



Fonte: Twitter (2014)

Imagem 9 – Operadores de buscas na plataforma do Twitter

| Operadores de Busca X | |
|--|---|
| Operador | Encontrar tweets... |
| twitter busca | contendo ambas as palavras "twitter" e "busca". Este é o operador padrão. |
| "diversão" | contendo exatamente a frase "diversão". |
| amor OU ódio | contendo apenas "amor" ou "ódio" (ou ambas). |
| cerveja-raiz | contendo "cerveja" mas não "raiz". |
| #haiku | contendo o marcador "haiku". |
| de:alexiskold | enviado por "alexiskold". |
| para:techcrunch | enviado para "techcrunch". |
| @mashable | falando sobre a pessoa "mashable". |
| "lazer" próximo:"são francisco" | contendo a frase exata "lazer" e enviado próximo a "são francisco". |
| near:NYC within:15mi | enviado a 15 milhas de "NYC". |
| super-herói since:2010-12-27 | contendo "super-herói" e enviado até "2010-12-27" (ano-mês-dia). |
| ftw until:2010-12-27 | contendo "ftw" e enviado até "2010-12-27". |
| filme -terror:) | contendo "filme" mas não "terror" e com uma atitude positiva. |
| voo :(| contendo "voo" e com uma atitude negativa. |
| tráfego ? | contendo "tráfego" e fazendo uma pergunta. |
| hilário filter:links | contendo "hilário" com links para URLs. |
| notíciassource:twitterfeed | contendo "notícias" e publicado através doTwitterFeed |

Fonte: Twitter (2014)

Imagem 10 – Busca avançada na plataforma do Twitter

The image shows the 'Busca Avançada' (Advanced Search) interface on Twitter. It is organized into several sections:

- Palavras (Words):** Includes input fields for 'Todas estas palavras', 'Exatamente esta frase', 'Qualquer uma destas palavras', 'Nenhuma destas palavras', and 'Estes marcadores'. There is also a dropdown menu for 'Escrito em' (Written in) set to 'Qualquer Idioma'.
- Pessoas (People):** Includes input fields for 'Destas contas', 'Para estas contas', and 'Mencionando estas contas'.
- Locais (Locations):** Shows 'Perto deste local' with a location pin icon and the text 'Localização desativada'.
- Datas (Dates):** Includes a date range selector with 'A partir desta data' and 'para'.
- Outro (Other):** Includes a 'Selezione:' section with checkboxes for 'Positivo :)', 'Negativo :(', 'Perguntas?', and 'Incluir retweets'.

A blue 'Buscar' button is located at the bottom left of the form.

Fonte: Twitter (2014)

O próprio Twitter facilita a recuperação de tweets, que são as mensagens postadas na plataforma do Twitter e que possuem até 140 caracteres, com a mesma etiqueta transformando todas tags em links em suas páginas oficiais. De modo geral, para facilitar esse processo de recuperação costuma-se usar etiquetas curtas e que sejam descritivas do assunto em questão, no entanto, etiquetas maiores são admitidas pelo sistema, como exemplo, de uma hashtag curta se vê ao lado a hashtag #estupro, ou como exemplo uma hashtag mais longa #NãoMereçoSerEstuprada. Quem tiver interesse nesse assunto pode fazer uma busca sobre

algumas dessas hashtags e encontrar tudo o que está sendo falado no momento sobre ela. Vale acrescentar que o único limite para o tamanho das hashtags no Twitter são os 140 caracteres admitidos pelo sistema, como é possível ver abaixo:

Imagem 11 – Testando limite de palavras em “hashtag” no Twitter



Fonte: Pacheco (2014)

Segundo Primo (2010), as hashtags se tornaram tão importantes para o Twitter, que em 30 de abril de 2009 foi incorporada a plataforma do Twitter a seção Trending Topics (TTs), que tem como base as hashtags que estão sendo mais difundidas na plataforma.

Além dos TTs, o Twitter também passou a publicar a lista das hastags mais utilizadas anualmente, ou seja, de um movimento espontâneo criado por um usuário, as hashtags foram oficialmente incorporadas pelo Twitter, e hoje contam com toda uma estrutura que facilita a busca e recuperação de mensagens (PRIMO, 2010).

A Web 2.0 permite hoje o uso de linguagens mais flexíveis e de padrões cada vez mais aceitos para representação da informação. De acordo com Campos (2001), todo movimento existente nos sistemas de recuperação de informação tem por objetivo possibilitar ao usuário acesso a informação. Para o autor, vários são os instrumentos para representar o conhecimento

em uma dada área do saber. Na próxima seção aprofundaremos mais sobre esses instrumentos, no qual na área da Biblioteconomia são denominadas LDs.

4.4 LINGUAGEM DOCUMENTÁRIA

Uma das principais características da LD é a substituição do texto do documento por sua descrição abreviada, utilizada como um artifício para recuperar o que é essencial no documento, ou seja, o tema central do documento.

“A linguagem é uma representação simbólica que expressa uma função psicossocial complexa e corresponde a uma manifestação intelectual e multiforme dos seres, que pode se manifestar de varias formas” (CINTRA, 1983, p. 7). A LN é a linguagem utilizada em nossas necessidades diárias, ou seja, é a linguagem de uso cotidiano. A LN normalmente se opõe a linguagem artificial ou controlada (LD), pois é uma forma de linguagem artificial que surge das dificuldades que a LN oferece para operar com precisão a descrição de documentos.

Novellino (1996, p. 38) afirma que no processo de representação da informação, uma das principais características é a “substituição de uma entidade linguística longa e complexa, o texto do documento, por sua descrição abreviada, de modo a evidenciar a essência do documento”. Portanto,

[...] As linguagens documentárias são instrumentos intermediários, ou instrumentos de comutação, através dos quais se realiza a ‘tradução’ da síntese dos textos e das perguntas dos usuários. A formalização das perguntas dos usuários é feita em linguagem do próprio sistema. É por esta razão que as LD podem ser concebidas como instrumentos de comunicação documentária. (CINTRA, 1994, p. 23).

Para Fujita (2004, on-line), “as LDs são um conjunto controlado de termos que visam à representação de conceitos significativos de assuntos dos documentos utilizados na fase de indexação e busca”. De acordo com a autora, as LDs são instrumentos de organização e comunicação da informação e tem como função ajudar o usuário a encontrar o que necessita.

Segundo Cintra (et al., 2002), grande parte das discussões teóricas sobre LDs estão inseridas no âmbito da análise documentária (AD), que é uma atividade metodológica específica da Documentação, tendo como objetivo analisar a representação da informação para posterior recuperação. “Nesse contexto as LDs são instrumentos intermediários através

dos quais se realiza a tradução da síntese dos textos e das perguntas dos usuários”. (CINTRA et al., 2002, p. 34).

Por princípio, as LDs estão estabelecidas através de convenções no conjunto do próprio sistema. No entanto, Cintra (et al., 2002), ressalta que o sistema não é tão estático e homogêneo quanto se propõe, uma vez que as LDs acabam sendo influenciadas pela flexibilidade da LN e diz ainda, que a partir da LN são retiradas as unidades que serão transformadas em unidades documentárias através de escolhas lexicais¹⁰.

As LDs são instrumentos de representação do conteúdo dos documentos informacionais com fins de classificação ou busca de informações. Os sistemas de classificação decimais, como a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU), os tesouros, dentre outros, são tipos de linguagens documentárias, que permitem agrupar documentos segundo o seu conteúdo, visando ao armazenamento e à recuperação da informação.

De acordo com Novellino (1996) nem sempre as LDs foram pensadas como instrumentos de indexação e recuperação, e no princípio as LDs possuíam um único objetivo, que era a padronização das entradas de assunto de catálogos ou índices. “As primeiras LDs foram as listas de cabeçalhos de assunto, logo após, a adesão ao vocabulário livre, isto é, a opção pela ausência de um controle do vocabulário usado para a indexação.” (NOVELLINO, 1996, p. 38). No entanto, depois se voltou ao controle do vocabulário, empregando-se as listas de termos autorizados.

Para Cintra (et al., 2002, p. 39) as linguagens documentárias são cada vez mais importantes nos dias de hoje “sendo importantes ferramentas na organização e distribuição de informação, imprescindíveis para agregar valor à informação especializada, auxiliando a tarefa de organizar tematicamente a informação”.

De acordo com Novellino (1996, p. 39), havia uma preocupação com a criação de um instrumento de representação da informação voltado para a recuperação, e, conseqüentemente, demonstrar ao usuário a estrutura da linguagem de representação dando origem aos tesouros, tesouros facetados e classauros.

O tesouro facetado e o classauro apresentam duas ordenações: a alfabética e a classificada, o que permite tornar visível ao usuário do tesouro, seja ele o indexador ou o usuário do sistema, a classificação a ele subjacente e que antes só era clara aos elaboradores do instrumento. Eles surgiram como

¹⁰ Definição de lexicologia - A Lexicologia está sendo tomada como disciplina que recobre parte da semântica e da sintaxe, já que trata da natureza e organização do vocabulário de uma língua. (CINTRA, 1983).

tentativa teórica e prática de reunir as vantagens da linguagem documentária verbal e dos sistemas de classificação facetados, assumindo que a teoria da classificação facetada seria a base para a estruturação de uma linguagem documentária verbal. (NOVELLINO, 1996 p. 40).

Entre os principais tipos de LDs, estão os sistemas de classificação bibliográfica, as listas de cabeçalhos de assunto e os tesouros. Em suas definições, ambos são instrumentos de indexação por assunto (esses instrumentos não fazem parte do escopo da pesquisa), e os tesouros são instrumentos de indexação por conceitos, como veremos a seguir.

4.4.1 Os Tesouros

O tratamento da informação tem como objetivo estabelecer uma representação de seu conteúdo com intuito de uma recuperação eficiente por parte do usuário. De acordo com Moreira (2005, p. 60), “um tesouro é um vocabulário de termos de determinada área do conhecimento, relacionados genérica e semanticamente. É uma forma de representação do conhecimento de determinada área do conhecimento”.

Os tesouros surgiram diante a necessidade de recuperação da informação e são uma linguagem controlada que tem como característica os relacionamentos entre os termos que o compõe (MOREIRA; MOURA, 2006).

Os tesouros constituem uma ferramenta de indexação já consolidada nas atividades de organização da informação empregada por muitos que exercem essas atividades. A flexibilidade de estabelecimento de novas relações entre seus termos, o estabelecimento de hierarquias e referências cruzadas conferem ao instrumento uma multiplicidade de usos, abrangendo os processos desde a indexação até o suporte para a efetiva recuperação dos documentos. (MOREIRA, 2005, p. 32).

De acordo com Gomes (1950 apud DODEBEI, 2002, p. 66), “o tesouro surge a partir de uma necessidade de tratamento de um grande volume de documentos especializados, de modo que, era preciso um instrumento com vocabulário mais específico dos que eram encontrados cabeçalhos de assunto”.

De acordo com Dodebei (2002, p. 67):

[...] nas tarefas de indexação e recuperação da informação, os tesouros têm como objetivos tentar solucionar problemas de distribuição de documentos nas classes de assunto por se tratar de um instrumento que relaciona termos

mais consistentes e apresenta uma estrutura sintética mais simples e uma rede de referências cruzadas mais complexa, de modo a permitir ao especialista uma forma mais fácil de encontrar a palavra-chave para sua busca.

Para Moreira (2005, p. 28) os tesouros são construídos de acordo com as especificidades de uma determinada área do conhecimento, e tendem a aprofundar três pontos fundamentais: “a garantia literária, a garantia de uso e a garantia estrutural - constituem o referencial para a construção de tesouros”.

A garantia literária sugere que as classes utilizadas em LDs devem se originar de classes existentes na literatura, ou seja, não se originam da classificação do conhecimento, deste modo, na garantia literária o termo escolhido para a representação dos assuntos deve ser derivado da própria literatura que se está representando.

A garantia do usuário, como o próprio nome sugere parte do princípio que os termos selecionados para um tesouro precisam estar de acordo com aqueles utilizados pelos usuários na recuperação de informação.

A garantia estrutural se dá quando os termos não são nem da garantia literária, nem da garantia do usuário. São termos que facilitam elos em uma hierarquia/estrutura ou colaboram para que seja possível dispor um conjunto mais específico de termos.

De acordo com Moreira e Moura (2006, on-line), “o tesouro avançou na estrutura e nas referências cruzadas, dando lugar às relações hierárquicas (verticais) e associativas (horizontais)”. Segundo os autores, os termos conferem a um tesouro uma multiplicidade de usos, como a recuperação da informação e a indexação. Nessa direção,

Moreira (2005, p. 122) apresenta de forma genérica, que a elaboração de tesouros se dá através de três tipos de relacionamentos: “o relacionamento de equivalência, o relacionamento hierárquico e o relacionamento associativo.”

As relações de equivalência acontecem quando existem sinônimos, ou quase sinônimos, e dessa forma, um termo pode ser representado por dois ou mais termos. Quando isso ocorre, um dos termos é escolhido como termo preferido e os demais como não preferidos.

De acordo com Currás (1995, p. 22), o “termo preferido é definido como um termo utilizado consistentemente para representar um conceito no processo de indexação, sendo também conhecido como descritor, ou termo principal”. Ainda de acordo com Currás os termos não preferido, chamados de não descritor ou secundário são utilizados como remissa, conduzindo o usuário ao termo preferido.

As relações equivalentes são representadas pelos indicadores **USE** e **UP**. O primeiro (USE) é utilizado antes do termo preferencial, remetendo o termo equivalente, e o segundo (UP), que significa Usado Por, tem a função de remeter os termos equivalentes para o termo preferencial, como no exemplo abaixo:

| | |
|---------------------|---------------------|
| MANDIOCA | Macaxeira |
| UP Macaxeira | USE MANDIOCA |

De acordo com Araújo (2011), as relações hierárquicas são apresentadas em forma de níveis hierárquicos de superordenação e subordinação entre os conceitos. Nessas relações o termo superordenado representa uma classe, enquanto os termos subordinados representam uma parte da classe.

Essas relações se dividem em três tipos: relacionamento genérico (gênero/espécie) representado pelos indicadores TG (termo geral) e TE (termo específico); relacionamento partitivo (todo/parte) representado pelos indicadores TGP (termo geral partitivo) e TEP (termo específico partitivo) e relacionamento de instância, representado pelos indicadores TGI (termo geral de instância) e TEI (termo específico de instância) (ARAÚJO, 2011). Assim observa-se:

- ✓ Relacionamento genérico (gênero/espécie): ocorre quando dois conceitos diferentes possuem as mesmas características idênticas e um deles possui uma característica a mais do que o outro (DAHLBERG, 1978 apud ARAÚJO, 2011, p. 29). Exemplo:

TG – Árvore
TE – Árvore frutífera

- ✓ Relacionamento partitivo (todo/parte): ocorre quando um conceito está intrinsecamente incluído em outro de forma que os termos podem ser organizados em hierarquias lógicas (ANSI/NISO Z39.19, 2003 apud ARAÚJO, 2011, p. 29-30). O todo representa o termo superordenado e a parte o termo subordinado. Exemplo:

TGP – Árvore
TEP – Caule
TEP – Folhas

- ✓ Relacionamento de instância: ocorre entre uma categoria de coisas ou eventos, expressos por um nome comum, e uma instancia individual dessa categoria, representada geralmente por um nome próprio (ANSI/NISO Z39.19, 2003 apud ARAÚJO, 2011, p. 30). Exemplo:

TGI – Regiões montanhosas

TEI – Alpes

TEI – Himalaia

É possível observar que:

[...] As relações associativas são aquelas ocorrências entre os termos que não apresentam nenhum relacionamento de equivalência ou hierarquia, porem são semanticamente ou conceitualmente associadas, necessitando de ser explicada dentro de um vocabulário controlado” (ANSI/NISO Z39.19, 2003 apud ARAÚJO, 2011, p. 31).

As relações associativas são representadas pelo indicador TR (termo relacionado), e também se encontra na literatura os indicadores TA (termos associado) e TC (termo correlato) e Segundo Currás (1995), as relações associativas podem acontecer entre termos que pertencem à mesma categoria ou em termos que pertencem à categorias diferentes. Exemplo:

Termos de mesma categoria:

TG – Bibliografias Nacionais

TR – Bibliografias especiais

Termos de categorias diferentes:

- ✓ Relação do todo com a parte.

TG – Edifício

TR – Portas

- ✓ Uma ocupação e a pessoa que a exerce.

TG – Contabilidade

TR – Contabilista

- ✓ Uma matéria-prima e seu produto.

TG – Farinha

TR – Pão

Nesse sentido,

As relações associativas constituem-se um desafio para os indexadores, pois, pelo fato de não apresentarem as clarezas das relações hierárquicas e de equivalência, exigem dos profissionais da informação um profundo conhecimento, tanto da estrutura do tesouro quanto do assunto abordado por ele, a fim de que seja possível viabilizar o estabelecimento desses vínculos conceituais. (MARRONI, 2006, p. 64).

De acordo com Lopes (2002, p. 47), “o tesouro pode ser compreendido como um conjunto de termos organizados hierarquicamente ou alfabeticamente, de modo a possibilitar a recuperação de informações temáticas restringindo as diversidades pelo uso da terminologia adequada”. Segunda a autora, quando uma base de dados utiliza o controle do vocabulário através do tesouro, cria a possibilidade, ao intermediário no planejamento da estratégia de busca, a recuperação, no campo específico de descritor, apenas daquelas palavras-chave listadas no tesouro da base de dados.

Realizadas as explicações sobre o instrumento tesouro, as próximas subseções irão abordar dois tipos de tesouros específicos que embasará a discussão dos dados coletados para essa pesquisa.

4.4.1.2 O tesouro jurídico do Superior Tribunal de Justiça (STJ)

Criado pela Constituição Federal de 1988, o Superior Tribunal de Justiça (STJ) é a corte responsável por uniformizar a interpretação da lei federal em todo o Brasil, seguindo os princípios constitucionais e a garantia e defesa do Estado de Direito (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014).

O STJ é a última instância da Justiça brasileira para as causas infraconstitucionais, não relacionadas diretamente à Constituição. Como órgão de convergência da Justiça comum, aprecia causas oriundas de todo o território nacional, em todas as vertentes jurisdicionais não especializadas. O STJ julga crimes comuns praticados por governadores dos estados e do Distrito Federal, crimes comuns e de responsabilidade de desembargadores dos tribunais de justiça e de conselheiros dos tribunais de contas estaduais, dos membros dos tribunais regionais federais, eleitorais e do trabalho. Julga também “habeas-corpus” que envolvam

essas autoridades ou ministros de Estado, exceto em casos relativos à Justiça Eleitoral. Pode apreciar recursos contra “habeas-corpus” concedidos ou negados por tribunais regionais federais ou dos estados, bem como causas decididas nessas instâncias, sempre que envolverem lei federal (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014).

No que tange ao interesse desse trabalho, o Tesouro Jurídico do STJ, gerenciado pela Secretaria de Jurisprudência, é uma lista de termos jurídicos e de conexão acompanhados das relações que se estabelecem entre eles. Pretende ser um instrumento adequado para: maximizar o uso da informação jurisprudencial, atingindo níveis crescentes de acessibilidade para usuários internos, ministros e servidores, outros órgãos do Poder Judiciário, advogados e público em geral; atingir maior uniformidade e flexibilidade no tratamento da informação; facilitar o intercâmbio da informação entre sistemas e bases de dados, possibilitando uma integração mais abrangente (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014).

O Tesouro Jurídico é um instrumento importante no tratamento da informação realizado pela Secretaria de Jurisprudência, porque permite a padronização da linguagem. A ideia é proporcionar ao usuário da jurisprudência uma maior precisão na busca da informação.

Os termos são organizados em ordem alfabética, acompanhados de suas relações e códigos das categorias nas quais estão inseridos. O Tesouro contém aproximadamente 12.500 termos, sendo constantemente atualizado, conforme requer a dinâmica do Direito (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014).

De acordo com o Tesouro Jurídico do STJ (on-line), os tipos de relações estabelecidas entre conceitos são: relação de equivalência, relação hierárquica, polierarquia e relação associativa e o tesouro é composto de descritores, não-descritores e modificadores como veremos a seguir.

- **Descritores:** são termos simples ou compostos (substantivos ou frases substantivadas) autorizados pelo Tesouro para representarem conceitos e proporcionar recuperação de informação, optando-se por utilizar a terminologia usualmente empregada pelos ministros, selecionada dos acórdãos incluídos na base de dados da jurisprudência do STJ;

- **Não-Descritores:** são termos que, embora representem os mesmos conceitos que os descritores, não são autorizados para uso no tratamento da informação, servindo apenas para indicar sinonímia no intuito de facilitar a recuperação da informação, em especial na página de Pesquisa de Jurisprudência onde há a opção de se realizar a pesquisa por sinônimos. Esses termos são representados pela indicação de termo correspondente, através da anotação USE;

- **Modificadores:** são termos autorizados pelo tesauro e utilizados para modificar ou complementar o descritor principal. Os modificadores tiveram muita importância quando se fazia a indexação dos documentos e, também, quando se evoluiu para o resumo estruturado e, depois, para as informações complementares.

O Tesauro Jurídico encontra-se organizado em categorias e subcategorias, utilizando-se, para essa estruturação, as subdivisões dos textos legais pertinentes. Essas categorias foram codificadas utilizando-se caracteres alfanuméricos. Para a padronização dos termos incluídos no Tesauro, foram adotados os seguintes critérios:

- a) Uso predominante de substantivos - uso de substantivos para representar os conceitos, não se permitindo o uso de verbos e adjetivos isoladamente, salvo pequenas exceções para alguns modificadores;
- b) Substantivo no singular ou no plural - uso de termos na forma singular, exceto quando esta modifica o sentido do termo ou quando o termo técnico é utilizado somente no plural;
- c) Empréstimos linguísticos - uso da tradução ou do termo estrangeiro dependendo do grau de incorporação na linguagem especializada;
- d) Gírias e nomes comerciais - uso como descritores somente nos casos em que não exista substituto aceito;
- e) Termos em latim - uso de termos em latim quando o conceito jurídico é mais conhecido desta forma;
- f) Siglas e abreviaturas - uso das siglas, fazendo-se remissão ao termo por extenso só quando facilmente reconhecido;
- g) Parênteses - inclusão, quando necessário, de parênteses "()" para adicionar um qualificador ao descritor;

h) Notas explicativas - inclusão de notas explicativas para esclarecer o significado de um descritor ou sua utilização na indexação (SUPERIOR TRIBUNAL DE JUSTIÇA, 2014).

Os tesouros são instrumentos específicos de uma determinada área do conhecimento que visam atender uma demanda de informação especializada, deste modo um mesmo termo pode se apresentar de forma hierárquica diferente, ou até mesmo ter uma definição diferente dependendo da área de conhecimento ao qual está inserido. A partir desta ideia foram consultados dois tesouros de áreas diferentes, de modo que se possa verificar a diferença de apresentação do termo “estupro” nesses dois instrumentos de representação da informação, como abordaremos a seguir.

4.4.1.2.1 O termo Estupro no tesouro jurídico do STJ

Em pesquisa efetuada dentro do tesouro jurídico do STJ pelo termo estupro foi possível identificar como se apresentam os termos, a subordinação, os termos gerais (TG), os termos específicos (TE) e seus termos relacionados (TR).

✓ **ESTUPRO**

TG1 Crime Contra a Liberdade Sexual

TG2 Crime Contra os Costumes

TG3 Crime

TG4 Delito

TG1 Crime Hediondo

TG2 Crime

TG3 Delito

TE1 ESTUPRO DE VULNERÁVEL

TE1 ESTUPRO PRESUMIDO

TR Aborto Necessário

TR Conjunção Carnal

TR Consentimento

TR Delegacia da Mulher

TR Grave Ameaça

TR Presunção de Violência

TR Violência

✓ **TE1 ESTUPRO DE VULNERÁVEL**

TG1 Estupro

TG2 Crime Contra a Liberdade Sexual

TG3 Crime Contra os Costumes

TG4 Crime

TG5 Delito

TG2 Crime Hediondo

TG3 Crime

TG4 Delito

TR Ato Libidinoso

TR Conjunção Carnal

TR Crime Sexual Contra Vulnerável

TR Deficiência Mental

TR Exploração Sexual de Vulnerável

TR Lesão Corporal Grave

TR Menor de Catorze Anos

TR Presunção de Violência

TR Violência Sexual

TR Vítima

✓ **TE1 ESTUPRO PRESUMIDO**

TG1 Estupro

TG2 Crime Contra a Liberdade Sexual

TG3 Crime Contra os Costumes

TG4 Crime

TG5 Delito

TG2 Crime Hediondo

TG3 Crime

TG4 Delito

TR Abuso Sexual

TR Conjunção Carnal

TR Delegacia da Mulher

TR Grave Ameaça

TR Violência

TR Violência Sexual

É possível observar na construção hierárquica do termo estupro apresentada acima, que aparecem como TG1 Crime Contra a Liberdade Sexual e Crime Hediondo, como TG2 Crime Contra os Costumes e Crime, como TG3 Crime e Delito e como TG4 Delito. Nota-se que há uma relação de *Polierarquia*, este tipo de relação permite que um termo específico, como no caso do termo “estupro”, possua uma relação direta com mais de um termo genérico.

Ex: TG1 - crime hediondo

TG1 - crime contra a liberdade sexual

4.4.1.3 O Thesaurus Brased – Thesaurus Brasileiro de Educação

O Thesaurus Brasileiro da Educação (Thesaurus Brased) começou a ser desenvolvido em 1980, sob a coordenação do professor Gaetano Lo Monaco e foi criado no intuito de facilitar a pesquisa em educação, sendo pioneiro na área de educação no Brasil. O diferencial do Brased em relação a outros vocabulários controlados de educação do país é que os termos são selecionados e estruturados dentro de uma matriz conceitual, elaborada a partir de uma análise crítica da realidade educacional e de seu contexto. A base conceitual na qual o Thesaurus Brased foi desenvolvido considera a educação em seu contexto global e interdisciplinar, o que permite ao estudioso analisá-la e compreendê-la em profundidade (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2011).

O Thesaurus Brased está inserido no Centro de Informação e Biblioteca em Educação – Cibec, que é um centro de documentação e de informação dedicado à conservação e disseminação de informações produzidas pelo INEP e possui um acervo que inclui documentos históricos, livros, inclusive obras raras, periódicos, obras audiovisuais em diversos formatos e outros de sua área de especialidade - a educação brasileira.

O Thesaurus Brased é um vocabulário controlado que reúne termos e conceitos, extraídos de documentos analisados no Centro de Informação e Biblioteca em Educação (Cibec), relacionados entre si a partir de uma estrutura conceitual da área. Estes termos chamados descritores são destinados à indexação e à recuperação de informações. (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2011).

4.4.1.3.1 O termo Estupro no Thesaurus Brased

Apresentou-se o termo “estupro” no Tesouro Jurídico e suas relações hierárquicas, e agora será abordado como se apresenta o mesmo termo no Tesouro Brasileiro de Educação, o Brased.

O referido tesouro insere a educação dentro de um contexto global, e de acordo com a matriz conceitual, o Brased é composto de quatro campos (ou subáreas), que delimitam a abrangência da área de Educação, a saber:

100 - Contexto da Educação: a educação do homem se realiza dentro da realidade global e em interação com esta; fora desta não há educação.

200 - Escola como instituição social: a escola é a educação institucionalizada; na sociedade politicamente organizada, de fato, encontraremos todas as condições para que a educação do homem socialmente aconteça.

300 - Fundamentos da Educação: a educação é o principal processo do desenvolvimento humano, que é interdisciplinar, isto é, muitas ciências fundamentam e integram no processo e a ação educativa.

400 - Educação: princípios, conteúdo e processo: o homem evolui interagindo constantemente com o meio: é a educação propriamente dita com seus princípios, conteúdo e processo. (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2011).

O Tesouro Brased possui seus termos estruturados dentro da lógica da sua matriz conceitual que contempla os quatro campos temáticos citados acima. Dentro da estrutura do tesouro, os termos são ordenados de acordo com as relações de hierarquia, de equivalência e de associações. Há ainda o campo denominado Identificadores e Especificadores de Informação, que classifica e complementa as informações relacionadas aos quatro primeiros

campos (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2011).

Abaixo se apresenta a estrutura do tesouro Brased e como está disponibilizado os desdobramentos hierárquicos do termo estupro.

100 CONTEXTO DA EDUCAÇÃO

110 Contexto ambiental

- 111 Meio Ambiente
- 112 Problemas Ecológicos
- 113 Política Ambiental
- 114 Meio Ambiente e Educação

120 Contexto humano

- 121 Pessoa
- 122 Família
- 123 Comunidade

130 Contexto social

TG Contexto social

TA Sociedade

TA Sociologia e Educação

TA Fatores Sócio educacionais

TA Sociologia

TE *Condições Sociais*

TR População

TR Realidade Social

TR Política Social

TG Condições Sociais

TA Condições Educacionais

TE *Problemas sociais*

TR Deficiências Sociais

TR Injustiças Sociais

TR Problemas Sociais

TG Problemas sociais

TA Problemas Sócio educacionais

TA Problemas do Ambiente Humano
TA Questões Morais
TA Problemática da Educação
TE *Criminalidade*
TR Delinquência
TR Desvios Sexuais
TR Desvios Sociais
TR Insegurança Social
TR Intolerância Social
 TG Criminalidade
 TE *Crime*
 TR Aumento da Criminalidade
 TR Corrupção
 TR Crime
 TR Criminoso
 TR Impunidade
 TR Reincidência No Crime
 TG Crime
 UP Delito
 TE *Delitos Sexuais*
 TR Crime por Imitação
 TR Crimes Públicos
 TR Roubo
 TR Tráfico de Drogas
 TR de Tóxicos
 TR Vadiagem
 TR Vandalismo
 TG Delitos Sexuais
 TA Orientação sexual
 TE *Estupro*
 TR Pornografia
 TR Proxenetismo
 ESTUPRO

O termo estupro se apresenta no tesouro como ultimo termo, sendo ele o mais especifico dentro da cadeia hierarquicamente apresentada. A hierarquia apresentada até o termo estupro vem se desdobrando desde o topo da cadeia que é *contexto da educação*, passando por contexto social, condições sociais, problemas sociais, criminalidade, crime, delitos sexuais e por fim estupro. Abaixo continua a estrutura geral do Tesouro Brased.

140 Contexto cultural

- 141 Meio Cultural
- 142 Política Cultural
- 143 Lazer e Educação
- 144 Esporte e Educação

150 Contexto político

- 151 Meio Político
- 152 Constituição Nacional
- 153 Governo da Nação

160 Contexto econômico

- 161 Meio Econômico
- 162 Mercado de Trabalho
- 163 Direito do Trabalho

170 Contexto mundial

- 171 Estrutura Política Mundial
- 172 Relações Internacionais
- 173 Problemática Mundial
- 174 Política Mundial

200 ESCOLA

210 Pesquisa da educação

- 211 Realidade da Educação
- 212 Informação e Educação
- 213 Educação Comparada

220 Estatística da educação

- 221 Estatística
- 222 Estatísticas do Contexto Educacional
- 223 Estatísticas Educacionais

- 224 Estatísticas Gerenciais
- 225 Estatísticas dos Profissionais da Educação
- 226 Indicadores Econômicos em Educação

230 Política da educação

- 231 Política
- 232 Educação como Ação Política
- 233 Política Mundial da Educação
- 234 Direito Educacional

240 Administração da Educação

- 241 Administração
- 242 Administração da Educação Escolar
- 243 Planejamento da Educação
- 244 Produtividade e Avaliação da Educação

250 Educandos

- 251 Tipos de Educandos
- 252 Vida do Educando
- 253 Assistência ao Educando

260 Profissionais da Educação

- 261 Administração de Recursos Humanos
- 262 Funções Docentes
- 263 Funções Não-Docentes
- 264 Magistério

270 Instituições de ensino

- 271 Natureza Jurídica das Instituições
- 272 Escolas
- 273 Instituições de Educação Básica
- 274 Escolas para Formação Profissional
- 275 Instituições de Ensino Superior
- 276 Instituições de Educação Especial
- 277 Centros de Educação de Adultos

280 Administração escolar

- 281 Administração das Instituições de Ensino
- 282 Gestão Administrativa da Escola
- 283 Gestão Pedagógica da Escola

284 Organizações Escolares

285 Registro Escolar

290 Economia da educação

291 Economia

292 Financiamento da Educação

293 Administração Financeira

294 Infraestrutura Escolar

300 FUNDAMENTOS DA EDUCAÇÃO

310 Cultura e educação

311 Cultura

312 Cultura e Religião

313 História das Civilizações

320 Filosofia e educação

321 Filosofia

322 Processo Filosófico

323 Filosofia Fundamental

324 Filosofia Prática

325 Filosofia das Ciências

330 Trabalho e educação

331 Trabalho Humano

332 Psicologia do Trabalho

333 Relação Educação/Trabalho

334 Administração do Trabalho

340 Psicologia e educação

341 Psicologia

342 Psicologia da Personalidade

343 Psicologia do Desenvolvimento

344 Psicologia Social

345 Psicologia Cognitiva

346 Psicologia da Dinâmica

347 Psicologia da Educação

350 Sociologia e educação

351 Sociologia

- 352 Estrutura Social
- 353 Processo de Interação Social
- 354 Sociologia da Educação

360 Comunicação e educação

- 361 Comunicação
- 362 Linguagem e Educação

370 Saúde e educação

- 371 Biologia Educacional
- 372 Excepcionalidade
- 373 Medicina e Educação
- 374 Psicologia Clínica
- 375 Educação Física

400 EDUCAÇÃO

410 História da educação

- 411 História Geral da Educação
- 412 História da Educação Brasileira

420 Filosofia da educação

- 421 Princípios da Educação
- 422 Crítica da Educação
- 423 Pedagogia e Educação
- 424 Finalidades da Educação

430 Educação escolar

- 431 Escolarização
- 432 Educação Escolar Básica
- 433 Educação Profissional
- 434 Educação Superior
- 435 Formação dos Profissionais da Educação

440 Modalidades de educação

- 441 Educação Formal
- 442 Educação Não-Formal
- 443 Educação Informal
- 444 Educação Especial
- 445 Educação de Jovens e Adultos

- 446 Educação a Distância
- 447 Modalidades de Educação Profissional
- 448 Educação de Grupos Atípicos

450 Curso e Currículo

- 451 Currículo
- 452 Curso
- 453 Plano de Ensino
- 454 Conteúdos Curriculares
- 455 Atividades Curriculares
- 456 Avaliação do Currículo

460 Processo de ensino-aprendizagem

- 461 Aprendizagem
- 462 Orientação Educacional
- 463 Ensino
- 464 Orientação Pedagógica
- 465 Organização do Trabalho Intelectual

470 Produtividade e Avaliação Escolar

- 471 Produtividade escolar
- 472 Avaliação Escolar

480 Meios de ensino

- 481 Tecnologia Educacional
- 482 Equipamentos Didáticos
- 483 Material Didático

900 IDENTIFICADORES E ESPECIFICADORES DE INFORMAÇÃO

910 Identificadores de denominação

- 911 Personalidades
- 912 Entidades
- 913 Lugares
- 914 Indicadores de Tempo
- 915 Grupos Humanos

920 Identificadores de conteúdo

- 921 Áreas do Conhecimento
- 922 Cursos

923 Profissões e Ocupações

924 Atividades Planejadas

925 Manifestações Socioculturais

930 Especificadores de forma de conteúdo

931 Documentos Não Oficiais

932 Documentos Oficiais

940 Especificadores de suporte de informação

941 Língua do Documento

942 Publicações

943 Publicações em Multimeios

5 ANÁLISE DO DADOS: O TERMO ESTUPRO NA LINGUAGEM DOCUMENTARIA x LINGUAGEM NATURAL

A representação da informação se dá em diferentes tipos de instrumentos e linguagens. Neste trabalho é feita comparação entre a LN utilizada na Rede Social Twitter com o uso das hashtags e com a LD, através de instrumentos de controle, mais precisamente dois tesouros, o do STJ e o Brased.

Atualmente, os tesouros são importantes instrumentos para os sistemas informatizados, tendo por objetivo principal o controle da polissemia, isto é, da pluralidade de sentido de determinadas palavras dentro de um contexto na LN, fazendo com que uma palavra que pode eventualmente assumir diversos significados passe a ter só um. Este controle terminológico diminui a polissemia existente na LN, fazendo com que indexadores e usuários consigam resultados mais eficazes.

5.1 LINGUAGEM NATURAL: #NÃO MEREÇOSERESTUPRADA

Diante do resultado da pesquisa do IPEA sobre a tolerância social à violência contra as mulheres (resultado que depois se mostrou errôneo), houve grande mobilização e repercussão no País, sobretudo nas Redes Sociais, onde rapidamente a jornalista Nana Queiroz deu início a uma campanha de repúdio contra o resultado da pesquisa com a hashtag #NãoMereçoSerEstuprada. Embora os temas centrais dessa polêmica gerada a partir do resultado da pesquisa do IPEA representada pela hashtag #NãoMereçoSerEstuprada tenha sido, a violência contra a mulher e o estupro, nas Redes Sociais pouco se viu hashtags com essas palavras.

Em LN, qualquer palavra, ou até mesmo conjunto de palavras podem ser utilizadas como hashtag, e estas facilitam a recuperação de informações em sistemas de busca como o Twitter Search. De modo geral, para facilitar esse processo de recuperação costuma-se usar etiquetas curtas e que sejam descritivas do assunto em questão, no entanto, etiquetas maiores são admitidas pelo sistema, como exemplo, de uma hashtag curta temos o termo #estupro, e, de uma hashtag mais longa temos o termo #NãoMereçoSerEstuprada. Vale acrescentar que o único limite para o tamanho das hashtags no Twitter são os 140 caracteres admitidos pelo sistema.

Em LN, no que diz respeito às Redes Sociais, os termos que representam determinado assunto podem variar muito, pois não há um compromisso por parte dos usuários na manutenção de um termo específico. No caso do #NãoMereçoSerEstuprada, outros termos em forma de hashtags aparecem nesse mesmo contexto, como por exemplo: #Sexismo, #Respeito, #MeuCorpoMinhasRegras, #Aculpanaoedavitima e #Respeiteasmulheres.

Estas palavras-chave fazem todo sentido dentro do contexto do #NãoMereçoSerEstuprada, no entanto, quando se faz a busca por essas hashtags os resultados são totalmente aleatórios, com sentido totalmente diferente, ou seja, ao contrário da LD que visa a redução dos significados em busca de uma maior especificidade, no caso das hashtags em LN os sentidos se perdem em um universo ilimitado.

Cintra (1983) aponta que a redução dos significados, por ser uma operação de corte num universo praticamente ilimitado, poderá se valer de alguns conceitos correntes em LN. Dentre eles a polissemia, a homonímia, a sinonímia e a antonímia. De fato, no âmbito da LN, as LDs apresentam como peculiaridade a existência de um vocabulário próprio que se caracteriza pela maior precisão de seus termos. Em geral a palavra, neste uso, tem um significado unívoco.

No caso da hashtag #NãoMereçoSerEstuprada, percebe-se o significado em torno da representação dessa marcação, que foi ocasionada pela pesquisa do IPEA (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a), porém em relação a essa representação nenhuma vinculação é feita a referida pesquisa realizada, nem com outros termos relacionados a questão da violência contra as mulheres.

5.2 LINGUAGEM DOCUMENTÁRIA: O TERMO ESTUPRO NOS TESAUROS DO STJ E BRASED

De fato a Web 2.0 e as Redes Sociais permitem hoje o uso de linguagens mais flexíveis e de padrões cada vez mais aceitos para representação da informação, e de acordo com Campos (2001) todo movimento existente nos sistemas tem por objetivo possibilitar ao usuário acesso a informação, e esse deve ser o principal objetivo do profissional da informação.

A LN pode ser entendida como sinônimo de discurso comum, pois é a linguagem usada habitualmente na fala e na escrita, porém em base de dados pode corresponder aos termos do título e resumo dos documentos. Os termos que constituem o vocabulário de uma

LD podem ser extraídos dos próprios documentos, na linguagem do autor em LN. Como principal característica a LD é utilizada para substituição do texto do documento em LN por sua descrição abreviada, utilizada como um artifício para recuperar o que é essencial no documento, ou seja, o tema central do documento.

Dessa forma, observa-se que o controle do vocabulário é um elemento essencial para a organização da informação, pois é uma linguagem artificial produzida com o objetivo de sistematizar o vocabulário do usuário para um vocabulário padronizado, e o mesmo não ocorre na LN.

Em LD os mesmos termos podem assumir diferentes conceitos dependendo da área de conhecimento em que está inserido. No caso dessa pesquisa, o termo estupro tem o mesmo conceito nos dois tesouros consultados, no entanto há uma diferença no contexto em que se apresentam.

No primeiro tesouro consultado, o do STJ, o termo apresenta uma relação hierárquica que exprime os graus de superordenação e subordinação entre os conceitos. O termo superordenado (TG - termo genérico) representa o conceito mais abrangente, do qual o termo subordinado (TE - termo específico).

O termo se apresenta dentro do contexto jurídico e dispõe de uma relação hierárquica de polierarquia, e este tipo de relação permite que um termo específico, como no caso do termo estupro, possua uma relação direta com mais de um termo genérico (TG) como o “crime contra a liberdade sexual” e “crime hediondo” e possui dois termos específicos (TE): o estupro presumido e o estupro de vulnerável.

Já no tesouro do Brased, pôde-se observar que dentro da hierarquização dos termos, de acordo com sua matriz conceitual, que ele é composto por quatro campos, sendo estes: contexto da educação, escola como instituição social, fundamentos da educação e educação. Nessas divisões, ou subdivisões, o termo estupro está estrutura de forma mais genérica no campo do “contexto da educação” e este vai se subdividindo em campos cada vez mais específicos até chegarmos no termo estupro, ou seja, o referido termo está dentro do contexto da educação, sendo este o termo mais genérico da cadeia, passando pelo contexto social, condições sociais, problemas sociais, criminalidade, crime, delitos sexuais até chegar ao termo estupro, sendo este o mais específico como vemos abaixo:

Contexto da educação

Contexto social

Condições Sociais

Problemas sociais

Criminalidade

Crime

Delitos Sexuais

ESTUPRO

De acordo Lancaster (1979 apud LOPES, 2002, p. 46) as LDs fornecem muita ajuda aos usuários que executam a busca, pois “possuem o controle de sinônimos e quase sinônimos, a separação dos homógrafos, o uso da pré-coordenação para evitar falsas coordenações e relações incorretas entre os termos e, ainda, a ligação de termos relacionados, todos com suas respectivas hierarquias”. Dessa forma, facilita o processo de recuperação.

No caso dos tesouros outros termos poderiam ser utilizados para representação da pesquisa do IPEA além do estupro, como por exemplo, “Violência Sexual”. Tanto no tesouro do STJ, quanto no Brased, este termo aparece como um termo relacionado. Nota-se que um termo na LD não existe por si só, isolado, ele sempre estará contextualizado com outro termo, no entanto na LN observada no Twitter, através das hashtags, os termos tagueados encontram-se isolados uns dos outros, mesmo quando estão dentro de um Tuite, pois ao clicar nas hashtags, cada uma direcionará para um resultado diferente, em contextos diferentes ou até contrários, ou seja, a recuperação do #NãoMereçoSerEstuprada só se dará em um determinado contexto e o resultado obtido na busca poderá não ser satisfatório ao usuário que busca informação.

Depois de alguns meses do resultado da pesquisa do IPEA e toda mobilização nas Redes Sociais foi feita nova pesquisa pela hashtag #NãoMereçoSerEstuprada no Twitter (19/11/2014) para averiguar como está a recuperação da informação contida na hashtag.

No resultado da busca constatou-se que ainda é possível recuperar a informação, no entanto percebeu-se que após alguns meses a campanha #NãoMereçoSerEstuprada perdeu força e foco e a hashtag supracitada é também utilizada para brincadeiras e deboches das mais diversas situações, como no caso noticiado em diversos veículos de informação, onde focas estariam abusando sexualmente de pinguins-rei na Ilha Marion, localizada no Atlântico Sul, como observa-se na imagem abaixo:

Imagem 12 – Internautas fazem piada com as focas ‘estupradoras’ de pinguins



Fonte: O Globo (2014)

Diversos internautas fizeram uma analogia entre o estupro dos animais e a campanha #NãoMereçoSerEstuprada que trata da violência contra a mulher, e usando no contexto do mundo animal, as mesmas desculpas esfarrapadas utilizadas para justificar a agressão às mulheres.

A LN, sobretudo nas Redes Sociais, é constituída por termos representados pelas hashtags e tais termos estão a um click de serem acessadas e os usuários podem interagir diretamente com os itens incluídos nessa representação, entretanto nem sempre os resultados obtidos são satisfatórios. Em contrapartida, na LD, o bibliotecário/indexador exerce um papel de mediador entre o usuário e a informação traduzindo o documento em termos que possam ser recuperados de forma mais precisa e rápida, poupando o tempo do usuário.

O uso da LN com certeza tem muitas dificuldades a serem superadas, e embora as LDs também apresentam algumas desvantagens, pois nem sempre os termos preferidos pelos indexadores são os termos utilizados pelos usuários em suas buscas, chega-se a conclusão que o uso das LDs em sistemas de busca e recuperação da informação é um instrumento confiável para garantir a recuperação da informação.

6 CONCLUSÃO

Com a crescente expansão e uso das redes sociais e dos serviços de informação e comunicação na Web faz-se necessário ao bibliotecário estar atento ao que acontece nessas redes e buscar estratégias para o tratamento da informação que nela circula.

Em LN, qualquer palavra, ou até mesmo conjunto de palavras podem ser utilizadas como hashtag, no entanto, este fato pode ser considerado fator dispersivo da informação, pois a mesma palavra pode circular de forma desordenada devido a ausência de controle do vocabulário. Sendo assim, cada pessoa pode representar da maneira que convier determinado assunto através do uso da hashtag e devido a falta de conhecimento técnico por parte dos usuários, a informação não é padronizada.

Em LN podem ser observadas algumas vantagens como, por exemplo: permitir o imediato registro da informação sem necessidade de consulta a uma linguagem de controle, o processo de busca pode ser facilitado, pois não há necessidades de treinamentos específicos como no uso de uma linguagem controlada e os termos são definidos pelos próprios usuários. Entretanto, conforme constatado nessa pesquisa, a LN através do uso da hashtag não padroniza a representação e, por conseguinte, influencia a recuperação.

As LDs apresentam um vocabulário próprio que se caracteriza pela maior precisão de seus termos, em geral a palavra ou termo tem um significado unívoco, e o mesmo não acontece com a LN. Dessa forma torna-se essencial o controle do vocabulário para a organização da informação, pois é uma linguagem artificial produzida com o objetivo sistematizar o vocabulário do usuário para um vocabulário padronizado.

Como exemplos de vantagens das LDs é possível citar o controle total do vocabulário de indexação, minimizando os problemas de comunicação entre indexadores e usuários, as relações hierárquicas e as remissivas do vocabulário controlado auxiliam tanto o indexador, quanto o usuário na identificação de conceitos relacionados e a redução no tempo de consulta para o usuário, pois a estratégia de busca será mais bem elaborada com o uso do tesouro.

Diante disso acredita-se que a linguagem natural não deve ser utilizada para organizar a informação efetivamente, pois a sinonímia e homonímia podem causar dispersão e desordem.

Tanto na LN, quanto na LD, se faz necessário dar condições favoráveis à busca do usuário. Podemos afirmar que, de modo geral, essa proposta de comparação tem o intuito de demonstrar que a LN tanto é capaz de facilitar o registro imediato de determinada informação, sobretudo nas Redes Sociais, dispensando uma pesquisa mais específica e agilizando o

processo de comunicação, quanto dificulta, pois sem a devida padronização certamente ocorrerá um alto índice de respostas negativas às buscas do usuário. Nas LDs, o índice de acertos na busca do usuário será satisfatório, pois trata de uma estrutura hierárquica mais definida e contextualizada.

Por fim, compreendemos que é importante incentivar pesquisas nesse contexto, pois o uso de linguagens de marcação são crescente nas redes sociais. Sendo assim, indica-se que a temática faça parte das agendas de pesquisa de bibliotecários e demais interessados na representação temática da informação.

REFERÊNCIAS

ARAÚJO, Ana Luíza Custódio de. **Vocabulário controlado básico**: evolução histórica. 2011. 57 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília. Disponível em: <<http://portal2.tcu.gov.br/portal/pls/portal/docs/2241055.PDF>>. Acesso em: 5 maio 2014.

BORBA, Diego do Santos; LAAN, Regina Helena Van der; CHINI, Bernadete Ros. Palavras-chave: convergências e diferenciações entre a linguagem natural e a terminologia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 2, p. 26-36, abr./jun. 2012. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000011800&dd1=13781>>. Acesso em: 8 maio 2014.

BRASIL. **ONU Mulheres**: Entidade das Nações Unidas para a Igualdade de Gênero e o Empoderamento das Mulheres. Disponível em: <<http://www.onumulheres.org.br/>>. Acesso em: 5 abr. 2014.

BRESSER, Deborah. Valendo! **#Naomerecoserestuprada**. Disponível em: <<http://entretenimento.r7.com/blogs/blog-da-db/valendo-naomerecoserestuprada-20140328/>> Acesso em: 30 mar. 2014.

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

CATARINO, Maria Elisabete; BAPTISTA, Ana Alice. Folksonomias: características das etiquetas na descrição de recursos da Web. **Inf. Inf.**, Londrina, v. 14, n. esp. 2009. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000007944&dd1=4c3bd>>. Acesso em: 24 abr. 2014.

CERQUEIRA, Daniel. **Organizadora de campanha contra estupro recebe ameaças na Web**: pesquisa constatou que a maior parte dos brasileiros acredita que as mulheres são responsáveis por sofrerem abusos sexuais: Depoimento em 31/03/2014. Entrevista concedida ao programa Fantástico da Rede Globo de Televisão, 2014. Disponível em: <<http://g1.globo.com/fantastico/noticia/2014/03/organizadora-de-campanha-contr-estupro-recebe-ameacas-na-Web.html>>. Acesso em: 5 abr. 2014.

CENTRO FEMININO DE ESTUDOS E ASSESSORIA. Disponível em: <<http://www.cfemea.org.br/>>. Acesso em: 5 abr. 2014.

CINTRA, Anna Maria Marques. Elementos de lingüística para estudos de indexação. **Ciência da Informação**, Brasília, v. 12, n. 1, 1983. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000002663&dd1=546be>>. Acesso em: 8 mai. 2014.

CURRÁS, Emília. **Tesouro, linguagens terminológicas**. Brasília: IBICT, 1995. Disponível em: <<http://livroaberto.ibict.br/handle/1/454>>. Acesso em 07 dez. 2014.

_____ et al. **Para entender as linguagens documentárias**. 2. ed. rev. e ampl. São Paulo: Polis, 2002. 92 p.

DAHLBERG, Ingetraut . Teoria do conceito. **Ciência da Informação**, v. 7, n. 2, p. 101-107, 1978. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/1680/1286>>. Acesso em: 5 maio 2014.

DANTON, Gian. **Metodologia científica**. Pará de Minas: Virtual Books Online M&M Editores, 2002. Disponível em: <[DISCOVER TWITTER. Disponível em: <<https://discover.Twitter.com/pt/learn-more>>. Acesso em: 3 maio 2014.](http://minhateca.com.br/clodomar/Baixar+Arquivos/LIVROS+PDF/Metodologia+e+produ*c3*a7*c3*a3o+cient*c3*adfica/Metodologia+Cient_fica+--+Gian+Danton,52568975.pdf.>>. Acesso em: 5 nov. 2014.</p>
</div>
<div data-bbox=)

DODEBEI, V. L. D. L. **Tesouro: linguagem de representação da memória documentária**. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. **Datagramazero**, Rio de Janeiro, v. 5, n. 4, 2004. Disponível em: <http://www.datagramazero.org.br/ago04/Art_01.htm>. Acesso em: 8 maio 2014.

GUIMARÃES, Eduardo. **Os estudos sobre linguagens: uma história das ideias**. Disponível em: <<http://www.comciencia.br/reportagens/linguagem/ling14.htm>>. Acesso em: 15 abr. 2014.

ILHEN, Junia de; ZAMORA, Maria Helena. Além do ato: os transbordamentos do estupro. **Revista Rio de Janeiro**, n. 12, jan./abr. 2004. Disponível em: <http://www.forumrio.uerj.br/documentos/revista_12/12_dossie_JuniaVilhena.pdf>. Acesso em: 2 maio 2014.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **SIPS - Sistema de indicadores de Percepção Social:** tolerância social à violência contra as mulheres. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/SIPS/140327_sips_violencia_mulheres_antigo.pdf>. Acesso em: 30 mar. 2014a. (versão original).

_____. **SIPS- Sistema de indicadores de Percepção Social:** tolerância social à violência contra as mulheres. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/SIPS/140327_sips_violencia_mulheres_novo.pdf>. Acesso em: 30 mar. 2014b. (versão após a errata).

_____. **Errata da pesquisa:** tolerância social à violência contra as mulheres. Portal IPEA, 04/04/2014. Disponível em: <http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=21971&catid=10&Itemid=9>. Acesso em: 5 maio 2014. (Errata).

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. Disponível em: <<http://portal.inep.gov.br/pesquisa-thesaurus>>. Acesso em: 3 mai. 2014.

KAUARK, F. S.; MANHÃES, F. C.; MEDEIROS, C. H. **Metodologia da pesquisa:** um guia prático. Itabuna: Via Litterarum, 2010.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos da metodologia científica.** 5. ed. São Paulo: Atlas, 2003.

LANCASTER, F.W. **Indexação e resumo:** teoria e prática. Brasília: Brinquet de Lemos, 1993.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão de literatura. **Ciência da Informação**, Brasília, v. 31, n. 1, jan./abr. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n1/a05v31n1.pdf>>. Acesso em: 7 maio 2014.

MARRONI, Gilza Núria Brandão. **Identificação e delimitação de relações associativas em tesouros:** um estudo de caso na área do direito do trabalho. 2006. 127 f. Tese (Mestrado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2006.

MOREIRA, Manoel Palhares. **Ambiente para geração e manutenção semi-automática de tesouros.** Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005. Disponível em:

<http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/VALA-6KHJPX/doutorado_manuel_palhares_moreira.pdf?sequence=1>. Acesso em: 5 maio 2014.

_____; MOURA, Maria Aparecida. Construindo tesouros a partir de tesouros existentes: a experiência do TCI - Tesouro em Ciência da Informação. **DataGramZero**, v. 7, n. 4, ago. 2006. Disponível em:
<http://www.dgz.org.br/ago06/Art_01.htm>. Acesso em: 7 maio 2014.

MOURA, Maria Aparecida. Folksonomias, redes sociais e a formação para o tagging literacy: desafios para a organização da informação em ambientes colaborativos virtuais. **Informação & Informação**, v. 14, n. esp., 2009. Disponível em:
<<http://www.brapci.ufpr.br/documento.php?dd0=0000007945&dd1=eb2be>>. Acesso em: 24 abr. 2014.

NARVAZ, Martha Giudice; KOLLER, Sílvia Helena. Famílias e patriarcado: da prescrição normativa à subversão criativa. **Psicol. Soc.**, Porto Alegre, v. 18, n. 1, jan./abr. 2006. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-71822006000100007> Acesso em: 1 maio 2014.

NAVARRO, Sandrelei. Interface entre lingüística e indexação: revisão de literatura. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, jan./jun. 1988.

NOVELLINO, Maria Salet Ferreira. A linguagem como meio de representação ou de comunicação da informação. **Perspect. Cienc. Inf.**, Belo Horizonte, v. 3, n. 2, p. 137-146, jul./dez. 1998. Disponível em:
<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/viewFile/602/371>>. Acesso em: 5 maio 2014.

_____. Instrumentos e metodologias de Representação da informação. **Inf. Inf.**, Londrina, v. 1, n. 2, p. 37-45, jul./dez. 1996.

O GLOBO. **Internautas fazem piada com as focas ‘estupradoras’ de pinguins do Atlântico Sul**. Disponível em: <<http://oglobo.globo.com/sociedade/tecnologia/internautas-fazem-piada-com-as-focas-estupradoras-de-pinguins-do-atlantico-sul-14590817>>. Acesso em: 19 de nov. 2014

OLIVEIRA, Karine Rios de; LEITE, Thiago André Rodrigues. Ferdinand de Saussure e Émile Benveniste: questões referentes à língua, à fala e à criação linguística. **Revista Digital Inventário**, out. 2011. Disponível em:
<<http://www.inventario.ufba.br/09/9/FERDINAND%20DE%20SAUSSURE%20E%20C3%20%89MILE%20BENVENISTE.pdf>>. Acesso em: 5 maio 2014.

PACHECO, Flávio.

Estou Fazendo Um Teste Para Descobrir Se Existe Algum Limite De Palavras Para a Formação De "hashtags No Twitter De Modo a Acrescentar Esse Conhecimento Ao Meu TCC. Rio de Janeiro, 19/05/2014, Twitter @flaviopacheco. Disponível em: <<https://Twitter.com/flaviopacheco>>. Acesso em: 19 maio 2014.

PATU, Gustavo. Análise: erro em pesquisa do Ipea joga luz sobre politização do instituto. **Folha de São Paulo**, Cotidiano, Brasília, 2014. Disponível em: <<http://www1.folha.uol.com.br/cotidiano/2014/04/1436292-analise-erro-do-ipea-joga-luz-sobre-politizacao-do-instituto.shtml>>. Acesso em 11 maio 2014.

PEREIRA, Débora de Carvalho; CRUZ, Ruleandson do Carmo. Folksonomia e tags afetivas: comunicação e comportamento informacional no Twitter. **Data Grama Zero**, v. 11, n. 6, dez. 2010. Disponível em: <http://www.dgz.org.br/dez10/Art_06.htm>. Acesso em: 24 abr. 2014.

PRIMO, A. O aspecto relacional das interações na “Web” 2.0. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 29., 2006, Brasília. **Anais eletrônicos...** Brasília: Intercom; Universidade de Brasília, 2006. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/1264>>. Acesso em: 5 maio 2014.

_____. **As tags no Twitter como informação contextual de afeto.** Dossiê Alex Primo (blog científico). 2010. Disponível em: <http://alexprimo.com/2010/03/09/as_tags_no_Twitter_como_informacao_conte/>. Acesso em: 5 maio 2014.

QUIVY, Raymond; CAMPENHOUDT, Luc Van. **Manual de investigação em ciências sociais.** Lisboa: Gradiva, 2005.

RECUERO, Raquel. **Redes sociais na internet.** Porto Alegre: Sulina, 2009. Disponível em: <<http://www.ichca.ufal.br/graduacao/biblioteconomia/v1/wp-content/uploads/redessociaisnainternetrecuero.pdf>>. Acesso em: 27 dez. 2013.

ROUSSEFF, Dilma. **A jornalista @nanaqueiroz se indignou c/ os dados da pesquisa do @ipeaonline sobre o machismo na nossa sociedade #respeiteasmulheres.** Brasília, 31 mar. 2014, Twitter @dilmabr. Disponível em: <<https://Twitter.com/dilmabr>>. Acesso em: 22 abr. 2014a.

_____. **Por ter se manifestado nas redes contra a cultura de violência contra a mulher, a jornalista foi ameaçada de estupro.** Brasília, 31 mar. 2014, Twitter @dilmabr. Disponível em: <<https://Twitter.com/dilmabr>>. Acesso em: 22 abr. 2014b.

_____. **Organizadora do protesto #NãoMereçoSerEstuprada, @nanaqueiroz merece toda a minha solidariedade e #respeito.** Brasília, 31 mar. 2014, Twitter @dilmabr. Disponível em: <<https://Twitter.com/dilmabr>>. Acesso em: 22 abr. 2014c.

_____. **Nenhuma mulher merece ser vítima de violência, seja física ou sob a forma de ameaça #respeiteasmulheres.** Brasília, 31 mar. 2014, Twitter @dilmabr. Disponível em: <<https://Twitter.com/dilmabr>>. Acesso em: 22 abr. 2014d.

_____. **O governo e a lei estão do lado de @nanaqueiroz e das mulheres ameaçadas ou vítimas de violência #respeiteasmulheres.** Brasília, 31 mar. 2014, Twitter @dilmabr. Disponível em: <<https://Twitter.com/dilmabr>>. Acesso em: 22 abr. 2014e.
TOMAÉL, Maria Inês; ALCARÁ, Adriana Rosecler; CHIARA, Ivone Guerreiro Di. Das redes sociais à inovação. **Ciência da Informação**, Brasília, v. 34, n. 2, 2005. Disponível em: <<http://www.scielo.br/pdf/ci/v34n2/28559.pdf>>. Acesso em: 27 dez. 2013.
SEXISMO: In: **Dicionário Preberam da Língua Portuguesa**, [em linha] 2008-2014). Disponível em: <<http://www.priberam.pt/dlpo/sexismo>>. Acesso em: 1 mai. 2014.

SOUSA, Brisa Pozzi de; FUJITA, Mariângela Spotti Lopes. Análise de assunto no processo de indexação: um percurso entre teoria e norma. **Inf. & Soc.: Est.**, João Pessoa, v. 24, n. 1, p. 19-34, jan./abr. 2014. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/16281>>. Acesso em: 28 out. 2014.

SUPERIOR TRIBUNAL DE JUSTIÇA. Disponível em: <<http://www.stj.jus.br/portal/site/STJ>>. Acesso em: 3 mai. 2014.

_____. **Thesaurus.** Disponível em: <<http://www.stj.jus.br/SCON/thesaurus/>>. Acesso em: 3 mai. 2014.

TUDO SOBRE REDES SOCIAIS. **Trending Topics.** Disponível em: <<http://www.tudosobreredesociais.com.br/2011/08/03/o-que-sao-trending-topics/>>. Acesso em: 3 mai. 2014.

TWITTER. **Blog.** Disponível em: <<https://blog.Twitter.com/2010/trend-or-not-trend>>. Acesso em: 3 mai. 2014.

_____. Disponível em: <<https://Twitter.com/>>. Acesso em: 3 abr. 2014.

_____. **Marcadores.** Disponível em: <<https://support.Twitter.com/articles/255508-o-que-sao-os-marcadores-simbolos-de#>>. Acesso em: 3 maio 2014.

WELLS, Miriam. **#Brazil women revolt against violent #sexism - my piece for @thesundaytimes**
http://www.thesundaytimes.co.uk/sto/news/world_news/article1396764.ece...
#NaoMerecoSerEstuprada @nanaqueiroz. Rio de Janeiro, 8 abr. 2014. Twitter @missmbc.
Disponível em: <<https://Twitter.com/missmbc>>. Acesso em: 22 abr. 2014.

WIKIPEDIA. **“hashtag”**. Disponível em: <[http://pt.wikipedia.org/wiki/”hashtag](http://pt.wikipedia.org/wiki/”hashtag”)>. Acesso em: 7 abr. 2014.