

Análise Gramatical de Etiquetas em Sites Colaborativos: Estudo de Caso Flickr

Grammar Analysis of Tags In Collaborative Systems: a Flickr Case Study

Análisis Gramatical de Etiquetas en Sitios Colaborativos: Estudio de Caso Flickr

*Guilherme Francis Noronha¹
Manoel Palhares Moreira²*

Resumo

A grande produção de documentos na internet já é realidade. Grande parte deles são disponibilizados pelos mesmos usuários que os consomem. A classificação em massa de documentos por especialistas da informação tornou-se impossível. Para tentar resolver esse problema surgem as etiquetas sociais em sistemas colaborativos. Os usuários seriam responsáveis por classificarem seus próprios documentos. Porém eles, em sua grande maioria, não dispõem de conhecimentos suficientes para indexação e recuperação correta de documentos. Surge então o problema de indexação errônea pelos usuários. Estudos tentam avaliar a utilidade dos sistemas colaborativos e se os usuários possuem consciência da importância da classificação correta de documentos. Esse trabalho tem como objetivo fazer um estudo de caso do Flickr. Foram analisados padrões de etiquetas tais como análise de gênero e número, uso de termos compostos, repetição e ausência de etiquetas nos documentos. Foram traçados padrões de comportamento dos usuários e como eles fazem uso etiquetas. Usou-se os padrões do NISO para análise de etiquetas e as classificações de Cañada para análise de usuários. Os resultados encontrados mostram que, apesar de haverem problemas no sistema colaborativo, tanto nas etiquetas quanto nos usuários, eles ainda funcionam como sistemas de indexação e recuperação de informação de documentos produzidos pelos próprios usuários.

Palavras-chave: Flickr. Sistemas Colaborativos. Análise de Etiquetas. Análise Gramatical. Análise de Usuário.

Abstract

The huge documents production on the internet is a fact. The great amount of these documents are provided by the same users who consume them. The mass documents classification by information specialists became impossible. Trying to solve this, the social tags in collaborative systems were created. The users became responsible to classify their own documents but in their vast majority, they do not have enough knowledge to index and retrieve documents correctly. Then comes up the problem of the documents that are wrongly indexed. Studies try to assess the usefulness of collaborative systems and whether the users have the consciousness of the importance of the right classification of documents. This paper has as its main goal to do a case study with the Flickr. Were analyzed patterns from the tags such as gender, plural, compound terms, repetition and absence of tags in documents. By means of this analysis, were identified patterns of users behaviors and how they use tags. Were used the NISO patterns to analyze the tags and classifications of Cañada to analyze the users. The results show that in spite of have problems in the collaborative systems both in tags and users these systems still work to index and retrieve information produced by the users.

¹ Doutorando e Mestre em Gestão e Organização do Conhecimento pela Universidade Federal de Minas Gerais

² Doutor em Ciência da Informação pela Universidade Federal de Minas Gerais.

Keywords: Flickr. Collaborative Systems. Tag Analysis. Grammatical Analysis. User Analysis.

Resumen

La gran producción de documentos en Internet es una realidad. Gran parte de estos documentos fueron producidos por los mismos usuarios que los consumen. La clasificación de los documentos por los especialistas de la información se hizo imposible. Las etiquetas sociales en sistemas colaborativos intentan resolver este problema. Los usuarios serían responsables por clasificar sus propios documentos. Pero ellos, en su gran mayoría, no disponen de conocimientos suficientes para indexación y recuperación de documentos. Esto lleva al problema de indexación errónea por los usuarios. Los estudios intentan evaluar la utilidad de los sistemas colaborativos y si los usuarios tienen conciencia de la importancia de la clasificación correcta de documentos. Este trabajo tiene como objetivo realizar un estudio de caso de Flickr. Se analizaron patrones de etiquetas como análisis de género y número, uso de términos compuestos, repetición y ausencia de etiquetas en los documentos. Se han identificado patrones de comportamiento de los usuarios y cómo se utilizan etiquetas. Se utilizaron los estándares del NISO para análisis de etiquetas y las clasificaciones de Cañada para análisis de usuarios. Los resultados encontrados muestran que, aunque existen problemas en el sistema colaborativo, tanto en las etiquetas como en los usuarios, todavía funcionan como sistemas de indexación y recuperación de información de documentos producidos por los propios usuarios.

Palabras clave: Flickr. Sistemas Colaborativos. Análisis de etiquetas. Análisis gramatical. Análisis de Usuario.

1 INTRODUÇÃO

A popularização da internet nos anos 2000 trouxe consigo uma explosão informacional nunca antes vivida na história. Esse grande volume de dados, apelidado pelos cientistas de Big Data, trouxe uma série de problemas em meio ao avanço tecnológico (GANDOMI; HAIDER, 2015). Um deles se relaciona diretamente com a forma como o usuário consome a própria informação. Boa parte dessa explosão informacional é gerada pelo próprio usuário que deixa de ser consumidor passivo e se torna um consumidor ativo da informação. Comportamento esse chamado de “prosumidor”, já antecipado por Kotler (1986). Ele produz e consome a própria informação. Esses aspectos são definidos como sistemas colaborativos, onde os usuários contribuem para a formação de um ecossistema (BAFOUTSOU; MENTZAS, 2002).

A quantidade de informação gerada torna impossível que um sistema especialista de classificação atue para indexar os documentos compartilhados por usuários. Alternativas como tesouros e taxonomias deixam de ser viáveis. Para contornar isso, surge então, o conceito de folksonomia, ou etiquetagem social, onde o próprio usuário é responsável por classificar o conteúdo que produz (SPITERI, 2007).

No entanto, o usuário, na maioria das vezes, não é especialista nem profissional ligado à Ciência da Informação que possua conhecimentos de classificação. Surge então os

problemas relacionados com mau uso de vocabulários e etiquetas.

Guy *et al.* (2006) identificaram uma série de problemas relacionados com mal uso de etiquetas em sistemas colaborativos. Símbolos, falta de convenções, palavras com erros de digitação, conjugação incorreta e mau uso de termos compostos são alguns dos problemas identificados pelos autores. Além disso, há o surgimento do uso de etiquetas como meios de formação de comunidade e busca por popularidade.

Pensando nesse fenômeno em sistemas colaborativos, o presente artigo levanta a seguinte pergunta de pesquisa: **Como o usuário brasileiro faz uso de etiquetagem social?** Para isso foi levantado o seguinte **objetivo geral**: realizar um estudo de caso que identifique o comportamento do usuário brasileiro por meio de análise gramatical, repetição e ausência de etiquetas usadas em sistemas colaborativos. E como **objetivos específicos** têm-se: (1) construir uma ferramenta automática de análise de etiquetas; (2) analisar a flexão em gênero e número dos termos nos documentos; (3) analisar se o usuário faz uso de termos compostos em múltiplas etiquetas; (4) analisar a ausência ou repetição de etiquetas em documentos diferentes; (5) comparar os dados obtidos com normas de vocabulários controlados para identificar se esses sistemas estão sendo usados de maneira correta; (6) identificar o comportamento de usuário dentro de sistema de classificação de Cañada (2012); (7) análise dos sistemas de recuperação da informação tendo em vista a forma como o usuário comum acessa a informação através de etiquetas: a Folksonomia acresceu algo a esse processo? Alterações gramaticais influenciam na forma como ele realiza essa busca?

A justificativa do presente trabalho é contribuir com a análise desses sistemas identificando a forma como o usuário brasileiro faz uso de etiquetas. A literatura sobre vocabulários controlados constrói padrões de bom uso de etiquetas para que sejam usadas como sistemas de recuperação de informação e é preciso analisar se os sistemas colaborativos possuem capacidade para isso.

O presente trabalho foi dividido em várias seções para melhor compreensão. Na **seção 2** contém a revisão de literatura de sistemas colaborativos, tipologias de etiquetagens e recomendações de uso de etiquetas. A **seção 3** identifica como tem sido os trabalhos na área de etiquetagem social nos últimos anos. A **seção 4** detalha a metodologia do trabalho. A **seção 5** apresenta os resultados encontrados e correlaciona-os com os métodos de Cañada (2012) e com as orientações da NISO. Por fim, a **seção 6** traz uma discussão dos resultados e propõe trabalhos futuros.

2 REVISÃO DE LITERATURA

O presente trabalho aborda o uso de etiquetas por usuários em sistemas colaborativos. Como revisão de literatura considerou-se noções de sistemas colaborativos, diretrizes para etiquetar documentos e o comportamento de usuários nesses sistemas.

2.1 SISTEMAS COLABORATIVOS

Os sistemas colaborativos são uma evolução natural da internet. O'Reilly (2005) chamou essa evolução de Web 2.0. A web tornou-se um sistema dinâmico onde os usuários, por meio de funcionalidades, conseguem atualizar e colaborar constantemente o sistema. O usuário é agora prosumidor de informação. Surge então o conceito de folksonomia, ou etiquetagem social. Cunhado por Wal (2007), é o resultado de formalizar a ação de classificação de documentos por pessoas não especializadas. Segundo o autor, etiquetagem social pode ser definida como o resultado da classificação livre da informação e objetos. Não há o conceito de taxonomia nesse cenário, pois não existem critérios para classificação de um documento.

Colaborações se tornaram comuns na internet por meio de compartilhamento de dados, sejam eles links, imagens, vídeos etc. As redes sociais, como Facebook, Instagram e Flickr são alguns exemplos de produtos onde usuários são prosumidores. A colaboração em massa dos usuários trouxeram problemas junto com a evolução dos sistemas. Um deles, é a ausência de vocabulários controlados que apresenta uma identificação maior com o usuário, mas excesso de flexibilização (PETERS, 2009).

A flexibilidade de etiquetas utilizadas na web exige pouco do usuário. Isso faz surgir problemas como ambiguidades, erros ortográficos, separação de termos compostos que mudam o significado do contexto e códigos que fazem o conteúdo ser recuperado apenas por um grupo específico de pessoas. Na maioria dos casos, a recuperação é imprecisa e com alta revocação. Então a organização e recuperação de informação precisou se remodelar para conseguir categorizar e classificar os documentos disponibilizados na internet sem nenhum controle.

A prática de etiquetagem em redes colaborativas têm ideais que vão além do conceito moderno de recuperação da informação apresentado por Baeza-Yates e Ribeiro Neto (2011). A participação ativa da comunidade em um sistema também cria um movimento de identidade e formação de nichos de usuários. Surge então o problema da etiquetagem social. São esforços dos usuários para que as etiquetas sejam usadas para conectar outras pessoas. Isso permite que pessoas recuperem documentos que elas etiquetam ao mesmo tempo que outras pessoas também tenham acesso (WAL, 2005).

Por exemplo, em sistemas colaborativos de imagens, como Instagram e Flickr, a etiqueta “TBT”, que significa “*Throw Back Thursday*” ou algo como quinta-feira da recordação, é usada para conectar usuários que postam fotos antigas às quintas-feiras. Outro exemplo de uso de etiquetas de forma social é a “*thepicoftheday*”, que significa a foto do dia, que conectam pessoas que querem ver a melhor foto tirada no dia por outras pessoas. Portanto, o uso de etiquetas na web 2.0 ganha novo sentido e uso, além do convencional usado em sistemas formais, como tesouros e taxonomias.

2.2 TIPOLOGIAS DE ETIQUETAGEM

Cañada (2012) afirma que existem motivações para as pessoas etiquetarem algum objeto dentro de um sistema colaborativo. Essas motivações se dividem em quatro estilos muito diferentes de etiquetagens: egoísta, amigável, popular e altruísta.

As etiquetagens egoístas são aquelas que não possuem propósito de recuperação de informação em massa, mas sim para benefício próprio. As etiquetas egoístas não fazem sentido para o documento marcado, apenas para o usuário que o marcou. Os documentos são etiquetados egoístamente com o intuito de ser recuperado somente pelo usuário no futuro. Já as etiquetas amigáveis têm o intuito de serem recuperados por nichos específicos de usuários que compartilham conteúdos entre si. Elas funcionam como códigos de comunicação entre a comunidade e pouco tem a ver com o verdadeiro conteúdo do documento. As etiquetas altruístas são as que mais se assimilam com o vocabulário controlado e o conceito de taxonomia definido na recuperação de informação clássica (PIRAQUIVE; AGUILAR; GARCÍA, 2009). Elas procuram descrever o conteúdo documentado e auxiliar a recuperação de informação em massa.

Por fim, as etiquetas populares são a de pior uso em sistemas de recuperação de informação. Elas são etiquetadas com o único propósito de ter mais audiência para o usuário. Para isso, eles se utilizam de marcação que chamam atenção, em vez de classificar corretamente os conteúdos.

2.3 RECOMENDAÇÕES DE USO DE ETIQUETAS EM VOCABULÁRIOS CONTROLADOS

National Information Standards Organization, ou NISO, é uma organização que desenvolve e mantém padrões de categorização de documentos bibliográficos e aplicações de indexação de bibliotecas. Possui um manual de referência para etiquetagem usando vocabulários controlados (NISO *et al.*, 2010). Embora sistemas colaborativos não se

apresentem como vocabulários controlados, as recomendações do NISO funcionam como critérios avaliativos para saber se a plataforma está sendo usada de forma adequada pelos usuários. Critérios como uso de etiquetas e sua extensão como marcação de conteúdo são alguns exemplos de aplicação (SPITERI, 2007).

2.3.1 Recomendações para número

As recomendações para flexão de número, singular ou plural, estão diretamente direcionadas à classe de palavras que a etiqueta possui. Para palavras que sejam quantificáveis, como “garoto”, “gato”, “cantor” etc, os plurais devem ser usados. As palavras não quantificáveis, conceitos abstratos e entidades únicas como “fluxo de água”, “budismo” e “torre Eiffel” devem permanecer no singular.

No entanto, há exceções bastante específicas onde a coexistência de singular e plural podem ser usadas em um mesmo documento. Quando os conceitos diferem de sentido, no singular e plural, ambas as etiquetas devem ser usadas. Por exemplo “dano” é uma etiqueta que representa estrago a um objeto e “danos” é uma etiqueta que pode representar um conjunto de pessoas de uma tribo dinamarquesa.

2.3.2 Recomendações para termo composto

Um termo é considerado composto se ele possui mais de uma palavra. Essa palavra pode estar conectada por hifens ou não. O único requisito é que esse conjunto expresse um único significado. Por exemplo, a etiqueta “Belo Horizonte” é um termo composto que expressa o nome de uma cidade.

Outra aceitação de termo composto é dada se um conjunto de palavras, ainda que sejam múltiplos substantivos, pertençam a uma mesma categoria superior. Por exemplo, a etiqueta “carros e caminhões” pode ser considerada termo composto por que ela representa termos dentro do conceito de automóveis.

2.3.3 Recomendações para gênero

O NISO não possui recomendações para flexões de gênero, pois o inglês é um idioma que não possui um sistema de concordância de gêneros (AUDRING, 2008). A flexão de gênero para o inglês acontece apenas para os pronomes que são comumente chamados de gêneros naturais, por exemplo, “ele”, “ela”, “seu”, “sua” etc.

Portanto, para o português, as regras a serem seguidas para flexões de gênero são as mesmas para número. As classes de palavras no feminino e masculino também podem ser

classificadas em quantificáveis e não quantificáveis. Os exemplos anteriores de garoto, gato e cantor podem ser substituídos por garota, gata e cantora.

Palavras que somente se flexionam com o uso de palavras auxiliares, como artigos e adjetivos, devem recorrer às recomendações de termos compostos. Por exemplo, as etiquetas “jacaré fêmea” e “a presidente” são categorias de etiquetas que se enquadram tanto em termo composto, quanto em flexão de gênero.

3 ANÁLISE DE ETIQUETAS EM SISTEMAS COLABORATIVOS: TRABALHOS CORRELATOS

Os trabalhos correlatos buscam identificar trabalhos similares que fazem análise de etiquetas em sistemas colaborativos. Há a discussão se a etiquetagem nesses sistemas contribuem para categorização e recuperação da informação ou se servem apenas para interação social e criação de nichos específicos. Com isso, procura-se compreender o comportamento do usuário que etiqueta o documento em um sistema colaborativo. Seja foto, música, texto ou uma URL. Com a análise do usuário vem também a análise de etiquetas. Sob perspectivas de diversas teorias de classificação e abordagens de autores distintos, esses dados são analisados e comentados. A seguir são listados exemplos de trabalhos similares publicados no Brasil nos últimos anos.

Santos (2013) diz que o propósito de etiquetar varia entre os usuários e os sistemas colaborativos usados. A exemplo das etiquetas “TBT” e “thepicoftheday”, elas são menos observadas em sistemas colaborativos onde as fotos não são o tipo principal de documento. O autor ressalta que nem sempre as etiquetas são usadas de modo popular e que o usuário busca organizar e recuperar seus documentos.

Silva (2015) analisou as etiquetas do site Last.fm usando os modelos propostos por Sen *et al.* (2006), que se categoriza em fatuais, subjetivos e pessoais, e Laplante (2015) que distribui as etiquetas por suas descrições. A análise segundo Sen *et al.* (2006) revelou que a maioria, 68%, das etiquetas são fatuais. A análise segundo Laplante (2015) revelou que as etiquetas representam, em sua maioria, os gêneros musicais a qual as músicas pertencem. E que, apesar disso, nem sempre as etiquetas são fiéis ao conteúdo etiquetado, um problema dos sistemas colaborativos.

Barbosa, Krebs e Sousa (2015) utilizou a dimensão expressiva de Manini (2002) associada com categorizações de localidade, tempo e descrições de imagem de Panofsky (2009) para analisar conteúdos relacionados ao National Geographic no Instagram. Os resultados encontrados pela autora sugerem que os usuários da rede não seguem os padrões de

etiquetagem definidos pela literatura. Nóbrega e Manini *et al.* (2016) analisam as redes sociais de fotografia Instagram e Flickr para analisarem o uso das etiquetas “#naovaitergolpe” e “#impeachment” como ferramentas de recuperação de informação e construção de sociedades em torno de uma polarização política. A análise levou em consideração as categorias de Cañada (2012), os níveis de conteúdo da imagem de Panofsky (2009) e a dimensão expressiva de Manini (2002). Os resultados apontam para usuários etiquetam tanto de forma altruísta quanto populista. Há o sentido de formar comunidades, mas também de recuperar a informação.

Krebs, Laipelt e Rosa (2018) analisam as etiquetas contidas na plataforma CiteULike como alternativa para criação de taxonomias para a área pediátrica. A validação ocorreu por meio de comparação com vocabulários controlados da área médica. Os autores concluíram que as etiquetas colaborativas são boas fontes para indexação de novos termos candidatos referentes à área. Massoni e Flores (2018) explora o uso de etiquetas por meio de um grupo específico no Flickr. Realizando a análise conteúdo, os autores verificaram que existe subjetividade nas motivações de etiquetagens dos usuários, mas que isso não acarreta em prejuízos nos sistemas colaborativos como sistemas de recuperação de informação. Verificou-se que a maioria das etiquetas representam, de fato, o conteúdo real das imagens. Lugares, condições climáticas, pessoas e sentimentos são alguns exemplos encontrados pelos autores. Além da representação temática, foram encontradas pelos autores representações descritivas como técnicas de fotografia, nome de equipamentos e autoria de imagens.

Os trabalhos citados nessa seção analisam diferentes redes sociais com diferentes propósitos. Todos os trabalhos identificaram aspectos sociais que as etiquetas representam em sistemas colaborativos. No entanto, eles afirmam que esse aspecto, por si só, não representa por completo sua natureza. As análises segundo as perspectivas de Cañada (2012), Manini (2002), Panofsky (2009), Sen *et al.* (2006) ou Laplante (2015) mostram que as características de recuperação de informação nesses sistemas continuam presentes. Os usuários categorizam os documentos de forma que retratam, ainda que parcialmente, o conteúdo documentado.

Quando comparado com sistemas rígidos de classificação, tais como as taxonomias, percebe-se que a etiquetagem social não obedecem padrões preestabelecidos. Porém sua relevância é percebida pelos autores que identificam semelhanças entre os métodos, podendo um complementar o outro.

4 METODOLOGIA

A caracterização da metodologia de pesquisa, segundo Gil (2008), considera o

seguinte trabalho como sendo básico, quantitativo, exploratório e estudo de caso. A pesquisa tem como objetivo a exploração do comportamento do usuário em redes colaborativas, por meio de um estudo de caso em uma rede social específica. A coleta de dados e as avaliações utilizam métodos estatísticos quantitativos para fazer análise do problema. A metodologia da pesquisa foi dividida em cinco etapas distintas: coleta de dados, construção da ferramenta de análise gramatical, elaboração da rotina de análise dos dados coletados, análise de fotos e de usuários. Cada etapa será brevemente explicada nas subseções posteriores.

4.1 COLETA DE DADOS

Foi decidido que o universo de pesquisa seria fotos do site Flickr. A escolha foi por causa da consolidação da plataforma como um site colaborativo. Também foi avaliado o seu potencial uso de etiquetas em fotos. As etiquetas são amplamente usadas para formar comunidades e nichos de fotógrafos, avaliadores e hobistas em cima do mesmo tema. Outro fator que influenciou na escolha do Flickr como estudo de caso é a disponibilização gratuita de uma API, um software de integração com o site que permite que sejam criadas rotinas automáticas de coleta. Para delimitar o universo de pesquisa, optou-se por pesquisar somente fotos e usuários concentrados em Minas Gerais. Para isso, usou-se a identificação de etiquetas de georreferenciamentos e limites de fronteira do estado por meio de coordenadas geográficas.

Uma vez delimitado a área de concentração de dados, foram realizadas entrevistas com estudantes, funcionários e pessoas voluntárias para definirem os temas de pesquisa. Esses temas são etiquetas iniciais de busca para a coleta de fotos. A opção por escolher temas diferentes é para evitar que a coleta seja feita somente em um nicho de usuários e, conseqüentemente, influencie nos resultados obtidos.

Temas como futebol, Mineirão e praia, em um total de trinta, foram escolhidos. Para cada tema foram coletados trinta usuários. Para cada usuário, trinta fotos foram coletadas e, enfim, as etiquetas dessas fotos foram extraídas. O universo amostral da pesquisa foi de 9000 fotos. Para facilitar a comparação entre as etiquetas coletadas, um tratamento prévio textual foi realizado. As etiquetas tiveram os caracteres colocados em caixa alta e também tiveram os acentos removidos.

Todo o método de coleta descrito acima foi feito de forma automática. Por meio de algoritmos, escritos em Java, as etiquetas foram extraídas das fotos e armazenadas em arquivos de texto para análise posterior.

4.2 ANÁLISE GRAMATICAL

O presente trabalho fez análises gramaticais de gênero e número. Para isso, foram implementadas regras de flexão de gênero e número encontrados na Gramática Houaiss da Língua Portuguesa (AZEREDO; 2009). Além disso, foram feitas identificações de termos compostos, ausência ou repetição de etiquetas em fotos diferentes pelo mesmo usuário. A **análise de gênero** considerou apenas as regras padrões de flexão de gênero. São elas:

Tabela 1 – regras de flexão de gênero.

Regra gramatical	Flexão	Exemplo
Palavras terminadas em “o”	Substitui-se o “o” por “a”	Advogado e advogada
Palavras terminadas em “ão”	Substitui-se o “ão” por “ã”, “oa” ou “ona”	Anão e anã; leão e leoa; grandão e grandona;
Palavras terminadas em “or”	Adiciona-se “a” ao final da palavra ou substitui-se o “or” por “eira”	Trabalhador, trabalhadora e trabalhadeira
Palavras terminadas em “e”	Substitui-se o “e” por “a”	Elefante e elefanta
Substantivos terminados em “ês”, “l” ou “z”	Adiciona-se “a” ao final da palavra	Marquês e marquesa; bacharel e bacharela; juiz e juíza

Fonte: Azeredo (2009).

Para a **análise de número** também foram consideradas apenas as regras de flexão de número contidas na gramática. São elas:

Tabela 2 – regras de flexão de número.

Regra gramatical	Flexão	Exemplo
Palavras terminadas em vogal	Adiciona-se o “s”	Advogado e advogados
Palavras terminadas em “z” ou “r”	Adiciona-se “es”	Algoz e algozes; castor e castores
Palavras terminadas em “m”	Substitui-se o “m” por “ns”	Miragem e miragens
Palavras terminadas em “ão”	Substitui-se por “ãos”, “ães” ou “ões”	Ancião e anciãos; escrivão e escrivães; anão e anões
Palavras terminadas em “s” que sejam oxítonas	Adiciona-se “es”	Marquês e marqueses
Palavras terminadas em “s” que sejam paroxítonas ou proparoxítonas	Permanecem invariáveis	Lápis e lápis
Palavras terminadas em “x”	Permanecem invariáveis	Clímax e clímax
Palavras terminadas em “al”, “el”, “ol” ou “ul”	Substitui-se a terminação por “is”	Varal e varais; corcel e corcéis; caracol e caracóis; azul e azuis
Palavras terminadas em “il” que sejam oxítonas	Substitui-se o “il” por “is”	Anil e anis
Palavras terminadas em “il” que sejam paroxítonas	Substitui-se o “il” por “eis”	Réptil e répteis

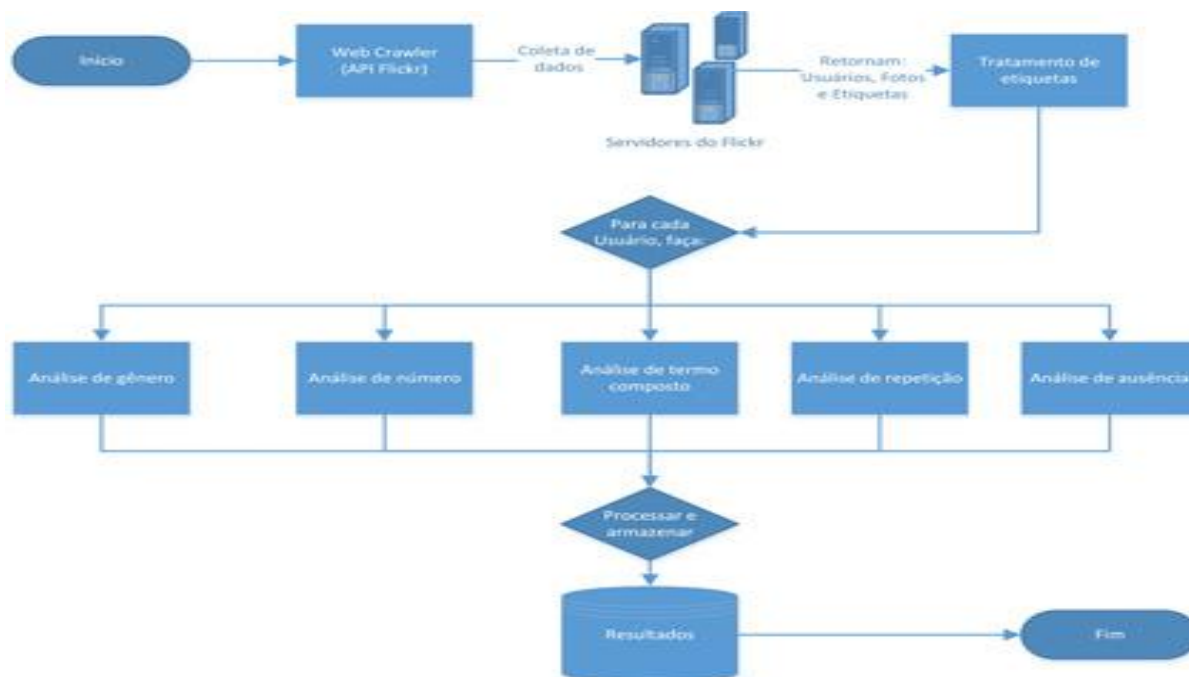
Fonte: Azeredo (2009).

Para os **termos compostos**, dois tipos de identificações foram feitas. A primeira é a

identificação de etiquetas hifenizadas como, por exemplo, joão-de-barro. A segunda é a identificação de etiquetas por espaços brancos como, por exemplo, Belo Horizonte. Para cada termo composto identificado, procurou-se por etiquetas que tivessem partes separadas do termo na mesma foto. A identificação de **repetição** e **ausência** de etiquetas foram feitas de maneiras distintas. A ausência de etiqueta é considerada em fotos com nenhuma etiqueta. E a repetição é caracterizada por um usuário que repete pelo menos uma etiqueta em, pelo menos, duas fotos.

4.3 ROTINA DE ANÁLISE DE DADOS

A rotina de análise de dados considera o usuário como papel central de atuação. Para cada tipo de análise, serão verificados dados de um mesmo usuário, podendo usar uma única foto ou pares de fotos. Usuários diferentes foram comparados entre si. Para análise sintática de uso de gênero, número e termos compostos, os comportamentos são verificados foto a foto. Ou seja, dentre todas as etiquetas colocadas em uma foto, são verificadas se o usuário colocou etiquetas diferentes para cada flexão. Por exemplo, em uma mesma foto o usuário coloca as etiquetas “azul” e “azuis”. Caso seja identificado esse comportamento, o algoritmo marca positivo para aquele caso naquela foto. A rotina de análise de repetição de etiquetas é feita por pares de fotos. Nela é verificada se o usuário repete pelo menos uma etiqueta em, pelo menos, duas fotos. Se o algoritmo detectar a repetição de etiqueta, então ele é interrompido e o usuário é marcado como positivo para o teste de repetição. Por fim, a rotina para análise de ausência é dada como positiva se uma foto não possui nenhuma etiqueta atrelada a ela. Foram verificados usuários que não colocam etiquetas em, pelo menos, uma foto e também usuários que não colocam etiquetas em nenhuma foto. Uma versão resumida da rotina pode ser encontrada na Figura 1:

Figura 1 – Fluxo de execução da metodologia.

Fonte: Elaborado pelos autores (2018).

4.4 ANÁLISE DAS FOTOS

A análise das fotos foi feita com uma comparação entre as recomendações do NISO e as etiquetas analisadas gramaticalmente. O objetivo dessa análise é encontrar más práticas de uso de sistemas colaborativos usando os padrões de etiquetas usados pelos usuários. As recomendações de número e gênero são para que apenas uma variação dessas etiquetas estejam presentes nas fotos. Ou seja, etiquetas flexionadas na mesma foto estão incorretas segundo as diretrizes de etiquetagem para sistemas de indexação e recuperação de informação.

As recomendações para termo compostos é que eles não sejam separados em etiquetas diferentes, pois sua separação pode ter um significado diferente do termo composto. Portanto fotos com etiquetas que tenham essa prática identificada serão consideradas incorretas segundo as recomendações do NISO.

O cálculo de fotos que são etiquetadas incorretamente é dada por meio de dois cenários distintos. O primeiro considera o pior cenário, onde os conjuntos analisados são distintos entre si. Nesse cenário, o resultado será a soma dos valores encontrados a um limite de 100% da amostra. O segundo caso representa o melhor cenário, onde todos os conjuntos menores estão contidos nos conjuntos maiores analisados. Nesse cenário, o resultado é composto pelo valor da maior amostra encontrada. A média será calculada pela soma dos dois

cenários.

4.4 ANÁLISE DOS USUÁRIOS

A análise usuários foi feita utilizando os critérios definidos por Cañada (2012). Considerou-se: (a) altruístas os usuários que flexionam gênero, número e separam termos compostos; (b) egoístas os usuários que não colocam etiquetas em, pelo menos, uma foto; (c) amigáveis os usuários que repetem as etiquetas.

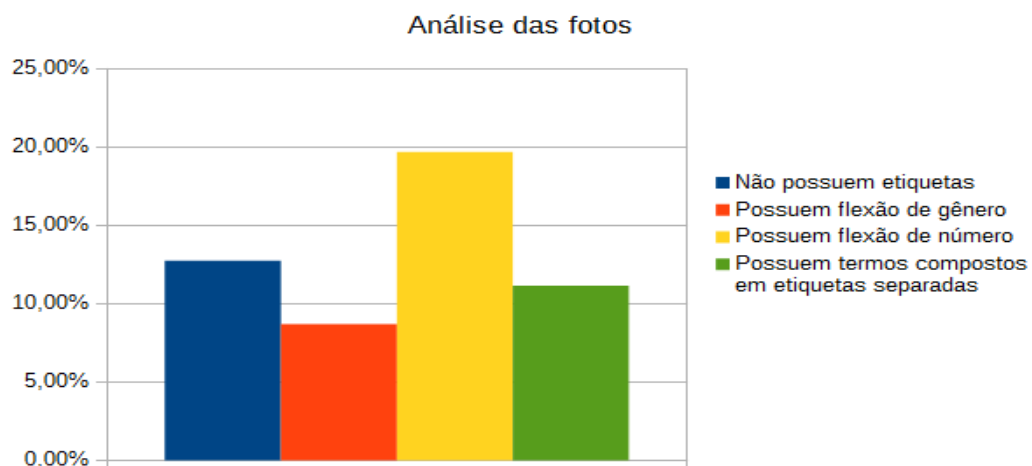
Note que a análise de etiquetas populares exigem que os documentos sejam analisados para saber se há ou não intuito apelativo por audiência. Essa análise não foi contemplada no presente estudo. O cálculo de classificação de usuários segue o mesmo usado para classificação de fotos.

5 RESULTADOS

A coleta de dados analisou ao todo 113.842 etiquetas distribuídas entre as 9000 fotos recuperadas pelo sistema. Ao todo obteve-se uma média de 12,64 etiquetas por foto. Foram feitas análises de fotos e usuários de acordo com as estratégias definidas nas subseções 4.4 e 4.5. Também foram realizadas análises primárias dos dados contidos nas fotos. Os resultados encontrados foram: (a) 12,73% das fotos não possuem nenhuma etiqueta; (b) 8,69% das fotos possuem pelo menos duas etiquetas que flexionam gênero; (c) 19,69% das fotos possuem pelo menos duas etiquetas que flexionam número; (f) 11,13% das fotos possuem etiquetas que separam outra etiqueta que é um termo composto.

A Figura 2 mostra graficamente a disposição dos resultados encontrados:

Figura 2 – disposição gráfica da análise de etiquetas coletas no Flickr.

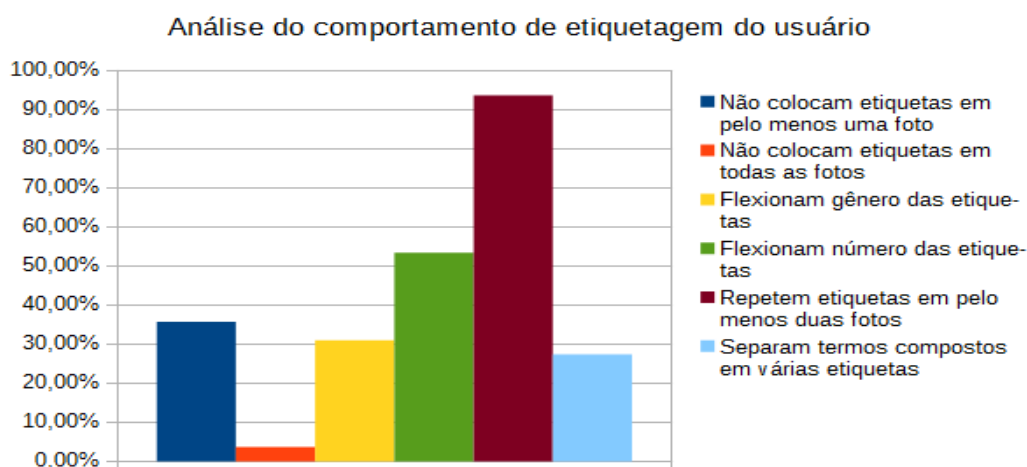


Fonte: Elaborado pelos autores (2018).

Em seguida foram feitas análises referentes ao comportamento dos usuários. A análise encontrou que: (a) 35,67% não colocam etiquetas em, pelo menos, uma foto; (b) 3,67% não colocam etiquetas em nenhuma foto; (c) 31% flexionam o gênero das etiquetas em suas fotos; (d) 53,33% flexionam o número das etiquetas em suas fotos; (e) 93,67% repetem uma etiqueta em suas fotos; (f) 27,33% separam os termos compostos em várias etiquetas.

A Figura 3 mostra graficamente a disposição dos resultados encontrados:

Figura 3 – Disposição gráfica do comportamento dos usuários no Flickr



Fonte: Elaborado pelos autores (2018).

5.1 RESULTADOS PARA ANÁLISE DE FOTOS

Considerando a análise das fotos descritas na subseção 4.4 temos que: (a) no pior cenário possível, o número de problemas de etiquetagem correta para fins de recuperação de informação representam 52,24% da amostra; (b) no melhor cenário possível são 19,69% de fotos com problemas; (c) a média de fotos com problemas representam 35,96% da amostra.

5.2 RESULTADOS PARA ANÁLISE DE USUÁRIOS

Podemos considerar os **usuários altruístas**, aqueles que flexionam etiquetas com intuítos de aumentar a recuperação de informação: (a) o melhor é composto de 100% dos usuários que possuem atos altruístas em, pelo menos, uma foto; (b) o pior cenário é composto por 53,33% dos usuários; (c) o cenário médio é de 76,66% de usuários altruístas.

Os resultados encontrados para **usuários amigáveis** foram de que 93,67%, levando em consideração que eles repetem etiquetas para recuperação pessoal ou para participar de comunidades específicas.

Por fim, os resultados encontrado para **usuários egoístas** são de: (a) 3,67% para

usuários extremamente egoístas, ou seja, não se preocupam com o sistema de recuperação de informação; (b) 35,67% de usuários que foram egoístas em, pelo menos, uma de suas fotos que foram analisadas.

5.3 ANÁLISE DO IMPACTO SOCIAL DO COMPORTAMENTO DO USUÁRIO

Os sistemas de recuperação de informação visam aproximar o usuário do objeto. Para que um conteúdo seja recuperado, é necessário que esses sistemas entendam as buscas realizadas pelos usuários. No entanto, a operação por trás dessas buscas denota rigidez muitas vezes excessiva, sem dar margens à flexibilidade que o instrumento poderia oferecer ao usuário. Por exemplo, um usuário que usa o Flickr para pesquisar a etiqueta “zangão” terá resultados diferentes se buscar por “abelha” ou “abelhas”. Considerando-se que todas as etiquetas estão intimamente relacionadas, é provável que o resultado das buscas das três etiquetas interessem ao usuário,

O papel social dos usuários dentro desses sistemas colaborativos é auxiliar na aproximação do objeto intencionado por quem faz a busca. Eles alimentam o banco de dados com opções extras de recuperação de informação para que os resultados relevantes cheguem até quem os procura.

A flexão de gênero e número colabora para que os sistemas de recuperação de informação sejam mais maleáveis na hora de retornar as buscas. Nesse sentido, 31% e 53,33% dos usuários da amostra colaboram na aproximação de outros usuários com os objetos dentro do sistema de recuperação de informação. Ao analisar as atitudes altruístas esse número sobe para, em média, 76,66%. Observa-se que o proprietário do objeto deseja que ele seja recuperado por alguém e tornar essa recuperação facilitada parece natural.

A repetição de etiquetas entre diferentes fotos ajudam a identificar uma comunidade em torno de interesses mútuos. Ela muitas vezes é intencional e preestabelecida em um grupo social. Por exemplo, a etiqueta “CatsOfInstagram” serve para aproximar pessoas com um mesmo interesse em comum: ver e postar fotos de gatos no Instagram. Mais de 90% da amostra reflete esse comportamento, que é tanto para fazer parte de um nicho, quanto pra criar seu próprio nicho de conteúdo. Esses usuários, considerados amigáveis, têm o papel social de aproximar objetos diferentes, mas de interesse comum, do usuário que busca um conteúdo de nicho.

Existem, no entanto, aspectos negativos a serem observados desse comportamento. A criação de nichos serve como palco para o movimento do filtro bolha, onde os sistemas analisam o comportamento do usuário para filtrar o tipo de informação que chega até ele e

deixá-lo permanentemente dentro de um nicho social (PARISER, 2012). A separação de termos compostos em várias etiquetas traz o aumento da revocação de sistemas sem, necessariamente, apresentar, ao usuário, resultados relevantes. Além disso, o número de fotos da amostra com problemas é de 35,96% e o número de usuários extremamente egoístas é de 3,67%. Essas atitudes afastam o usuário do objeto e produz conteúdo que nunca mais vai ser recuperado pelos sistemas de informação.

6 DISCUSSÃO E TRABALHOS FUTUROS

Os resultados mostram que os problemas encontrados nas fotos possuem uma chance média de 35,96% de acontecer. Esse número, que no pior caso pode ser de 52,24%, mostra que os sistemas colaborativos estão longe de ser um sistema de recuperação ideal para documentos criados pelos usuários. No entanto, o cenário contrário pode ser visto de forma otimista. Isso indica que a média de fotos sem problemas em suas etiquetas é de 64,04%. Os sistemas colaborativos, ainda que imperfeitos, podem sim ter boa parte de seus documentos recuperados por outros usuários. Os resultados reforçam as demais análises sobre o assunto que foram discutidos na seção 3.

Os resultados que analisaram o comportamento do usuário identificaram que o comportamento altruísta é visto em média de 76,66% dos usuários, o que é bom para a manutenção desses sistemas. A taxa de usuários amigáveis é de 93,67%. Considerando as análises de Cañada (2012), ambos esses tipos de usuários contribuem para o sistema colocando etiquetas que podem ser recuperadas por usuários do sistema ou comunidades específicas. Os usuários extremamente egoístas representam apenas 3,67% da amostra, um valor considerado bom para os padrões de qualidade em sistemas colaborativos. Já os usuários que tiveram comportamento egoísta em, pelo menos, uma de suas fotos, o número sobe 35,67%.

No entanto, vale ressaltar que o comportamento egoísta não é predominante entre os usuários do Flickr. A incidência de, pelo menos, uma foto sem etiquetas é comum nesses sistemas, uma vez que eles também servem como plataformas de backup para os usuários. Nesse caso, o usuário não enxerga o sistema colaborativo como uma plataforma de indexação de documentos públicos e sim como um espaço particular para salvar seus dados. Os resultados encontrados no presente estudo para o comportamento do usuário corroboram com os achados na seção 3. O uso principal dos sistemas colaborativos como plataformas de recuperação de informação ainda é maioria entre os usuários. Apesar de eles não seguirem estritamente as recomendações de indexação sugeridas por órgãos competentes, eles ainda se

preocupam em descrever corretamente o conteúdo que eles produzem nesses sites. A análise de flexão gramatical para verificar tanto o padrão de etiquetas quanto dos usuários apresentou uma forma diferente de avaliação. É importante ressaltar que a flexão de gênero e número nem sempre contribuem com a recuperação de informação e precisam ser evitadas. A separação de termos compostos em várias etiquetas também podem ser problemática, pois os termos separados podem ter significados diferentes do termo composto.

Porém, esse comportamento pode ser interpretado como uma mistura de altruísmo e populismo. Os usuários querem ter um alcance maior de seus documentos, mas sem perder a essência da classificação correta. A repetição de etiquetas em fotos diferentes traz um comportamento preocupado do usuário em recuperar suas próprias fotos e, ao mesmo tempo, estabelecer-se em comunidades que se identificam por meio de etiquetas específicas. Dessa forma, entende-se que as contribuições sociais para a recuperação de informação e aproximação do usuário com o conteúdo relevante são satisfatórias. A análise gramatical reforçou a identificação de um movimento altruísta em sistemas de informação, onde os usuários colaboram com o conteúdo ao mesmo tempo em que estabelecem relações profícuas com a comunidade.

Apesar de os resultados encontrados serem considerados satisfatórios pelos autores, a presente pesquisa encontrou uma série de limitações. Uma delas está na coleta de dados. Devido a automatização da coleta, não há como precisar se um usuário foi selecionado mais de uma vez, pois há a possibilidade que ele esteja enquadrado em diferentes temas que foram selecionados aleatoriamente.

Os algoritmos de análise foram construídos usando regras para as flexões padrões que são encontradas na língua portuguesa. Casos específicos de flexão, tanto de gênero quanto de número foram excluídos da análise. Portanto sugere-se como trabalhos futuros a incorporação dessas análises como um todo. É preciso avaliar se esses casos podem interferir na percentagem dos grupos. Outra análise que é sugerida, é a verificação desse comportamento em outros sistemas colaborativos além do Flickr. Será que o mesmo padrão é identificado em redes diferentes?

Conclui-se que os sistemas colaborativos ainda é assunto de relevância para ser estudado, tanto pela Ciência da Informação, quanto pela Ciência da Computação. A tecnologia cada vez mais acessível e a participação dos usuários nesse novo movimento de produção de documentos transforma o cenário de indexação e recuperação. As etiquetas são parte importante desse processo e precisam ser estudadas.

REFERÊNCIAS

AUDRING, J. Gender assignment and gender agreement: evidence from pronominal gender languages. **Morphology**, Springer, v. 18, n. 2, p. 93–116, 2008.

AZEREDO, J. C. d. **Gramática Houaiss da língua portuguesa**. [S.l.]: Publifolha, 2009.

BAFOUTSOU, G.; MENTZAS, G. Review and functional classification of collaborative systems. *International journal of information management*. Elsevier, v. 22, n. 4, p. 281–305, 2002.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval: the concepts and technology behind search** (ACM Press Books). [S.l.]: Addison-Wesley Professional Harlow, 2011.

BARBOSA, F.; KREBS, Luciana Monteiro; SOUSA, Rodrigo Silva Caxias de. Folksonomia: análise de etiquetagem de imagens da national geographic brasil no instagram. **Informação & Informação**, v. 23, n. 3, 2015. Disponível em:

CAÑADA, J. **Tipologías y estilos en el etiquetado social**. [S. l. : s. n.], 2012.

GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*. Elsevier, v. 35, n. 2, p. 137–144, 2015.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GUY, M. *et al.* Tidying up tags. **D-lib Magazine**, v. 12, n. 1, 2006.

KOTLER, P. Prosumers: a new type of consumer. **The Futurist**, v. 20, p. 24–28, 1986.

KREBS, L. M.; LAIPELT, R. C. F.; ROSA, S. S. da. O uso da folksonomia na atualização de vocabulários controlados da área da pediatria. **Prisma. com**, n. 36, p. 59–77, 2018.

LAPLANTE, A. Tagged at first listen: An examination of social tagging practices in a music recommender system. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 20, n. esp.1, p.33 – 54, fev. 2015.

MANINI, M. P. **Análise documentária de fotografias: um referencial de leitura de imagens fotográficas para fins documentários**. 2002. Tese (Doutorado em Ciência da Informação e Documentação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2002.

MASSONI, L. F. H.; FLORES, A. B. A cidade representada em tags: Explorando a folksonomia no flickr. **PontodeAcesso**, v. 11, n. 3, p. 133–147, 2018.

NISO, N. I. S. O. U. et al. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. [S. l.]: NISO Press, 2010.

NÓBREGA, Isabella de Oliveira; MANINI, Miriam Paula. #impeachment ou #naovaitergolpe: uma análise sobre a folksonomia na indexação de imagens fotográficas em

redes sociais da web 2.0. **Biblioline**, v. 12, n. 4, 2016.

O'REILLY, T. What is web 2.0. **O'reilly**, 30 set. 2005. Disponível: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>. Acesso em:

PANOFSKY, E. **Significado nas artes visuais**. São Paulo: Perspectiva, 2009.

PARISER, E. **O filtro invisível: o que a internet está escondendo de você**. Rio de Janeiro: Zahar, 2012.

PETERS, I. Folksonomies. Indexing and retrieval in Web 2.0. **De Gruyter Sauer**, 2009. DOI: <https://doi.org/10.1515/9783598441851>

PIRAQUIVE, F. N. D.; AGUILAR, L. J.; GARCÍA, V. H. M. Taxonomía, ontología y folksonomía, ¿ qué son y qué beneficios u oportunidades presentan para los usuarios de la web? **Universidad & Empresa**, v. 11, n. 16, p. 242–261, 2009.

SANTOS, H. P. Etiquetagem e folksonomia: o usuário e sua motivação para organizar e compartilhar informação na web 2.0. **Perspectivas em Ciência da Informação**, v. 18, n. 2, p. 91–104, 2013

SEN, S. *et al.* Tagging, communities, vocabulary, evolution. 20TH ANNIVERSARY CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK: ACM, 2006 Nova York. **Proceedings [...]**. Nova York, 2006, p. 181–190. <https://doi.org/10.1145/1180875.1180904>

SILVA, L. M. d. C. Folksonomia: análise das tags no site Last. fm. **RI FURG**, 2015.

SPITERI, L. F. The structure and form of folksonomy tags: The road to the public library catalog. **Information technology and libraries**, v. 26, n. 3, p. 13–25, 2007.

WAL, T. V. Folksonomy definition and Wikipedia. **vanderwal.net**, 2 Nov. 2005. Disponível em: <http://www.vanderwal.net/random/entrysel.php?blog=1750>

WAL, T. V. Folksonomy coinage and definition. **vanderwal.net**, 2 Feb. 2007. Disponível em: <http://vanderwal.net/folksonomy.html>