



A Importância da Indexação e Categorização para a Análise de Grande Volume de Dados¹

Veronica A. Ribeiro HAACKE²

Fabio Gomes GOVEIA³

Lia Scarton CARREIRA⁴

Universidade Federal do Espírito Santo, Vitória, ES

RESUMO

Um grande volume de dados, dentre textos, imagens e vídeos, é compartilhado cotidianamente pelos indivíduos nos sites de redes sociais. A partir deste contexto, encontramos uma série de estudos que buscam compreender esse fenômeno. Entre eles, está a pesquisa inserida no projeto Visagem, a qual busca analisar e compreender as imagens compartilhadas nos sites de redes sociais. Contudo, para que sejam realizadas análises detalhadas e profundas das imagens, em rede, se faz necessário o processo de indexação e categorização, como forma de organizar e facilitar as suas visualizações. Neste artigo, busca-se, portanto, apresentar alguns apontamentos sobre esses aspectos a fim de levantar questões e identificar sua importância para o estudo dos modos contemporâneos de produção e compartilhamento de dados na Internet, em especial ao que tange à imagem.

PALAVRAS-CHAVE: Indexação; Folksonomia; imagens; redes sociais.

INTRODUÇÃO

Com os avanços tecnológicos e a expansão da internet, o espaço comunicacional foi intensamente modificado. A capacidade de compartilhar informação que antes era de exclusividade dos meios comunicacionais ditos tradicionais (jornal impresso, televisão e rádio), passou a ser benefício de qualquer um que possua acesso à internet e às tecnologias básicas (computador, celular, tablets, máquinas fotográficas e etc). Assim, a informação que antes seguia o fluxo UM-TODOS passou a ser compartilhada pelo TODOS-TODOS, produzindo narrativas-monstro. Isto é, narrativas produzidas pela multidão, modo pelo qual se demonstra suas ideias e discursos. Dessa maneira, uma

¹ Trabalho apresentado no DT 1 – Jornalismo do XIX Congresso de Ciências da Comunicação na Região Sudeste, realizado de 22 a 24 de maio de 2014.

² Estudante de Graduação 5º. semestre do Curso de Jornalismo da Ufes, email: veronica.rhaacke@gmail.com

³ Orientador do trabalho. Professor do curso de comunicação social da Ufes, email: fabiogv@gmail.com

⁴ Co-orientadora e pesquisadora associada ao Labic Ufes, email: liacarreira@gmail.com



grande quantidade de conteúdo é produzido e compartilhado, gerando um grande volume de dados.

Os dispositivos móveis (aparatos tecnológicos portáteis como celulares e câmeras fotográficas) influenciaram largamente na quantidade desses dados. Eles facilitaram o registro daquilo que se vê e se experiencia, já que muitas pessoas carregam seus dispositivos cotidianamente a todo lugar. Registrar visualmente o que aconteceu, e que de alguma forma provocou a vontade de compartilhamento, se tornou algo mais cotidiano do que somente a fala e a escrita. Além disso, muitos celulares possibilitam o acesso à internet, fazendo com que esse compartilhamento de informação e conteúdo seja realizado imediatamente, concomitante ao acontecimento. Agora, qualquer um que tenha esses artefatos em mãos pode ser um indivíduo que se torna fonte de informação, produzindo e compartilhando conteúdos.

Além dos dispositivos móveis, os sites de redes sociais potencializaram um maior fluxo de compartilhamento desses conteúdos. Criados com a ideia de criar redes de pessoas, que possam compartilhar e trocar conteúdos e afetos, os sites de redes sociais se transformaram em um grande palco de publicações, com textos, áudios, vídeos e imagens. O usuário pode compartilhar, na rede online, entre uma foto e outra, sua opinião sobre algo que leu/viu/ouviu; publicar um texto seu ou de outra pessoa e etc. O usuário passou a ser compartilhador essencial de informação. Passou a ter mais liberdade para dizer o que pensava. Somando-se a isso, o usuário obteve a possibilidade de atingir um maior número de pessoas com as suas publicações e produzir um maior volume de conteúdo.

“Em seguida, vem a fase mais madura da web, a 2.0, caracterizada pelo nascer dos perfis, quando o usuário deixa de ser dono da sua “home” e passa a ter a sua própria “timeline”, passando a interagir fortemente na rede e publicar seus interesses nas redes sociais. [...] O surgimento dessas novas subjetividades, com o nascer dos blogs, redes sociais e portais de notícias, fazia emergir as transformações no jornalismo, com a urgência da notícia. Com o poder de publicação na mão de todos, muitas mudanças tornaram-se evidentes no rumo da história. Os conflitos da nossa época se organizariam com base na comunicação, na qual aqueles que estivessem configurados em rede, sairiam à frente.” (CALMON, Priscilla. 2013)

Diante dessas transformações de compartilhamento de informações, alguns estudos surgiram com a ideia de analisar esse fenômeno de grande volume de informações e o modo como são compartilhados nos sites de redes sociais. Essas análises são muito importantes, pois estudam os processos contemporâneos e habituais



de produção de conteúdo em . O acesso do indivíduo aos meios de produção e compartilhamento de informações, envolve mudanças tanto culturais como sociais. Afinal, o cotidiano das pessoas foi transformado com a oportunidade de se relacionar e compartilhar conteúdos com uma grande quantidade de indivíduos através da internet.

Um outro fator importante a ser estudado é como a imagem se faz importante na contemporaneidade. A produção de imagens teve um grande aumento, isso devido tanto às facilidades tecnológicas quanto aos meios de compartilhamento. Atualmente, existem aplicativos/sites próprios para o compartilhamento de imagens (como *Flickr* e *Instagram*), o que fomenta ainda mais essa veiculação de imagens. Assim, as imagens estão cada vez mais presentes como uma forma de “comprovação” do real, de que o indivíduo estava presente naquele momento, estava em determinado lugar ou até mesmo como maneira de mostrar o que vestiu ou comeu. E a imagem se torna ainda mais importante em um ambiente globalizado como a internet. Como ela não precisa de um idioma para ser “lida”, pode ser interpretada independentemente da sua nacionalidade, não limitando assim, territórios geográficos e linguísticos para seu entendimento.

Os sites de redes sociais, que já eram uma fonte alternativa de informação, tiveram grande relevância durante as manifestações iniciadas em junho de 2013 no Brasil. Além de boa parte das manifestações terem sido combinadas através de eventos do site *Facebook*, lá encontram-se as imagens, textos e vídeos de quem estava participando das manifestações e que compartilhava informações. No *Facebook*, além de outros sites como *Twitter* e o aplicativo *Instagram*, era possível ter uma visão mais ampla do que estava acontecendo. Muitos textos e imagens que eram postados nos sites de redes sociais influenciaram o modo de entendimento dos usuários sobre o fato que estava ocorrendo. A televisão e o jornal impresso, por muitas vezes apresentou apenas uma versão, um lado do acontecimento, provocando críticas e indignação dos participantes e ativistas. Enquanto isso, na Internet, havia uma gama de versões e desdobramentos do ocorrido, reafirmando sua multiplicidade de atores e visões. Foi possível, através da análise desses conteúdos, observar o ponto de vista de vários indivíduos, participantes ou não dos movimentos.

A partir de uma leitura desses dados, ao serem coletados e visualizados, podemos perceber como as imagens foram parte importante para a construção das narrativas das manifestações. Uma grande quantidade de imagens foi compartilhada, sendo elas em grande parte imagens convocatórias, fotos do próprio movimento, fotos das pessoas que participavam do protesto, além de mostrarem a violência e a repressão presentes nas



ruas. No *Twitter*, a *hashtag* “#VemPraRua”, que possuía o caráter convocatório continha mais de 88 mil imagens⁵ associadas em grande parte ao movimento.

Diante de tantas imagens compartilhadas, surge a ideia da pesquisa do projeto Visagem, inserido no Laboratório de estudos sobre Imagem e Cibercultura (Labic): analisá-las e poder estudar os seus padrões e as suas dissonâncias, quem são os usuários que compartilham essas imagens, como elas podem mostrar a importância dos sites de redes sociais para o compartilhamento de informações e como elas são capazes de conduzir a narrativa do contexto em que estão inseridas.

Parece que estamos sofrendo de excesso de informação ou abundância de dados. E a boa notícia é que pode haver uma solução fácil pra isso, e é usarmos mais os nossos olhos. E assim visualizando informação, para que possamos ver os padrões e conexões que importam e então projetar a informação para que faça mais sentido, ou conte uma história, ou nos permita focar apenas na informação que for importante (MCCANDLESS, 2010)

Este artigo tem como principal objetivo observar a importância da indexação e da categorização para a análise de um grande volume de dados. Com um maior fluxo de trocas de conteúdos na internet, mudanças na construção das narrativas contemporâneas e na quantidade de dados compartilhados ocorrem. Para a análise desse grande volume de dados, se faz necessário o processo de categorização, tanto como forma de facilitar a análise, como forma de possibilitar que esses dados sejam posteriormente retomados através de buscas usando as categorias. Por isso, mostramos a seguir algumas pesquisas que usaram a indexação e a categorização de dados, com ênfase na pesquisa de imagens do projeto Visagem.

HISTÓRICO E IMPORTÂNCIA DA INDEXAÇÃO DE DADOS

A atividade de indexação é uma prática bastante antiga, já que o tratamento da informação era realizado antes mesmo da denominação indexação. Pode-se dizer que a indexação surgiu com a atividade de criação de índices como instrumento de armazenagem para posterior recuperação da informação. Ela é uma importante ferramenta de busca, principalmente com fins em pesquisa documentária. A biblioteconomia é uma das áreas que usa a indexação como processo para categorizar e

⁵ A extração das imagens da hashtag foi realizada no período de um mês (15 de junho a 15 de julho). Para que não houvesse uma grande ocorrência de imagens desnecessárias, como peças de publicidade ou elementos gráficos do design do site, foram estabelecidos padrões para as imagens que seriam consideradas válidas: as imagens deveriam ter um tamanho mínimo de 200x200 pixels e um tamanho em disco de pelo menos 15 kb. Também só seriam aceitas imagens nos formatos .bmp; .jpg; .jpeg; .tiff e .tif para a extração.



organizar seu grande volume de documentos., por exemplo. Essa indexação de documentos e informações é realizada por pessoas especializadas e com o conhecimento específico e segue um mecanismo de organização especializado, com padrões como o uso da linguagem, os pontos a serem utilizados e a estruturação dos índices.

Assim, essa prática que existe desde as origens da biblioteca e com grande importância no século XX, pois foi capaz de tratar da explosão documental do momento, é caracterizada pelo processo de identificar, descrever ou caracterizar o conteúdo de um documento. Dessa maneira, as informações poderão ser retomadas através de buscas nos índices.

Dessa forma, o processo de indexação permite que os dados não se percam. É através dele que conseguimos ter acesso aos dados, por maior e mais variados que sejam, e posteriormente, poder acessá-los através de buscas. Ao indexar, destacamos um documento e realizamos uma análise prévia para inseri-lo em uma categoria. Quando trabalhamos com uma quantidade muito grande de dados, como a pesquisa do projeto Visagem, podemos encontrar a questão de como visualizá-los e, posteriormente, analisá-los. Os dados podem acabar se perdendo neles mesmos, e se “bagunçando” quando estão em grande quantidade. Para solucionar essa questão, a categorização das imagens se faz essencial.

Além do benefício de organização dos dados, a indexação e a posterior categorização, possibilitam que eles sejam analisados como um todo ou apenas como parte de um grupo de determinada categoria. Em nossa pesquisa, ao analisar um conjunto de dados como um todo, é possível analisar questões de cores, brilho e saturação, podendo ter uma visão geral de como as imagens interagem e quais os padrões que elas possuem. Já analisando-as em grupos menores, os dados se tornam mais claros e mais fáceis de interpretar, assim as análises podem ser mais profundas e específicas em torno de um determinado objeto que está presente na imagem.

Um importante item para a categorização de dados gerados online é a *hashtag*. É usada em sites como o *Facebook*, *Tumblr*, *Pinterest*, *Tumblr*, *Twitter* e o aplicativo *Instagram*. As *hashtags* são palavras-chaves adicionadas ao símbolo # (cerquilha). Elas se transformam em hiperlinks dentro da rede e são indexáveis pelo mecanismo de busca. Ao clicar em uma *hashtag*, é possível encontrar, ordenados, conteúdos que foram publicados com a mesma *hashtag*. Sendo assim, ela exerce a função de indexar aquilo que está anexado à ela, sejam textos ou imagens. Ela dá uma representação para aquilo



que está indexando e é criada pelo usuário que compartilha o conteúdo ao qual ela pertence.

A *hashtag*, enquanto processo de indexação, facilita a organização, pois agrupa os conteúdos nos quais ela foi inserida. Como vimos, isso é muito importante para facilitar a busca de temas que estão sendo discutidos nos sites de redes sociais. No Twitter, é possível ver quais as *hashtags* estão sendo mais usadas no momento, através dos *Trending Topics*⁶.

Há também a existência das *tags* afetivas. São *hahstags* que tem como ideia expressar um sentimento. Os usuários adicionam essas *hashtags* às suas mensagens com a ideia de mostrar seu estado emocional. A essência do uso desse tipo de *tag* não é possibilitar a organização para a posterior busca, apenas mostrar o que se sente. Alguns exemplos são *hashtags* como: #fome, #sono, #prontofalei, #desabafo, #morri, entre outras que tem como principal objetivo mostrar o seu posicionamento afetivo/sentimental.

O sinal #, nestes casos, não é um convite para a busca de outras mensagens com a mesma *tag*. Tampouco é um compromisso com a coletividade, à medida que o interagente não visa agrupar sua mensagem com outras sobre o mesmo tema, facilitando a busca temática. O objetivo nestes casos é expor publicamente um dado afetivo que acrescenta informações ao que acaba de ser escrito, ou que ressignifica o retweet da mensagem de um terceiro (PRIMO apud PEREIRA e CRUZ, p.7, 2010).

Podemos perceber então que as *hashtags* exercem três papéis importantes: elas indexam o conteúdo; criam comunidades por agruparem assuntos que possuem a mesma *hashtag*, possibilitando que os usuários interajam com aqueles que estão falando sobre o mesmo assunto, além de expressar emoções, com as *tags* afetivas.

Assim, quando trabalhamos com uma quantidade muito grande de dados, podemos encontrar a questão de como visualizá-los e, posteriormente, analisá-los, ao explorar essas funções das *hashtags*. Estas, torna-se, portanto, elementos chaves para as pesquisas desse porte e contribuem significativamente tanto para os processos de coleta (que envolve igualmente um processo de arquivamento para documentação e memória posterior), quanto para os processos de visualização, que podem tomar como base as palavras-chaves em destaque.

⁶Uma lista em tempo real das *hashtags* mais usadas. O recurso de *Trending Topics* usa por padrão a abrangência total, mas também é possível filtrar por países



FOLKSONOMIA: A INDEXAÇÃO DA MULTIDÃO

Com o avanço da internet e o aumento do número de usuários, o processo de armazenamento de informações mudou. Muitas pessoas passaram a compartilhar conteúdos utilizando plataformas da internet como local de armazenamento para esses conteúdos e documentos. Assim surge a Folksonomia, como uma vertente da indexação (junção das palavras taxonomia e *folk* - povo), termo criado por Thomas Vander Wal em 2004. Trata-se, portanto, de um mecanismo de representação, organização e recuperação de informações que não é feito por especialistas ou que segue uma norma preestabelecida e documentada. Ao contrário, é o próprio usuário que compartilha o conteúdo é quem vai organizá-lo, podendo adicionar *tags*/palavras-chaves que acreditam ser a mais propícias para representar aquilo que está disponibilizando na internet. Ademais, não há regras de linguagem ou de estruturação na Folksonomia, ou seja, o indivíduo possui liberdade para representar, organizar e recuperar a informação da forma que desejar e achar mais conveniente.

Apesar das diferenças nos processos que compõem a Folksonomia, ela cumpre os objetivos de um processo de indexação na representação de um documento, agora na visão do usuário. A questão que se coloca se refere à participação do usuário. Nos processos de indexação convencional, o usuário não é um participante ativo, mas o pressuposto da garantia de usuário faz com que ele não seja esquecido. (RODRIGUES e MOREIRA, p.89, 2012)

Esse sistema de categorização é adotada por muitos sites de compartilhamento, como blogs, *Delicious*⁷ e *Flickr*⁸. Além desses sites, a Biblioteca do Congresso dos EUA também utilizou a Folksonomia para a classificação do seu acervo. Através de uma parceria com o site *Flickr*, usuários puderam adicionar *tags* para as imagens do acervo americano. Assim, a Folksonomia permite essa ideia de colaboração e coletividade por trás da categorização. É, portanto, um olhar mais plural sobre um determinado conteúdo.

⁷O *Delicious* oferece um serviço on-line, que permite adicionar e pesquisar bookmarks sobre qualquer assunto. Mais do que um mecanismo de buscas para encontrar o que se quer na web, é uma ferramenta para arquivar e catalogar os sites preferidos para acessá-los de qualquer lugar.

⁸O *Flickr* é um site da web de hospedagem e partilha de imagens fotográficas. Permite interatividade entre os usuários.



Outro exemplo, é uma pesquisa realizada para entender como a Folksonomia estabelece suas relações, a qual analisou o processo de etiquetagem de imagens do *Flickr*. O principal objetivo da pesquisa era observar as estratégias que os usuários utilizavam para aderir uma *tag*/etiqueta à imagem. Para isso, foram analisados nove usuários, duas imagens de cada usuário e duas etiquetas de cada imagem.

Por ser complexa a análise de imagens, as *tags* que foram adicionadas às imagens foram divididas em três grupos: pré-iconográficas; iconográficas e iconológicas. O primeiro, foi atribuída às etiquetas que davam uma descrição genérica dos objetos e ações que estavam representados na imagem; o segundo são as etiquetas que nomeiam e identificam a imagem e os elementos identificados estão presentes na imagem; e o terceiro grupo está constituído das *tags* que estabelecem um assunto, identificam um evento, nomeiam ou identificam elementos da imagem sem que eles estejam representados na imagem. São *tags* que estabelecem relações criadas pelos próprios usuários.

Através dessa pesquisa, foi possível perceber as estratégias que basearam os usuários no uso de suas *tags*, como esse usuário se relaciona com a imagem analisada e a interpretação que ele fez para, escolher a etiqueta para categorizar a imagem. Somando-se a isso, foi percebido que cada usuário possui um padrão para categorizar as imagens.

Porém, ao mesmo tempo que a liberdade e a multiplicidade de visões podem ser vistas como pontos positivos da Folksonomia, elas também trazem dificuldades. Uma grande quantidade de *tags*, que muitas vezes podem ser utilizadas para a mesma informação⁹, pode dificultar a busca por determinado conteúdo. Da mesma forma, as *tags* podem não representar a mesma coisa para os usuários que categorizam e para aqueles que realizam as buscas. Portanto, informações e conteúdos podem não ser encontrados devido ao termo que foi usado para a sua categorização. Além desses fatores, existem as inconstâncias no vocabulário, como uso de gírias, abreviações, palavras grafadas com erros ortográficos, polissemia¹⁰ e sinonímia¹¹ que dificultam o acesso aos conteúdos.

⁹Tagspace : uma mesma tag é usada para representar dados diferentes

¹⁰ Uma determinada palavra ou expressão adquirir um novo sentido além de seu sentido original.

¹¹ São palavras distintas que possuem o mesmo significado



PROCESSO DE CATEGORIZAÇÃO DO PROJETO VISAGEM

O Laboratório de Estudos de Imagem e Cibercultura (Labic), criado em 2007, tem como objetivo desenvolver pesquisas e atividades de extensão sobre os impactos da cultura digital nos processos e nas práticas de comunicação contemporânea. Trabalhando com um grande volume de dados, o laboratório possui dois focos principais de pesquisa: a análise semântica dos textos e a análise das imagens extraídas dos sites de redes sociais.

Para realizar a pesquisa de um termo, para a posterior extração, nos apoiamos na Folksonomia já realizada através das *hashtags*. Sem as *hashtags* não conseguiríamos achar o grande volume de textos e imagens dos sites de redes sociais e tampouco acessar as comunidades criadas e os temas já indexados pelos indivíduos.. A pesquisa se baseia com a extração dos *tweets* através do software *yourTwrapperKeeper*. Através dele, conseguimos extrair as redes de *tweets* que pertencem a uma determinada *hashtag*. Depois de extraídos, é criado um arquivo que contém todos os *tweets* publicados no período de extração, com todas as informações disponíveis sobre eles, como a data de publicações e o número de republicações, informações que podem ser essenciais para outras análises.

Para análise, posterior à extração dos *tweets*, nos apoiamos sobre o arquivo que contém todas as informações sobre eles e usamos um *script* em java para separar os *tweets* que contém links. Depois dessa separação, o *script* realiza o *download* das imagens contidas nos links. Com as imagens extraídas, realizamos o processo de visualização através do software *ImageJ*, que permite visualizar e analisar grandes números de imagens e perceber como elas se correlacionam.

Como primeira parte do processo de análise e interpretação, categorizamos as imagens de um *dataset*¹². Essa categorização se faz essencial como maneira de organizar e recuperar as informações das imagens, através do termo da categoria. Além de podermos analisar todas as imagens presentes em uma *hashtag*, como um todo, também podemos analisar as imagens de uma determinada categoria, podendo traçar semelhanças, disparidades e os padrões das imagens categorizadas.

Inicialmente, a categorização era realizada manualmente através de uma tabela do *Excel*, onde eram inseridos o número da imagem e a sua respectiva categoria. Um processo lento e complicado, por se tratar de um grande volume de imagens. Após

¹² Conjunto de dados dos tweets, como imagens, tempo, geolocalização, retweets, indivíduo que compartilhou.



estudos e aperfeiçoamentos, conseguimos iniciar a categorização através de um software, *MatLab*.

O *Matlab* (abreviatura de MATrix LABoratory - Laboratório de Matrizes) é um software de simulação matemática que realiza operações matriciais, constrói gráficos em duas ou três dimensões, auxilia no processamento de sinais, além de manipular outras funções especializadas. Ele trabalha com uma linguagem de programação de alto nível, em um ambiente interativo, para o desenvolvimento de algoritmos, análise e visualização de dados e computação numérica. Próprio para as áreas técnicas e científicas, o software tem funções de tratamento numérico de alto desempenho, capazes de resolver problemas computacionais técnicos de forma mais eficiente do que as tradicionais linguagens de programação. Além do ambiente interativo, outra facilidade do *Matlab* é a possibilidade de execução de arquivos texto contendo uma sequência de instruções definidas pelo usuário. Esses arquivos texto, que têm extensão '.m', podem ser criados e editados dentro ou fora do seu ambiente.

As categorias foram inseridas no *Matlab*, cada uma com uma letra como referente. Assim, ao abrir as imagens no software, para categorizá-las, era preciso apertar a letra para aquela categoria. Ao clicar “enter”, a categorização da imagem era efetuada e uma tabela do *Excel* criada, com o número da imagem e a sua respectiva categoria. Para a análise das 8 mil imagens extraídas da *hashtag* “#passelivre” e “#passe_livre”, foram usadas as seguintes categorias: “multidão”, “pessoas”, “cartazes”, “retrato”, “polícia”, “bandeira”, “confronto”, “reação/resistência”, “humor”, “reprodução de jornais, revistas, sites e televisão”, “trânsito/caos”, “publicidade”, “outros”, “montagem” e “informativo/convocatória”. Essas imagens foram categorizadas a partir do objeto que elas representavam.

Com essa sistematização de categorias, o trabalho ficou muito mais rápido e simples. Afinal, as imagens já possuíam seu nome, e era preciso apenas clicar na letra referente à categoria escolhida para a imagem e apertar enter. Não precisava mais abrir cada imagem manualmente para depois abrir o arquivo do *Excel* e inserir o texto na tabela. Assim, o tempo gasto para a categorização de um grande volume de imagens se tornou muito menor em comparação à primeira categorização.

Com a separação das imagens em categorias, podemos ter uma outra visão do *dataset*. Podemos ver quais são as imagens que mais estão presentes, qual a categoria que é realçada. Assim, conseguimos interpretar melhor qual é o tipo de imagem que é



mais compartilhada. Além disso, se quisermos, podemos criar visualizações com as imagens de apenas uma categoria e ver quais são os seus padrões.

Mesmo com esse avanço que o software *MatLab* nos possibilitou, ainda precisamos de dispositivos que nos permita realizar a categorização totalmente automatizada, para assim, conseguirmos realizá-la de maneira mais ágil e eficiente, para posteriormente realizarmos uma análise mais densa das imagens que já foram categorizadas. Pensa-se futuramente em explorar, dessa forma, técnicas baseadas em *redes neurais*, sistemas computacionais que nos permitiriam identificar elementos através do reconhecimento de sua estrutura visual.

CONSIDERAÇÕES FINAIS

Através deste artigo, podemos perceber que o fluxo e o volume de informações e conteúdos se transformaram devido aos aparatos tecnológicos, tanto para a produção (computadores, celulares e máquinas fotográficas), quanto para o compartilhamento (internet e sites de redes sociais). O volume de dados, entre textos e imagens, se tornou muito maior, e por isso, se fez necessário métodos para analisá-los, como a indexação e a categorização.

A indexação e a categorização facilitam o processo de análise de grande volumes de dados, pois os organiza e permite que suas informações sejam retomadas posteriormente através de buscas. Uma dessas formas de indexação é a Folksonomia. Realizada por pessoas que não possuem um estudo específico para a indexação, ela se faz presente nos sites de redes sociais. Os mesmos indivíduos que compartilham o conteúdo, podem também atribuí-los palavras-chaves que acreditam ser as mais propícias para categorizá-lo. Uma dessas formas de folksonomia são as *hashtags*.

As *hashtags* além de criarem uma ideia de comunidades, englobando informações que possuem as mesmas palavras-chaves, elas também possibilitam que através de pesquisas, sejam possíveis encontrar esses conteúdos. E é através dessa função da *hashtag* que conseguimos extrair os dados, através do software *yourTwapperKeeper*, para a pesquisa do projeto Visagem.



REFERÊNCIAS

CALMON, Priscila. Narrativas controversas: As tramas emergentes da ciberguerra do #Wikileaks, Vitória, 2013

MCCANDLESS, David. Palestra Ted Talks. Disponível em <<http://comunicacaochapabranca.com.br/?p=14857>> Acesso em: 2 de abril de 2014

RODRIGUES, André Augusto; MOREIRA, Manoel Palhares. Folksonomia: análise da etiquetagem de imagens no Flickr, Recife, 2012

PEREIRA, Débora; CRUZ, Ruleandson. Folksonomia e tags afetivas: comunicação e comportamento informacional no Twitter, Minas Gerais, 2010