

Linguagem Natural no Twitter e Linguagem Documentária em Tesouros: da hashtag #NãoMereçoSerEstuprada ao descritor estupro

Natural Language Twitter and Documentary Language in Thesaurus: the hashtag #NãoMereçoSerEstuprada the descriptor rape

Brisa Pozzi de Sousa

Doutoranda em Ciência da Informação pela Universidade Federal de Minas Gerais – UFMG
Professora Assistente do Departamento de Estudos e Processos Biblioteconômicos da Universidade Federal do Estado do Rio de Janeiro – Unirio.
E-mail: brisapozzi@gmail.com

Flávio Pacheco da Silva

Bacharel em Biblioteconomia pela Universidade Federal do Estado do Rio de Janeiro – Unirio.
E-mail: flavio.pacheco@yahoo.com.br

Resumo

Aponta a linguagem natural no Twitter a partir da hashtag #NãoMereçoSerEstuprada e realiza comparação entre a linguagem documentária de dois tesouros pelo descritor estupro, sendo o Thesaurus Brasileiro da Educação (Brased) e o Tesouro Jurídico do Superior Tribunal de Justiça (STJ). A pesquisa é descritiva e bibliográfica, de cunho qualitativo, embasada na investigação teórica. Discorre sobre vantagens e desvantagens no uso dessas linguagens e demonstra que na natural qualquer palavra, ou até mesmo conjunto de palavras podem ser utilizadas como hashtag para representar qualquer assunto. No entanto, observa-se que o controle do vocabulário é um elemento essencial para a representação de assuntos, pois é uma linguagem artificial produzida com o objetivo de sistematizar o vocabulário controlando, como por exemplo, a polissemia. Os resultados demonstram que a linguagem natural no Twitter não padroniza a representação do assunto sendo considerada fator dispersivo, ao contrário das linguagens documentárias que apresentam vocabulário caracterizado pela precisão dos descritores.

Palavras-chave: Linguagem natural. Linguagem documentária. Twitter - hashtag. Descritor estupro. Tesouro Brased. Tesouro STJ.

Abstract

Points natural language on Twitter from the hashtag # NãoMereçoSerEstuprada and performs comparison between the two indexing language thesaurus descriptor for rape, and the Brazilian Thesaurus of Education (Brased) and the Legal Thesaurus Superior Court of Justice (STJ). The research is descriptive and bibliographical, of qualitative nature, based on theoretical research. It discusses advantages and disadvantages of using these languages in natural and demonstrates that any word, or even whole words may be used as hash tag to represent any subject. However, it is observed that the control of vocabulary is an essential element for the representation issue, it is an artificial language produced in order to systematize the vocabulary control, such as polysemy. The results show that the natural language Twitter does not standardize the representation of the subject under consideration dispersive factor, unlike the documentary language vocabulary characterized by the feature that the accuracy of descriptors.

Keywords: Natural language. Documentary language. Twitter - hashtag. Descriptor rape. Thesaurus Brased. Thesaurus STJ.

Introdução

Neste artigo serão apresentados os conceitos da Linguagem Natural (LN) e seus desdobramentos na Rede Social Twitter, através da marcação denominada *hashtag*, e realizar-se-á comparação com a Linguagem Documentária (LD) e apresentação hierárquica do descritor estupro em dois diferentes contextos documentários: no Thesaurus Brasileiro de Educação (Brased) e no Tesouro do Superior Tribunal de Justiça (STJ).

A motivação da escolha do descritor partiu do resultado da pesquisa do IPEA sobre tolerância social à violência contra as mulheres (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a), pois, 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 a representação da *hashtag* #NãoMereçoSerEstuprada.

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 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 Topics*¹ do Twitter. A mobilização foi muito rápida, entretanto não é objetivo do artigo a discussão do resultado da pesquisa, mas sim delinear análise sobre a representação do assunto estupro, tanto em LN, precisamente na Rede Social Twitter, quanto na LD, com uso de dois tesouros.

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 linguagens documentária e natural. Portanto, o objetivo baseia-se na investigação da representação do descritor estupro e o resultado com uso de dois instrumentos de controle de vocabulário, o Brased e o Tesouro do STJ, e suas possíveis relações com a LN empregada no Twitter pelo uso da *hashtag*.

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, 2010).

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 (COELHO, 2011).

A escolha pelos dois tesouros decorre da possível configuração jurídica e educacional que o resultado da pesquisa do IPEA pode suscitar discussões e, assim, a representação de assuntos. Não se descartam outras áreas e tão pouco constitui objetivo esgotar abordagens de análise, porém nos domínios dos tesouros selecionados foi possível encontrar modelagem que atende a proposta de investigação. Em consonância com as questões descritas, pretende-se demonstrar as relações que aproximam ou distanciam esses dois tipos de linguagens - LN e LD - especificamente no que tange o descritor estupro.

Dessa forma, após a introdução apresenta-se a contextualização da pesquisa do IPEA sobre tolerância social à violência contra as mulheres. Em seguida, dispõe-se a metodologia e o referencial teórico e, posteriormente, a análise dos dados coletados entre as linguagens documentárias através de dois tesouros selecionados e a LN utilizada na rede social Twitter. Por fim, segue o delineamento e as constatações que não visam minimizar as discussões, mas sim despertar o interesse pelo contexto levantado.

Contextualização da análise: a pesquisa do IPEA tolerância social à violência contra as mulheres

Pesquisa divulgada no início do ano de 2014 pelo Instituto de Pesquisa Econômica Aplicada (IPEA), na data de 27 de março de 2014, sobre 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 este foi o estímulo para a pesquisa realizada no Brasil (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a).

Sendo assim, um grupo de trabalho foi formado nacionalmente 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. No caso da pesquisa realizada pelo IPEA foi feito 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 para 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 que age com violência contra sua esposa (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a).

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 reduzida 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.

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 (SEXISMO, c2013).

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 seus direitos sobre a mulher não são mais irrestritos e excluem formas abertas e extremas de violência (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014a).

O coordenador da pesquisa declarou em entrevista a um site de programa televisivo que a principal conclusão diante dos 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 [...] centenas de vítimas simplesmente não vão prestar queixa à polícia, porque elas vão achar que [...] fizeram alguma coisa, que facilitaram e vão ser mal vista na sociedade. (ORGANIZADORA..., 2014).

Em face da 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. A seguir é traçado panorama de como se deu a repercussão e, após, o embasamento do contexto para nossa investigação.

A Repercussão: Mulher Nenhuma Merece Ser Estuprada

A partir do polêmico resultado, a jornalista Nana Queiroz realizou campanha de repúdio nas Redes Sociais com a *hashtag*⁴ (representada pelo símbolo #, de nome cerquilha) seguida da palavra-chave #NãoMereçoSerEstuprada.

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. A foto feita em Brasília traz ao fundo o Planalto Central (BRESSER, 2014; ORGANIZADORA..., 2014).

Considera-se que a partir desse fato houve grande adesão e manifestação da sociedade, principalmente nas Redes Sociais, e a *hashtag* #NãoMereçoSerEstuprada ganhou força. Em

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, c2014a).

pouco tempo já estava entre os *Trending Topics* do Twitter (COELHO, 2011; ORGANIZADORA..., 2014).

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, 'Web'sites e aplicativos. (DISCOVER TWITTER, [c2014], p. web).

A campanha ganhou adesão de artistas e personalidades de destaque, como políticos, e foi divulgada em vários meios de comunicação no Brasil, e até mesmo no exterior, mas concomitantemente a repercussão foi publicada a correção de dados da pesquisa do IPEA.

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, 2014b; 2014c) 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, 2014c).

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 passa a 26%, sendo que 70% não concordam e 3,4% são neutros (INSTITUTO DE PESQUISA ECONÔMICA APLICADA, 2014c).

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 atenção da

opinião pública, no entanto, apesar da mudança de percentual ser muito grande, de 65% para 26%, os números permanecem altos.

Porém, como o foco do trabalho não consiste em investigar questões que envolvem as respostas, os resultados da pesquisa tiveram a função de despertar o interesse na investigação da representação do assunto estupro. De um lado temos a repercussão nas mídias sociais, como no Twitter, que faz uso da LN e, de outro, os ambientes físicos como as Bibliotecas, que utilizam os instrumentos de controle de vocabulário, como os tesouros e outras LDs, para representar assuntos.

Portanto, a seguir será descrita a metodologia empregada no percurso da pesquisa e, após, de forma breve é realizada abordagem sobre linguagem, a fim de alcançar a contextualização da LN e da LD no escopo que se propõe.

Procedimentos metodológicos

A pesquisa é classificada como descritiva, pois visa descrever as características das linguagens natural e documentária e estabelecer relações entre ambas através de observação sistemática (MARCONI; LAKATOS, 2003).

Também se caracteriza como pesquisa bibliográfica e qualitativa. Realiza-se a investigação teórica do descritor estupro embasada na análise em dois tesouros - no Brased e no STJ - e fundamenta-se comparação com a linguagem natural empregada na Rede Social Twitter através da *hashtag* #NãoMereçoSerEstuprada. Tem-se que ressaltar que foram inseridos, neste trabalho, apenas o recorte de um descritor dos referidos tesouros - estupro - não ampliando a pesquisa para outros termos associados.

A investigação possibilitou o levantamento geral de aspectos pertencentes às Redes Sociais e foi possível visualizar como um assunto pode apresentar formas diferentes de representação, dependendo do domínio que está inserido.

A Linguagem

De acordo com Dahlberg (1978) a linguagem constituiu-se na capacidade do homem em dar nome aos objetos que estavam ao seu redor e comunicar-se com os iguais, de modo que permitiu-lhe se relacionar 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).

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 inteiramente no conjunto de uma sociedade.

Já 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).

Por isso, 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 expressamos à experiência segundo padrões da cultura da qual participamos, e como toda linguagem tem por objetivo comunicar algo a um receptor.

No contexto da representação de assuntos os termos geralmente atribuídos na LN 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.

Linguagem Natural no Contexto da Web 2.0

Com a crescente expansão das Redes Sociais qualquer acontecimento se torna acessível a quem se conecte a internet, podendo haver, inclusive, interação entre quem transmite e quem recebe. Acompanha-se, também, o crescente uso de computadores e até mesmo de *smartphones*⁵ pela população, o que possibilita meios de acesso aos ambientes que transmitem informações.

Com isso, transformam-se comportamentos e ideias com ferramentas interativas e esta nova e promissora fase da internet chama-se *Web 2.0*⁶ sendo baseada na construção coletiva do conhecimento. Nessa direção,

[...] 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) e Moura (2009), três dimensões fundamentais, sendo dimensão sócio-comunicacional, dimensão linguístico-discursiva e a dimensão de produção de sentidos. Assim, de forma sucinta, 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.

Qin (2008⁷ apud MOURA, 2009) sinaliza que esses ambientes podem ser considerados espaços sociais semânticos e identifica a existência de comunidades que se organizam a partir

⁵ 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.

⁶ 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).

⁷ QIN, J. **Folksonomies and taxonomies where the two can meet**. 09 out. 2008. Disponível em: <<http://nkos.slis.kent.edu/2008workshop/JianQin.pdf>>. Acesso em: 15 set. 2009.

do compartilhamento de informações. Esses espaços são criados por usuários autônomos que compartilham diferentes tipos de informação e distintas visões do mundo que os cerca. Portanto, as *tags* ou *hashtags*, ou seja, as etiquetas

[...] são a base dos espaços sociais semânticos e podem ser estudadas em função das palavras (linguística), dos símbolos programáticos (Ciência da Computação), das significações (Semiótica e Psicologia), da lógica simbólica (Filosofia), dos metadados (Biblioteconomia e Ciência da Informação). (QIN, 2008⁷ apud MOURA, 2009, p. 29).

No contexto da *Web 2.0* e com o uso cada vez mais frequente das *tags* ou *hashtags* surge a *folksonomia*, e esta objetiva à representação livre da informação em LN, sem adotar regras ou políticas de indexação e, dessa forma, torna-se o resultado da etiquetagem dos recursos da web, em um ambiente social, pelos próprios usuários com o objetivo de uma futura recuperação.

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, 2006⁸ apud PEREIRA; CRUZ, 2010, p. web).

De acordo com Catarino e Batista (2009) 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.

⁸ NORUZI, A. Folksonomy: why do we need controlled vocabulary?. **E- prints in Library and Information Science**, p. 7, 2006. Disponível em: <<http://eprints.rclis.org/archive/00011286/>>. Acesso em: 28 mar. 2010.

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, a característica do texto curto é o que mantém o serviço do Twitter rápido e relevante, incentivando as pessoas a *tweetarem*⁹ de modo fácil e somente com a essência da ideia que estão tentando comunicar (DISCOVER TWITTER, [c2014]).

As Redes Sociais, de um modo geral, representam um espaço de compartilhamento de opiniões, experiências e conteúdos. Dentro das Redes, cada usuário se torna um produtor de conteúdo de informação para todas as pessoas e em um *tweet* é possível ver fotos, vídeos e incluir links de notícias, blogs, *websites* e aplicativos.

No Twitter os usuários usam o símbolo marcador # antes de uma palavra-chave ou frase relevante (sem espaços) nos *tweets* que postam e, dessa forma, a marcação tem a finalidade de categorizar os *tweets* e facilitar a sua localização em uma busca na própria rede social. 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 serem muito populares são, muitas vezes, “Assuntos do Momento”, portanto *Trending Topics* (TWITTER, c2014b).

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

De acordo com Primo (2010, p. web) “a tag ou ‘hashtag’ é um recurso que foi criado dentro do micro blog Twitter de forma espontânea pelos seus usuários”. O autor corrobora que a *hashtag* é um tipo de palavra-chave determinada pelo sinal # (cerquilha) seguido da palavra ou conjunto de palavras escritas juntas e sucessivas, sem nenhum outro tipo de pontuação. Desde a primeira vez que foi utilizada serviu para 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 fez sobre os incêndios que estavam acontecendo em San Diego.

No Brasil, as *hashtags* também são muito usadas, como é o caso da campanha pela não violência contra mulher, #NãoMereçoSerEstuprada, por exemplo. Ainda de acordo com Primo

⁹ Ato de enviar mensagens pela plataforma do Twitter.

(2010, p. web) “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 conjuntos de palavras podem ser utilizadas como *hashtag*, e estas palavras-chave facilitam a recuperação de assuntos 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.

O próprio Twitter facilita a recuperação de *tweets* (mensagens postadas na plataforma que possuem até 140 caracteres) com a mesma etiqueta transformando todas *tags* em *links*. 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, tem-se #estupro e como exemplo de uma *hashtag* mais longa #NãoMereçoSerEstuprada. Acrescenta-se que o único limite para o tamanho das *hashtags* no Twitter são os 140 caracteres admitidos.

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

Além dos *Trending Topics*, o Twitter também passou a publicar a lista das *hashtags* mais utilizadas anualmente, e a partir de um movimento espontâneo criado por um usuário o Twitter conta atualmente com estrutura que facilita a busca e recuperação de mensagens (PRIMO, 2010).

A *Web 2.0* permite o uso de linguagens mais flexíveis e de padrões cada vez mais aceitos para representação da informação. Campos (2001) direciona que todo movimento existente nos sistemas de recuperação de informação tem por objetivo possibilitar ao usuário acesso a informação. Para a autora, vários são os instrumentos utilizados para representar o conhecimento em uma dada área do saber. Sendo assim, na próxima seção o foco é a LD.

Linguagem Documentária

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

Para Fujita (2004, p. web), “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.

Cintra (et al., 2002) direciona que 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 dos assuntos dos documentos. Os sistemas de classificações decimais, como a Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU), os tesouros, dentre outros, são tipos de linguagens documentárias que permitem agrupar documentos segundo o seu conteúdo, visando o armazenamento e à recuperação da informação.

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

Entre os principais tipos de LDs, os tesouros são instrumentos de indexação por conceitos. Surgiram diante da necessidade de recuperação da informação sendo uma linguagem

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).

controlada que tem como característica os relacionamentos entre os termos que o compõe (MOREIRA; MOURA, 2006).

De acordo com Gomes (1950¹¹ apud DODEBEI, 2002, p. 66), “o tesouro documentário surgiu da necessidade de manipular grande quantidade de documentos especializados. Era preciso trabalhar com um vocabulário mais específico e com uma estrutura mais depurada do que aquela presente nos cabeçalhos de assunto [...]”

Dodebei (2002, p. 67) também destaca que:

O emprego de tesouros nas tarefas de indexação e recuperação da informação tenta resolver o problema da alocação de documentos em classes de assuntos, não só por sua capacidade de controlar o vocabulário, mas porque é um instrumento que relaciona os descritores/termos de forma mais consistente, apresentando uma estrutura sintética simplificada e uma complexa rede de referências cruzadas. Isto permite ao especialista localizar com mais facilidade a palavra-chave requerida para uma busca.

Para Moreira (2005) 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.

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 e, deste modo, 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 os utilizados, pela comunidade usuária, 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, p. web) “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 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

¹¹ GOMES, H. E. (Org.). **Manual de elaboração de tesouros monolíngues**. Brasília: PNB, 1990.

relacionamentos: “o relacionamento de equivalência, o relacionamento hierárquico e o relacionamento associativo”.

As relações de equivalência acontecem quando existem sinônimos, ou quase sinônimos e, assim, 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 baseando-se em Currás (1995) 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), significa Usado Por, tem a função de remeter os termos equivalentes para o termo preferencial, como no clássico exemplo:

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

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”. Segundo a autora, quando uma base de dados utiliza o controle do vocabulário, como o tesouro, cria a possibilidade para o planejamento da estratégia de busca no campo específico de descritor, através de palavras-chave listadas em ordem alfabética.

O instrumento tesouro é constituído em domínios específicos de assuntos e, portanto, dois domínios embasarão a discussão pretendida.

O Tesouro Jurídico do Superior Tribunal de Justiça (STJ)

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. (BRASIL, [200?]).

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 domínio do Direito (BRASIL, [200?]).

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:

- 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 tesouro 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. (BRASIL, [200?], p. web).

O Tesouro Jurídico do STJ encontra-se organizado em categorias e subcategorias que foram codificadas utilizando-se caracteres alfanuméricos.

Como já apontado, Os tesouros são instrumentos específicos de uma determinada área ou domínio e, Deste modo, um mesmo termo pode ser representado de forma hierárquica diferente, pois dependerá da área de inserção. Para essa análise foram consultados dois tesouros de áreas diferentes, de modo que se possa verificar a apresentação do descritor estupro em ambos. Assim, no Tesouro Jurídico do STJ foi possível identificar a apresentação, a subordinação, os termos gerais (TG), os termos específicos (TE) e seus termos relacionados (TR).

Observa-se que a hierarquização do descritor estupro aparece 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 e este tipo de relação permite que um termo específico, como no caso do termo estupro, possua relação direta com mais de um termo genérico.

O Thesaurus Brased – Thesaurus Brasileiro de Educação

O Thesaurus Brased teve construção iniciada em 1980, sob a coordenação do professor Gaetano Lo Monaco e foi criado no intuito de facilitar a pesquisa em Educação, sendo pioneiro da área 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 (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, c2011).

No Brased o contexto global da educação está de acordo com a matriz conceitual, composta 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, c2011).

Na estrutura do referido Thesaurus, 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, c2011).

No campo “100 - Contexto da Educação” o descritor estupro é apresentado como último, sendo o mais específico dentro da cadeia hierarquicamente apresentada, pois se desdobra desde

o topo da cadeia em “contexto da educação”, passando por “contexto social”, “condições sociais”, “problemas sociais”, “criminalidade”, “crime”, “delitos sexuais” e, por fim, “estupro”.

Análise dos dados: o descritor Estupro na Linguagem Documentaria X Linguagem Natural

Diante do resultado da pesquisa do IPEA sobre tolerância social à violência contra as mulheres houve grande mobilização e repercussão no país, sobretudo nas Redes Sociais, onde a jornalista Nana Queiroz deu início a campanha de repúdio contra o resultado da pesquisa com a *hashtag* #NãoMereçoSerEstuprada. Embora os temas centrais da polêmica gerada a partir do resultado da pesquisa do IPEA, representada pela *hashtag* #NãoMereçoSerEstuprada, tenha sido a violência contra a mulher e o estupro, na Rede Social Twitter pouco se constata *hashtags* com essas palavras.

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 usuário 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 do proposto, ou seja, ao contrário da LD que visa a redução dos significados em busca de maior especificidade, no caso das *hashtags* em LN, os sentidos se perdem ilimitadamente.

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, nem com outros termos relacionados a questão da violência contra as mulheres.

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 de qualquer sistema de informação.

A LN pode ser entendida como sinônimo de discurso comum, pois é a linguagem usada habitualmente na fala e na escrita, porém em sistemas de informação pode corresponder aos termos do título e resumo dos documentos, por exemplo. Já a LD visa substituir o texto do documento em LN, por sua descrição abreviada, utilizada como um artifício para recuperar o que é essencial no documento, ou seja, o assunto central.

Dessa forma, observa-se que o controle do vocabulário é um elemento essencial para a representação 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 do domínio que está inserido. No caso desta pesquisa, o termo estupro tem o mesmo conceito nos dois tesouros consultados, no entanto há uma diferença no contexto em que se apresentam.

No tesouro do STJ o descritor apresenta uma relação hierárquica que exprime os graus de superordenação e subordinação e o termo superordenado (TG - termo genérico) representa o conceito mais abrangente em relação ao subordinado (TE - termo específico).

O descritor no contexto jurídico dispõe de uma relação hierárquica de polierarquia e este tipo de relação permite que um termo específico, como 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 Brased pôde-se observar que na hierarquização dos termos, de acordo com a matriz conceitual, é composto por quatro campos, sendo: contexto da educação, escola como instituição social, fundamentos da educação e educação. Nessas divisões, o termo estupro está estruturado de forma mais genérica no campo do “contexto da educação” e é subdividindo em campos cada vez mais específicos. Portanto, o referido está inserido no contexto da educação, sendo o mais genérico da cadeia, mas é direcionado por contexto social, condições sociais, problemas sociais, criminalidade, crime e delitos sexuais, sendo estupro o mais específico, como é possível observar abaixo:

Contexto da educação
Contexto social
Condições Sociais
Problemas sociais
Criminalidade
Crime
Delitos Sexuais
Estupro

Outros descritores poderiam ser utilizados para representação da pesquisa do IPEA, como por exemplo, “Violência Sexual”. Tanto no tesouro do STJ, quanto no Brased, este termo aparece como termo relacionado. Nota-se que um termo na LD não existe por si só, isolado, e sempre estará contextualizado com outro termo, no entanto na LN observada no Twitter, através da *hashtag*, 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 necessite da informação.

Depois de meses do resultado da pesquisa do IPEA e toda mobilização nas Redes Sociais foi realizada nova pesquisa pela *hashtag* #NãoMereçoSerEstuprada no Twitter (na data de 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 #NãoMereçoSerEstuprada perdeu força e foco, e a *hashtag* supracitada é também utilizada, por exemplo, para dissipar outras informações, como no caso sarcástico noticiado que focas estariam abusando sexualmente de pinguins-rei na Ilha Marion, localizada no Atlântico Sul (O GLOBO, 2014).

A LN, sobretudo nas Redes Sociais, é constituída por termos representados pelas *hashtags* 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 indexador exerce papel de mediador entre o usuário e o assunto, traduzindo o documento em termos que possam ser recuperados de forma mais precisa e rápida.

Portanto, percebe-se que o uso da LN em relação à LD apresenta, por exemplo, desvantagem na recuperação do assunto representado. Na LD também se deve levar em conta que nem sempre os termos preferidos pelos indexadores serão os utilizados pelos usuários nas

buscas. Entretanto, o emprego da LD em sistemas de busca e recuperação da informação é um instrumento confiável que necessita ser constantemente atualizado.

Além disso, pela hierarquia do tesouro é possível compreender as conexões do descritor com os termos relacionados, o que facilita o conhecimento de um domínio. Como apontado, a LN não proporciona esse entendimento.

Delineamento de Conclusões

Em LN, qualquer palavra, ou até mesmo conjuntos 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*, mas a descrição não é padronizada.

Conclui-se que o uso da LN desencadeia algumas vantagens, tais como, 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, a LN através do uso da *hashtag* não padroniza a representação e, por conseguinte, influencia a recuperação.

Por sua vez, as LDs apresentam vocabulário próprio que se caracteriza por maior precisão de termos e, em geral, o descritor possui significado unívoco, mas o mesmo não acontece com a LN. Dessa forma, torna-se essencial o controle do vocabulário para a representação da informação, pois é uma linguagem artificial produzida com o objetivo de sistematizar o vocabulário do usuário para um vocabulário padronizado.

Contata-se que tanto na LN, quanto na LD, se faz necessário dar condições favoráveis à recuperação, mas 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 determinado assunto, sobretudo nas Redes Sociais, dispensando uma pesquisa mais específica e agilizando o processo de comunicação, quanto dificultar, pois sem a devida padronização certamente ocorrerá um alto índice de respostas negativas na recuperação.

Por fim, compreende-se a necessidade de outras pesquisas sobre a temática, pois o uso de linguagens de marcação é crescente nas Redes Sociais suscitando pontos importantes para a representação e busca da informação. Enveredar por esse escopo fortalecerá as discussões.

Referências

BRASIL. Superior Tribunal de Justiça. Secretaria de Jurisprudência do Superior Tribunal de Justiça. **Consulta ao vocabulário jurídico controlado**. Brasília, DF: A Secretaria, [200?]. Disponível em: <http://www.stj.jus.br/SCON/thesaurus/ajuda_thes.jsp>. Acesso em: 03 maio 2014.

BRESSER, D. Valendo! #Naomerecosestuprada. **Blog da DB**. 28 mar. 2014. Disponível em: <<http://entretenimento.r7.com/blogs/blog-da-db/valendo-naomerecosestuprada-20140328/>>. Acesso em: 30 mar. 2014.

CAMPOS, M. L. A. **Linguagem documentária: teorias que fundamentam sua elaboração**. Niterói: EDUFF, 2001.

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

CINTRA, A. M. M. Elementos de linguística 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: 08 maio 2014.

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

COELHO, M. O que são e como funcionam os Trending Topics: recurso mede popularidade de assuntos no Twitter. **Home iG Tecnologia**, 25 ago. 2011. Disponível em: <<http://tecnologia.ig.com.br/o+que+sao+e+como+funcionam+os+trending+topics/n1597175643026.html>>. Acesso em: 03 maio 2014.

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

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, Brasília, v. 7, n. 2, p. 101-107, 1978. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/1680/1286>>. Acesso em: 05 maio 2014.

DISCOVER TWITTER. [c2014]. Disponível em: <<https://discover.Twitter.com/pt/learn-more>>. Acesso em: 03 maio 2014.

DODEBEI, V. L. D. L. M. **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: 08 maio 2014.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **SIPS – Sistema de indicadores de percepção social**: tolerância social à violência contra as mulheres. 27 mar. 2014a. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/SIPS/140327_sips_violencia_mulheres_antigo.pdf>. Acesso em: 30 mar. 2014.

_____. **SIPS – Sistema de indicadores de percepção social**: tolerância social à violência contra as mulheres. 04 abr. 2014b. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/SIPS/140327_sips_violencia_mulheres_novo.pdf>. Acesso em: 30 mar. 2014.

_____. **Errata da pesquisa**: tolerância social à violência contra as mulheres. Portal IPEA, 04 abr. 2014c. Disponível em: <http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=21971&catid=10&Itemid=9>. Acesso em: 05 maio 2014.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Thesaurus brasileiro da educação**. c2011. Disponível em: <<http://portal.inep.gov.br/pesquisa-thesaurus>>. Acesso em: 03 maio 2014.

LOPES, I. L. 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: 07 maio 2014.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2003.

MARTELETO, R. M. A metodologia de análise de redes sociais (ARS). In: VALENTIM, M. L. P. (Org.). **Métodos qualitativos de pesquisa em ciência da informação**. São Paulo: Polis, 2005. p. 81-100.

MOREIRA, M. P. **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.

_____; MOURA, M. A. Construindo tesouros a partir de tesouros existentes: a experiência do TCI - Tesouro em Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 7, n. 4, ago. 2006. Disponível em: <http://www.dgz.org.br/ago06/Art_01.htm>. Acesso em: 07 maio 2014.

MOURA, M. A. Folksonomias, redes sociais e a formação para o tagging literacy: desafios para a organização da informação em ambientes colaborativos virtuais. **Informação & Informação**, Londrina, v. 14, n. esp., 2009. Disponível em:

<<http://www.brapci.ufpr.br/documento.php?dd0=0000007945&dd1=eb2be>>. Acesso em: 24 abr. 2014.

NARVAZ, M. G.; KOLLER, S. H. Famílias e patriarcado: da prescrição normativa à subversão criativa. **Psicologia & Sociedade**, 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: 01 maio 2014.

O GLOBO. **Internautas fazem piada com as focas ‘estupradoras’ de pinguins do Atlântico Sul**. 18 nov. 2014. Disponível em: <<http://oglobo.globo.com/sociedade/tecnologia/internautas-fazem-piada-com-as-focas-estupradoras-de-pinguins-do-atlantico-sul-14590817>>. Acesso em: 19 nov. 2014

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. **Fantástico**, 31 mar. 2014. Disponível em: <<http://g1.globo.com/fantastico/noticia/2014/03/organizadora-de-campanha-contr-estupro-recebe-ameacas-na-web.html>>. Acesso em: 05 abr. 2014.

PEREIRA, D. C.; CRUZ, R. C. Folksonomia e tags afetivas: comunicação e comportamento informacional no Twitter. **DataGramZero**, Rio de Janeiro, 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: 05 maio 2014.

_____. As tags no Twitter como informação contextual de afeto. **Blog Dossiê Alex Primo**. 09 mar. 2010. Disponível em: <http://alexprimo.com/2010/03/09/as_tags_no_Twitter_como_informacao_conte/>. Acesso em: 05 maio 2014.

RECUERO, R. **Redes sociais na internet**. Porto Alegre: Sulina, 2009.

SEXISMO. In: Dicionário Priberam da Língua Portuguesa, c2013. Disponível em: <<http://www.priberam.pt/dlpo/sexismo>>. Acesso em: 01 maio 2014.

TWITTER. **O que são os marcadores (símbolos de "#")?**. c2014a. Disponível em: <<https://support.Twitter.com/articles/255508-o-que-sao-os-marcadores-simbolos-de#>>. Acesso em: 03 maio 2014.

_____. **Trending topics**. c2014b. Disponível em: <<https://Twitter.com/>>. Acesso em: 03 abr. 2014.

_____. To trend or not to trend... **Blog**. 08 dez. 2010. Disponível em: <<https://blog.Twitter.com/2010/trend-or-not-trend>>. Acesso em: 03 maio 2014.

Artigo submetido em: 30 jan. 2015
Artigo aceito em: 01 set. 2015