

A folksonomia do YouTube¹

Lúcio Siqueira Amaral Filho²

Resumo

Uma das características marcantes da internet é a enorme quantidade de informações presentes na rede. Isso porque a web 2.0 exponencializa a possibilidade dos usuários construírem seus próprios conteúdos, publicando e compartilhando-os com os outros usuários da rede. Com o aumento do número de informações cresce a dificuldade de encontrá-las, sendo necessário uma organização dessas informações. Surgem, assim, as *tags* (etiquetas), palavras associadas a conteúdos na rede para facilitar sua futura recuperação. A partir disso, este trabalho pretende delinear algumas concepções teóricas a respeito da folksonomia (etiquetagem) de vídeos no YouTube, onde usuários utilizam palavras-chave para que seus vídeos sejam facilmente encontrados em futuras buscas. Para tanto, foi feita uma pesquisa no site pelas expressões "*ipod ad*" e "*ipod commercial*", com o intuito de saber quais os critérios de etiquetagem dos vídeos mais visualizados dos resultados.

Palavras-chave: folksonomia, YouTube, web 2.0, iPod, vídeo.

Abstract

A key feature of internet is the huge of information on the web. This is because the web 2.0 exponentiating the ability of users build their own content, publishing and sharing them with other users on the network. With the increasing of information, grows difficulty to find them, requiring an organization of such information. Thus, appears the tags, words associated with the web content to facilitate their future recovery. From this, this paper attempts to outline some theoretical concepts about the folksonomy of videos on YouTube, where users use keywords to your videos to found them easily in future searches. The work was done a site search with the words "ipod ad" and "ipod commercial", in order to know the criteria for tagging the most viewed videos.

Keywords: folksonomy, YouTube, web 2.0, iPod, video.

¹ Artigo científico apresentado ao eixo temático "Cibercultura", do III Simpósio Nacional da ABCiber. O presente trabalho faz parte da dissertação de mestrado que se encontra em fase de elaboração, sob orientação do prof. Dr. Francisco Menezes.

² Mestrando em Comunicação e Linguagens da Universidade Tuiuti do Paraná - UTP, professor do Curso de Publicidade e Propaganda da Celer Faculdades e coordenador da Agência Experimental Looping, da mesma instituição. E-mail: lucio@lucioamaral.com.br.

Panorama

O atual estágio em que se encontra a internet foi chamado por Tim O'Reilly (2005) de web 2.0 no ano 2000, onde ele declara que após a explosão da *bolha.com* a internet tornou-se mais importante do que nunca, sendo utilizada como plataforma, pois cada usuário pode controlar seus dados através da arquitetura da participação da rede. Antes disso, a internet era baseada no sistema um-para-todos, onde um usuário publicava suas informações para todos aqueles que acessavam seu site.

Com a amplificação da participação na rede, potencializada pela web 2.0, passou a existir uma participação ativa e constante dos usuários. Isso possibilitou a distribuição e o gerenciamento de conteúdos on-line, criados por qualquer tipo de usuário (sejam eles vídeos, fotos, textos, poemas, pensamentos, desejos, músicas, entre tantas outras criações) quase que instantaneamente, pois todos os usuários tornaram-se possíveis produtores de conteúdo. Isto é, o que antes só poderia ser feito através de programas instalados em computadores, agora pode ser elaborado remotamente, no ambiente virtual. Maria Clara Aquino (2009) afirma que essa publicação de conteúdos potencializa a "liberdade de emissão", fundindo os papéis de emissor e receptor da informação por meio dessas novas ferramentas tecnológicas interacionais.

Apesar de ser o principal diferencial entre essas duas gerações da web, a participação não é um fenômeno novo quando se trata de mídias. A televisão e o rádio utilizam-se dela há muito tempo, quando os ouvintes ligam para a rádio relatando algum tipo de problema que a sua comunidade está enfrentando; ou até mesmo em participações ao vivo nos programas de televisão, com espectadores ligando para fazer perguntas à pessoa que está sendo entrevistada naquele momento. Outra forma dessa participação do ouvinte é através de ligações com pedidos de músicas para a rádio. Com o passar do tempo e com o avanço das tecnologias da informática, essas ligações foram lentamente substituídas pelos e-mails e pelas mensagens enviadas por programas síncronos e assíncronos³.

Essa evolução dos processos de participação e colaboração cria, segundo André Lemos (2004), um ambiente que faz com que haja um agrupamento de pessoas com interesses comuns para trocas de informações, músicas, fotos, etc., tornado-se, assim, um fenômeno social. Esse ambiente pode ser representado pelo constante crescimento

³ Programas assíncronos são os de correio eletrônico, onde existe um significativo espaço de tempo separando a mensagem enviada e a mensagem recebida; já os síncronos são aqueles em que a mensagem recebida e a enviada são separadas por um pequeno espaço temporal, como o MSN Messenger, o Skype e o ICQ.

do número de blogs pra publicação de opiniões, informações e produção de conteúdos; pelo Flickr ou Picasa, onde os usuários hospedam suas fotografias e imagens, existindo a possibilidade de outras pessoas poderem fazer comentários a respeito delas; pela Last.fm, uma rádio online que afirma ter o maior catálogo de músicas da internet, que conecta usuários de acordo com os seus gostos musicais, funcionando como um sistema de recomendações que permite a troca de música entre eles; pela Wikipédia, uma enciclopédia escrita de maneira colaborativa; pelo YouTube, o maior site de hospedagem de vídeos da rede; entre tantos outros exemplos.

Muitas informações, e cada vez mais

Com o aumento da publicação de conteúdos pelos usuários, o número de informações presentes na rede também teve um acréscimo, ocasionando assim um excesso de informações. A lógica é simples: quanto mais usuários emissores, maior será o número de mensagens. Em 2005, Gulli e Signorini (in FRAGOSO, 2007), calcularam pelo menos 11,5 bilhões de páginas indexáveis, ou seja, aquelas páginas que são acessíveis às ferramentas de buscas. Conforme pesquisa realizada pela empresa americana Netcraft, no mês de agosto de 2009 existiam 225.950.957 sites ativos na internet⁴.

Com o crescimento do número de informações cresce também a dificuldade de saber onde encontrá-las e de como chegar até elas, “lá há o que não sei onde nem como encontrar” (VAZ, 2008, p. 228). A partir disso surge uma necessidade ainda maior de localização das informações disponibilizadas na rede, tarefa realizada pelos buscadores.

Buscadores, ferramentas ou sistemas de busca são mecanismos que atuam recuperando informações na rede. Eles armazenam em seus bancos de dados os conteúdos dos sites presentes na rede e disponibilizam-nos para consultas. Ou seja, a busca não é realizada especificamente na internet, mas nos bancos de dados desses buscadores. A organização dos resultados é dada pela quantidade de ocorrências da expressão buscada pelo usuário. Os buscadores nada mais são do que sites que fornecem ao usuário fácil acesso às informações buscadas.

⁴ Disponível em http://news.netcraft.com/archives/2009/08/31/august_2009_web_server_survey.html.

Os robôs, utilizados pelos buscadores, são programas que automatizam a navegação na internet, fazendo o trabalho que seria impossível de ser elaborado por seres humanos. Eles navegam por todos os links da rede para então criar os bancos de dados dos buscadores. A eficiência do buscador se dá na sua capacidade de “vasculhar” o seu banco de dados e comparar os caracteres digitados na busca com aqueles presentes nele. Quanto mais rápida e melhor for essa consulta, trazendo resultados significativos no seu relatório, melhor será o mecanismo.

Segundo Montardo (2007), os links podem ter o mesmo sentido de autoridade, ou seja, quanto mais links direcionarem para uma determinada página, maior será o seu grau de importância. Ela traz um comparativo com artigos: quanto maior o número de referências a um dado artigo, melhor será sua autoridade. Fazendo a consideração que as referências, e também os links, devem ser de boa qualidade.

Conforme autores como Feitosa, Céndon, Lawrence e Giles, Antoniou e van Harmelen e Gulli e Signorini (in AMARAL e AQUINO, 2008), os buscadores possuem problemas da ordem da(s):

- a) pouca relevância nos resultados;
- b) pouca atualização em seus bancos de dados;
- c) diferentes formas de indexação e varredura de seus robôs;
- d) diferentes formas de organização dos resultados; e
- e) diferentes tamanhos das páginas indexadas, tempo de processamento e restrição nos termos das buscas.

Esse crescimento descontrolado da quantidade de informações na rede, aliado aos problemas encontrados em recuperá-las de uma forma prática e confiável, tornou necessário o desenvolvimento de técnicas que apurassem facilmente as pesquisas em meio a essa sobrecarga informacional da rede, auxiliando os usuários no processo de recuperação informacional. Amaral e Aquino (2008) ainda citam Dreyfus, afirmando que não existe uma hierarquia baseada em classes na organização da rede, e sim uma interconexão generalizada.

Etiquetar

A partir disso tudo, em 2003 o site del.icio.us⁵ foi criado com base nas *tags*⁶ feitas pelos usuários. Em cada favorito marcado pelos usuários, eles possuem a possibilidade de associar uma *tag*, ou seja, associar uma palavra-chave para identificar e futuramente recuperar esse favorito em uma possível pesquisa. Após o del.icio.us, o Flickr⁷ surgiu com a mesma proposta que o del.icio.us, de associar palavras-chaves, as *tags*, às fotos que os usuários disponibilizam no site. Dessa forma, as *tags* começaram a ser usadas para a busca de informações na internet.

Em 2004 Thomas Vander Wal criou o termo folksonomia para denominar as estruturas categóricas criadas pelos usuários. Segundo Wal (2007),

é o resultado do taggeamento livre que as pessoas dão a informações e objetos (qualquer coisa com URL) para sua futura recuperação. O taggeamento é feito em um ambiente social (usualmente compartilhado e aberto a terceiros). A folksonomia é criada pelo ato da etiquetagem da pessoa que está consumindo a informação.⁸

A principal característica, então, da folksonomia, é essa participação ativa do usuário nos processos de etiquetagem das informações, coisa apenas possibilitada com o desenvolvimento da web 2.0. Aquino (2007) chama de hipertexto 2.0 o novo hipertexto surgido pelos links formados pelas *tags*.

Gouvêa e Loh (2007) citam vantagens e desvantagens na utilização de *tags*. As vantagens seriam:

- a) recuperação de documentos;
- b) compartilhamento de conhecimento e forma de organização;
- c) facilidade para associar classes já existentes, classificando conteúdos pertencentes a mais de um assunto;
- d) alternativa para a web semântica⁹;
- e) definição de conteúdos secundários;
- f) criação descentralizada, causando uma estrutura menos rígida e mais flexível;
- g) facilidade de uso, visto que as *tags* são criadas pelos usuários.

Por sua vez, as desvantagens seriam:

⁵ Disponível em <http://delicious.com/>, é um site de social bookmarking, ou em uma tradução literal, marcadores sociais, que consiste em socializar os favoritos adotados pelos usuários do serviço.

⁶ Tags são palavras-chaves associados a informações.

⁷ Disponível em <http://www.flickr.com/>, é um site de compartilhamento de fotos.

⁸ Tradução livre do autor.

⁹ Web semântica é uma extensão da atual web em que a informação é dada com o seu real significado, possibilitando que pessoas e computadores possam trabalhar em uma boa cooperação. (BERNERS-LEE e MILLER, 2002)

- a) problemas nas relações de significados entre as *tags* e seus conteúdos;
- b) problemas ortográficos variados, desde a má grafia até erros de sintaxe;
- c) problemas nos usos das *tags* genéricas e específicas.

Apesar de suas desvantagens, a utilização da etiquetagem feita pelos usuários tem sido adotada em sites com os mais variados conteúdos, com o intuito de organizar e separar as informações para sua possível busca futura. Mas essa busca pela organização das informações não é uma necessidade surgida com a sobrecarga informacional presente na rede. Ela já vem desde os primórdios da humanidade, quando o homem sentiu a necessidade de identificar as coisas presentes no mundo. Com o tempo e a evolução dos estudos, foi criado um método para organizar e sistematizar informações, a taxonomia. Ela é

uma classificação de entidades em forma hierárquica, de acordo com relações presumidas a partir dos objectos do mundo real que tais entidades representam. Uma ontologia deste tipo apresenta os termos numa hierarquia que apresenta relacionamentos entre os objectos e classes, subclasses e classes-pai (GUARINO, WELTY, 1998 citado por CATARINO, 2009, p. 71).

A folksonomia, portanto, é uma derivação, ou até uma evolução da taxonomia. Sua origem, na língua inglesa, significa uma classificação dada pelos usuários (*folks* = pessoas + taxonomia).

O YouTube

YouTube tem origem na língua inglesa. *You* = você, *tube* = tubo, uma gíria usada para a televisão. Uma livre tradução para o português ficaria “Você Televisão” ou “Televisão Você”, dando a entender que qualquer pessoa pode ser capaz de ter seu canal de televisão, com os vídeos que desejar, fazer a sua transmissão televisiva. Seu endereço eletrônico é www.youtube.com.

O YouTube¹⁰ é um site de publicação e compartilhamento de vídeos que permite aos usuários e visitantes assistir, compartilhar e tecer comentários a respeito dos vídeos ali hospedados. Fundado em fevereiro de 2005, ele se intitula a comunidade de vídeos on-line mais popular do mundo, possibilitando que as pessoas se conectem,

¹⁰ YouTube tem origem na língua inglesa. *You* = você, *tube* = tubo, ou gíria para televisão. Uma livre tradução para o português ficaria “Você Televisão” ou “Televisão Você”, dando a entender que qualquer pessoa pode ser capaz de ter seu canal de televisão, com os vídeos que desejar. Seu endereço eletrônico é www.youtube.com.

informem e inspirem, além de atuar como uma plataforma de distribuição de conteúdos. Em novembro de 2006 o YouTube foi adquirido pelo Google¹¹, e complementa a missão deste, de organizar e disponibilizar as informações do mundo para todas as pessoas.

Informação e entretenimento são as principais características do site, haja vista que pode ser encontrada uma imensa variedade de tipos e categorias de vídeos, que vão desde programas exibidos em redes de televisão e nele publicados¹², até vídeos amadores de momentos em família¹³, passando por flagrantes¹⁴ e mesmo séries¹⁵ criadas para veiculação exclusiva no YouTube. Com isso, ele acaba servindo como uma espécie de repositório da memória audiovisual de várias épocas, uma vez que muitas pessoas vão até ele para buscar vídeos dificilmente encontrados em outros locais (PAIVA, 2008). Com ele pode-se reviver os tempos saudosos de quando se criança, assistindo aqueles programas e séries adorados na infância, conhecer os comerciais que inauguraram a fase de ascensão e reconhecimento mundial da publicidade brasileira, rever episódios marcantes daquela novela quase esquecida e seus atores que já não atuam mais, recuperar momentos emblemáticos da história brasileira e mundial, entre tantas outras possibilidades.

O YouTube, acima de tudo, é possibilidade. Possibilidade de resgatar e assistir conteúdos audiovisuais antigos; possibilidade de ver o episódio da novela ou do seriado favorito perdido; possibilidade de mostrar para familiares e amigos distantes o vídeo do filho caminhando pela primeira vez; possibilidade de se fazer visível para o mundo; possibilidade de se tornar autor do seu próprio conteúdo audiovisual. Os usuários tornam-se, agora, cada vez mais atores e produtores de narrativas audiovisuais. Eles podem criar aquele comercial da sua marca favorita que nunca foi feito, mostrando a sua capacidade criativa e técnica. Basta apenas ter um dispositivo

¹¹ O Google é um site de buscas na internet. A sua missão "é oferecer a melhor opção de busca na Internet tornando as informações mundiais acessíveis e úteis. O Google, desenvolvedor do maior mecanismo de busca do mundo, oferece o caminho mais rápido e fácil de encontrar informações na web. Com acesso a mais de 1,3 bilhão de páginas, o Google oferece resultados relevantes para usuários de todo o mundo, normalmente em menos de meio segundo. Hoje, o Google responde a mais de 100 milhões de consultas por dia." (SOBRE..., 2009)

¹² Matéria sobre o adiamento das provas do Enem/09, veiculada no Jornal Nacional de 01/10/2009. (<http://www.youtube.com/watch?v=G2WQjihjVvk>)

¹³ Menino de 7 anos voltando do dentista com seu pai, no carro, ainda sob o efeito da anestesia. (<http://www.youtube.com/watch?v=txqiwrbyGrs>)

¹⁴ Daniela Cicarelli flagrada fazendo amor com seu então namorado em uma praia da Espanha. (<http://www.youtube.com/watch?v=2ZxzCCEqjI0>)

¹⁵ O que é isso?, uma série produzida exclusivamente para o YouTube, com 16 websódios em 2 temporadas. (<http://www.oquequeeisso.com.br>)

digital de captura de imagens para produzir o vídeo e enviá-lo para o YouTube, que todas as pessoas, de qualquer parte do mundo terão acesso a ele.

As tecnologias digitais, ao favorecerem a produção de conteúdo, incentivam o indivíduo a esquadrihar as diversas mídias em busca da informação desejada e a engendrar um processo de colaboração entre indivíduos que se reúnem em comunidades virtuais, listas de discussão, *blogs* para buscar, produzir e partilhar informações adicionais sobre seus produtos culturais favoritos (RÉGIS, 2008, p. 35-36).

A folksonomia do YouTube

Para atender os objetivos deste trabalho, foram definidos os termos “*ipod ad*” e “*ipod comercial*” para serem utilizados em uma busca por vídeos no site global do YouTube¹⁶. A opção pela língua inglesa deve-se ao fato dela ser a mais utilizada no mundo todo. O resultado para “*ipod ad*” retornou cerca de 1.020 vídeos, enquanto que “*ipod comercial*” retornou 2.860. Para refinar e diminuir os resultados, a fim de facilitar o trabalho, foram escolhidos os cinco vídeos mais visualizados de cada busca. *Ipod ad*: iPod Touch Ad - Nick Haley (2.275.958 visualizações); New Apple iPod Nano ad - 1234 by Feist (2.050.259 visualizações); Ipod Ad (1.080.933 visualizações); Napoleon Dynamite iPod Ad (768.524 visualizações); iPod + iTunes TV Ad 1 of 2 (699.434 visualizações). *Ipod comercial*: iPod Commercial - Shake it! Metro Station (2.370.650 visualizações); iPod Commercial - Disturbia Rhianna (1.910.548 visualizações); Black Eyed Peas - Hey Mama iPod Comercial (613.956 visualizações); Jerk It Out (586.198 visualizações); The Fratellis - Flathead iPod Commercial (433.708 visualizações).

Em uma primeira análise constata-se que IA tem mais visualizações que IC¹⁷, 6.875.108 perante 5.915.060. De todos os dez vídeos nenhum de IA se repete em IC; quatro deles não são comerciais ditos “oficiais”, sendo que três são paródias (Napoleon Dynamite iPod Ad, iPod Commercial - Shake it! Metro Station e iPod Commercial - Disturbia Rhianna) e um é um clipe musical (Jerk It Out) que nada tem a ver com iPod, a não ser a informação colocada por seu autor: “Better than iPod commercial”. O que chama atenção nos resultados é que os dois vídeos mais visualizados de IC são as paródias do usuário *WasteTimeChasingCars*, muito mais vistos que os outros três.

¹⁶ As buscas foram realizadas no dia 4 de outubro de 2009.

¹⁷ A partir daqui, os resultados da busca por “iPod ad” serão nomeados simplesmente IA, enquanto que os por “iPod comercial” serão IC.

Wander Wal classifica a folksonomia em dois tipos: largas e estreitas. As largas, utilizadas pelo del.icio.us, são aquelas em que muitos usuários utilizam um número pequeno e comum a todos de *tags* para os objetos. Neste tipo, quem etiqueta os objetos não são seus criadores, mas os usuários que o acessam. Já as estreitas dizem respeito a uma ou um pequeno número de pessoas etiquetando objetos para sua posterior busca. O maior representante é o Flickr, onde cada *tag* utilizada é singular, cabendo ao criador do objeto a sua etiquetação.

O YouTube apresenta uma folksonomia estreita, visto que no momento em que o usuário envia um arquivo para o site ele tem a possibilidade de adicionar uma *tag* ao vídeo. Caso não seja adicionada nenhuma *tag* pelo usuário, automaticamente o nome do vídeo é utilizado como *tag* pelo YouTube. Abaixo são demonstradas as *tagclouds*¹⁸ de IC e IA¹⁹.



Figura 1 - Tagcloud de IA

¹⁸ Tagcloud, ou nuvem de tags, é uma visualização de tags, onde as mais utilizadas possuem tamanho e cores que as destacam das demais.

¹⁹ Tagclouds criadas a partir do site <http://www.wordle.net/create>.

Referências bibliográficas

AMARAL, A, AQUINO, M.C. Práticas de folksonomia e social tagging no Last.fm. In: *Anais do IHC 08 - VIII Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais*, Porto Alegre, 2008. Disponível em: http://www.inf.pucrs.br/ihc2008/pt-br/assets/files/Praticas_Folksonomia_Social_Tagging_Lastfm.pdf

AQUINO, Maria Clara. *Hipertexto 2.0, folksonomia e memória coletiva: um estudo das tags na web*. E-Compós, Brasília n. 9, nov. 2007.

AQUINO, Maria Clara. Os blogs na web 2.0: representação e recuperação coletivas de informação. In: AMARAL, Adriana; MONTARDO, Sandra; RECUERO, Raquel (orgs.). *Blogs.com: estudos sobre blogs e comunicação*. São Paulo: Momento Editorial, 2009. p. 237-256.

August 2009 Web Server Survey. Disponível em http://news.netcraft.com/archives/2009/08/31/august_2009_web_server_survey.html. Acessado em 22 de setembro de 2009.

BERNERS-LEE, Tim e MILLER, Eric. *The semantic web lifts off*. ERCIM News, n.51, 2002. Disponível em http://www.ercim.org/publication/Ercim_News/enw51/berners-lee.html. Acessado em 25 de setembro de 2009.

CATARINO, Maria Elisabete. *Integração da folksonomia nos metadados: identificação de novos elementos como contributo para a descrição de recursos em repositórios*. 2009. 257 f. Tese (Doutorado em Tecnologias e Sistemas de Informação) - Departamento de Sistemas de Informação - Universidade do Minho, Braga, Portugal. 2009.

FRAGOSO, Suely. *Quem procura acha? O impacto dos buscadores sobre o modelo distributivo da World Wide Web*. Disponível em <http://www.eptic.com.br/arquivos/Revistas/v.%20IX,n.3,2007/ASuelyFragoso.pdf>. Acessado em 21 de setembro de 2009.

GOULART, Rodrigo e MONTARDO, Sandra. *Histórico dos mecanismos de busca e suas implicações em comunicação e marketing*. Trabalho apresentado no GT História da Mídia Digital, no V Congresso Nacional de História da Mídia. São Paulo, 2007.

GOUVÊA, Cleber e LOH, Stanley. *Folksonomias: identificação de padrões na seleção de tags para descrever conteúdos*. RESI – Revista Eletrônica de Sistemas de Informação. Ed. 11, n. 2, 2007. Disponível em <http://revistas.facecla.com.br/index.php/reinfo/article/view/214/118>. Acessado em 25 de setembro de 2009.

GULI, A e A. SIGNORINI. *The Indexable Web is more than 11.5 billion pages*. International Conference on the WWW 2005, 10 a 14 de Maio, 2005, Chiba, Japão. Disponível em <http://www.cs.uiowa.edu/~asignori/web-size/size-indexable-web.pdf>. Acessado em 21 de setembro de 2009.

PAIVA, C. C. . *YOUTUBE como vetor de Modernização Tecnológica e Desenvolvimento Social*. ABCIBER. In: II Simpósio ABCiber. São Paulo: PUCSP, 2008. v. 1. p. 1-15.

RÉGIS, Fátima. *Tecnologia de comunicação, entretenimento e competências cognitivas na cibercultura*. Revista Famecos. Porto Alegre, 2008. p. 32-37.

VAZ, Paulo. *Mediação e tecnologia*. In: MENEZES, Francisco; MACHADO, Juremir (orgs.). *A genealogia do virtual: comunicação, cultura e tecnologias do imaginário*. Porto Alegre: Sulina, 2008. p. 216-238.

WAL, Wander. *Explaining and showing broad and narrow folksonomies*. Disponível em http://www.personalinfocloud.com/2005/02/explaining_and_.html. Acessado em 4 de outubro de 2009.

WAL, Wander. *Folksonomy coinage and definition*. Disponível em <http://www.vanderwal.net/folksonomy.html>. Acessado em 25 de setembro de 2009.