Classificação do conteúdo dos jornais online, segundo a lógica da Web Semântica

Leonardo Rocha
ESDI-UERJ
R. Evaristo da Veiga 95, Lapa,
Rio de Janeiro, RJ
leoburla@estacazero.com

Washington Lessa
ESDI-UERJ
R. Evaristo da Veiga 95, Lapa,
Rio de Janeiro, RJ
wdlessa@esdi.uerj.br

André Monat
ESDI-UERJ
R. Evaristo da Veiga 95, Lapa,
Rio de Janeiro, RJ
andresmonat@yahoo.com.br

RESUMO
Este artigo tem por objetivo mostrar o panorama atual dos jornais *online* brasileiros, no que diz respeito à classificação do seu *conteúdo*, identificando problemas acarretados pela utilização limitada dos recursos próprios da Web.

A partir de uma indicação do cenário mundial, serão levantadas algumas das iniciativas mais representativas e bem recebidas no âmbito internacional, que visam a potencialização do jornalismo digital – seja através da aplicação sistemática de conceitos inovadores de classificação do *conteúdo digital* ou de novas técnicas de rotulação/anotação da informação – ampliando as dimensões da estrutura narrativa e, por conseguinte, propiciando ao leitor uma experiência de imersão mais rica e profunda durante o consumo do *conteúdo*.

Concluído o levantamento, será proposta uma solução adequada à realidade dos jornais *online* brasileiros, capaz de ser implantada dentro de seus produtos, trazendo os subsídios necessários para que os resultados sejam compatíveis com aqueles alcançados pelas exemplares iniciativas externas, tanto do ponto de vista funcional quanto do tecnológico.

Author Keywords
Jornalismo online, classificação, conteúdo digital, experiência de uso, tecnologia, Web Semântica.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.
Copyright 2009 ACM 978-1-60558-246-7/09/04...$5.00.

INTRODUÇÃO

Apesar da Internet já existir há aproximadamente 16 anos no país, pouco se fez em relação à classificação do *conteúdo* dos jornais *online* brasileiros. No início, muitos veículos optaram pela simples transposição de conteúdos impressos para o *online*, sem tratamento editorial e mercadológico; os jornais impressos, caracterizados pela composição em capa e cadernos e pela limitação de espaço (do papel), tinham suas edições transferidas para a versão *online* sem qualquer alteração do *conteúdo*. Hoje, embora essa transposição não mais ocorra, os jornais *online* brasileiros ainda se baseiam nos jornais impressos; suas páginas iniciais assemelham-se às capas de jornal em papel e as matérias são associadas a *editorias* que, analogamente às versões impressas, funcionam como grandes diretórios, tais quais os cadernos de seus predecessores, sendo um reflexo da divisão/especialização do trabalho dentro dos jornais.

Vale observar que o conceito de *editoria*, que por diversas vezes será abordado ao longo do artigo, não se refere aos grupos especializados de profissionais que, dentro de uma organização jornalística, cobre somente assuntos específicos de sua competência. O termo diz respeito tão somente ao conceito de *diretório* – normalmente representado sob a forma de seção – sendo este parte integrante do primeiro nível da estrutura de navegação dos *websites* de notícias.

No que concerne aos papéis existentes dentro da estrutura organizacional de um jornal *online*, serão utilizadas apenas as denominações *produtor de conteúdo digital*, ou simplesmente *produtor de conteúdo*, que equivale ao repórter que faz o levantamento e constrói o *conteúdo digital* em si, ao jornalista que redige a matéria ou que, de alguma forma, dá corpo ao fato noticioso, seja ele fotógrafo, ilustrador/infografista, vídeografista ou editor de áudio/vídeo; e *gestor de conteúdo digital*, ou tão somente *gestor de conteúdo*, cuja atribuição é classificar o *conteúdo digital*, ou garantir a sua correta classificação, mantendo-o

organizado da melhor forma possível, de modo a facilitar as operações de busca e recuperação de informações.

O que aqui se pretende discutir é a forma como o *conteúdo* resultante do jornalismo digital é até hoje, no Brasil, tratado, categorizado e inter-relacionado. No cenário mundial, já é possível encontrar algumas iniciativas que buscam o rompimento desse paradigma, adotando uma abordagem que explora o potencial da Web – de modo a aproveitar todos os recursos tecnológicos existentes e garantir uma série de benefícios, tanto para os *produtores* e *gestores de conteúdo digital* quanto para o seu público consumidor – distanciando-se consideravelmente da lógica dos jornais impressos.

É preciso adotar um novo modelo de solução, aproveitando-se de algumas das iniciativas internacionais – que já passaram por inúmeras adaptações e melhorias, fruto da longa experiência e da busca constante por inovações no meio digital – para, finalmente, romper com o modelo de classificação do *conteúdo* ainda utilizado pelos jornais *online* brasileiros, que, se é injusto apontá-lo como incorreto ou inapropriado, poder-se-ia dizer ultrapassado, tanto em relação à prática do mercado globalizado quanto das necessidades dos leitores e, até mesmo, dos *produtores* e *gestores de conteúdo digital*, cuja dinâmica de trabalho exige uma operação cada vez mais ágil, a partir de um volume de informações em extraordinária expansão.

Ao se estabelecer uma proposta compatível com as potencialidades que se configuram na Web, intensifica-se o fluxo de consumo do *conteúdo*, entre outras coisas, permitindo ao leitor uma experiência de imersão mais rica e profunda, sendo esta o resultado da produção de conhecimento propiciada pelas novas formas de classificação e relacionamento dos fatos noticiosos em meio digital.

Tais consequências podem levar ao aumento do tempo de permanência do leitor, à expansão da oferta de espaços publicitários e de patrocínio em áreas dedicadas a nichos de consumo qualificados e, por conseguinte, ao crescimento da receita dos jornais *online* brasileiros, isto sem contar com a percepção de valor que poderá ser agregada às suas marcas pelo próprio público, dado o avanço destes veículos como fontes de informação.

CENÁRIO ATUAL

Para diagnóstico adequado dos jornais *online* no Brasil, é preciso, antes de mais nada, considerar a trajetória destes veículos até o momento presente, bem como os elementos fundamentais que os caracterizam, de modo a estabelecer uma generalização que sirva de ponto de partida à solução que será proposta mais à frente.

Sobre o referido assunto, há os trabalhos de Bardoel & Deuze (2001) [1] e Palacios (2002) [10], que estabelecem como características do jornalismo desenvolvido para a Web a *multimidialidade*, a *interatividade*, a *hipertextualidade*, a *personalização*, a *memória* e a *atualização contínua*, conforme as seguintes definições:

Multimidialidade:

Para Palacios (2003) [12], o conceito diz respeito à:

> [...] convergência dos formatos das mídias tradicionais (imagem, texto e som) na narração do fato jornalístico. A convergência torna-se possível em função do processo de digitalização da informação e sua posterior circulação e/ou disponibilização em múltiplas plataformas e suportes, numa situação de agregação e complementaridade (Palacios, 2003, p:3) [12].

Interatividade:

Entre as demais características, Bardoel & Deuze (2001) [1] consideram que a notícia *online* possui:

> [...] a capacidade de fazer com que o leitor/usuário sinta-se mais diretamente parte do processo jornalístico. Isto pode acontecer de diversas maneiras: pela troca de e-mails entre leitores e jornalistas, através da disponibilização da opinião dos leitores, como é feito em sites que abrigam fóruns de discussões, através de chats com jornalistas, etc (Bardoel & Deuze, 2001, p:5) [1].

Machado (1997) [6], por sua vez, ressalta que:

> A interatividade ocorre também no âmbito da própria notícia, ou seja, a navegação pelo hipertexto também pode ser classificada como uma situação interativa. Adota-se o termo multi-interativo para designar o conjunto de processos que envolvem a situação do leitor de um jornal na Web (Machado, 1997, p:146) [6].

Dessa forma, é possível identificar dois elementos distintos presentes no conceito de *interatividade* que são (a) a navegação *per se* através do hipertexto, o que já equivale a uma situação interativa e (b) as atividades de mídia social, que correspondem aos processos de comunicação entre usuários e seu engajamento no enriquecimento da narrativa jornalística, estejam estes na condição de leitores ou jornalistas, promovidos por ferramentas e aplicativos específicos para este fim.

Faz-se necessário observar que o termo “mídia social”, ou *social media*, é empregado para designar o uso do meio eletrônico para interação social, permitindo a criação, o compartilhamento e o intercâmbio do *conteúdo* gerado pelo usuário. Cada vez mais apoiada por tecnologias ubíquas e escaláveis, a mídia social traz mudanças consideráveis à forma como as organizações, comunidades e indivíduos se comunicam. Como exemplos de produtos de mídia social, temos Facebook, Twitter e Flickr.

Também é válido esclarecer que por narrativa jornalística, ou narrativa do fato jornalístico, entende-se não apenas a reportagem, que é o gênero privilegiado da narração

jornalística, mas também a notícia pura e simples de um fato qualquer, visto que a notícia não é um acontecimento em si, mas sim a sua narração. Desse modo, as diferenças entre notícia e reportagem se limitam às formas de condução da narração. Como ingredientes imprescindíveis à narrativa há personagens, ação dramática e descrições de ambiente.

Não há como falar em narrativa jornalística sem considerar a estrutura organizacional das informações. Sobre isso, Fidalgo (2004) [4] explica:

> Ao princípio, a informação é dada em traços gerais, consistindo preferencialmente na resposta sumária às perguntas que tradicionalmente enformam um *lead*: quem, o quê, quando, onde, porquê. Depois vêm as notícias subsequentes completando e pormenorizando a informação. De certo modo, poder-se-ia entender essa sucessão de notícias como uma extensão da noção de pirâmide invertida, usada na feitura de uma notícia, a um conjunto de notícias sobre o mesmo evento (Fidalgo, 2004, p:2) [4].

Assim, ao narrar um fato jornalístico, o *produtor de conteúdo* deve, em princípio, tentar responder às questões essenciais próprias do *lead* – ou seja: o que, quem, quando, onde, como e porque – ainda que a urgência do imediatismo, este próprio do meio *online*, se sobreponha às exigências da objetividade e da verificabilidade.

Na Internet, a narrativa jornalística conta com diversos formatos de mídia. Texto, imagem, áudio e vídeo em formato digital são utilizados de forma integrada e complementar, de modo a enriquecer e pormenorizar a informação, gerando o que aqui já foi definido como *multimidialidade*.

Hipertextualidade:

O hipertexto está relacionado com os conceitos e as propriedades da hipermídia, que, para Machado (1997) [6], é:

> [...] uma forma combinatória, permutacional e interativa de multimídia, em que textos, sons e imagens (estáticas e em movimento) estão ligados entre si por elos probabilísticos e móveis, que podem ser configurados pelos receptores de diferentes maneiras (Machado, 1997, p:146) [6].

A partir do exposto por Machado (1997) [6], fica clara a relação entre os conceitos de *interatividade* e *hipertextualidade*, uma vez que a situação interativa abrange também a navegação propiciada pela difusão de *links*, cujo objetivo é a interconexão de diferentes formatos de mídia.

Personalização:

Conceito que está diretamente relacionado à *interatividade*, sendo fruto da relação do leitor com o produto noticioso, suas escolhas e sua linearidade dentro de uma narrativa particular. Para Zuin & Correia (2008) [16], a personalização está:

> [...] próxima da dissertação, uma vez que o usuário tem condições de fazer e/ou modificar o processo, ou ainda construir de acordo com sua mediação (Zuin & Correia, 2008, p:13) [16].

Memória:

O armazenamento do *conteúdo* em banco de dados permite que o usuário acesse qualquer *conteúdo* e estabeleça seus próprios fluxos de leitura, criando assim uma relação particular com a questão da temporalidade. Sobre isto, Machado (2004) [7] diz o seguinte:

> O uso das bases de dados permite a ininterrupção do fluxo informacional, o que torna possível e confiável o acesso a informações. [...] o texto webjornalístico, incrementado pelo banco de dados, pode viabilizar vários caminhos a determinados conteúdos e o usuário pode acessá-los a qualquer momento (Machado, 2004, p:5) [7].

Atualização contínua:

O jornalismo na Web permite atualizações de *conteúdo* em *tempo real*. Além disso, o seu *conteúdo* pode ser acessado a qualquer instante e em qualquer lugar via Internet. Sobre esse assunto, Palacios (2003) [12] esclarece que:

> A rapidez do acesso, combinada com a facilidade de produção e de disponibilização, propiciadas pela digitalização da informação e pelas tecnologias telemáticas, permitem uma extrema agilidade de atualização do material nos jornais da Web. Isso possibilita o acompanhamento contínuo em torno do desenvolvimento dos assuntos jornalísticos de maior interesse (Palacios, 2003, p:4-5) [12].

Uma vez levantados os elementos que caracterizam o jornalismo *online*, faz-se necessária uma investigação sobre o seu processo evolutivo e as mudanças que dizem respeito à narrativa jornalística. Nesse contexto, temos os trabalhos de Pavlik (2001) [13], Mielniczuk (2003) [9] e Schwingel (2005) [15].

Pavlik (2001) [13] aponta três estágios de evolução para o jornalismo digital: o primeiro, caracterizado pela transposição do conteúdo impresso para a Internet; o segundo, pela agregação de recursos e criação de *conteúdos* originais; e o terceiro – que, segundo ele, estaria começando a emergir, na época da elaboração do seu trabalho – marcado por um produto totalmente exclusivo para a Internet.

Já Mielniczuk (2003) [9], vai um pouco mais adiante e propõe que a trajetória dos produtos jornalísticos desenvolvidos para a Web seja dividida em três momentos: produtos de primeira geração ou fase de transposição,

produtos de segunda geração ou fase de metáfora e produtos de terceira geração ou fase do jornalismo digital. Na fase transpositiva, a disponibilização de informações jornalísticas na Web fica restrita à possibilidade de ocupar um espaço, sem explorá-lo como um meio que apresenta características específicas. Na fase metafórica, o jornal impresso funciona como uma referência para a elaboração das interfaces dos produtos e começam a ocorrer experiências na tentativa de explorar as características oferecidas pela rede. O jornalismo digital de terceira geração é descrito pela autora como "sites jornalísticos que extrapolam a ideia de uma versão para Web de um jornal impresso já existente".

Schwingel (2005) [15], por sua vez, complementa o trabalho de Mielniczuk (2003) [9] apresentando indícios de uma nova etapa do jornalismo digital – ou seja, a quarta geração de produtos jornalísticos desenvolvidos para a Web – na qual ocorre a consolidação da utilização de bancos de dados complexos através da utilização de ferramentas automatizadas e diferenciadas (sistemas de apuração, edição e veiculação de informações) na criação e manutenção de produtos jornalísticos. Tais ferramentas estabeleceriam vínculos entre diferentes plataformas (Web, e-mail etc) e ambientes distintos (*chats*, fóruns etc), utilizando-se de tecnologias também diferenciadas (PHP, Pearl, XML etc), de acordo com seus interesses e conveniências.

É necessário, contudo, esclarecer que a noção de banco de dados aplicada ao jornalismo digital vai além dos conceitos de depósito integrado de dados, de coleção de documentos ou de repositório para busca e recuperação de informações. Em conformidade com o princípio da transcodificação, desenvolvido por Manovich (2001) [8] – segundo o qual todos os objetos digitais podem ser traduzidos para outros formatos – Machado (2004) [7] demonstra como os bancos de dados são uma forma cultural com estatuto próprio no jornalismo digital. Para o autor, o banco de dados desempenha as funções simultâneas e complementares de (a) formato para a estruturação da informação, (b) suporte para modelos de narrativa multimídia e (c) memória dos *conteúdos* publicados, o que resulta num formato para esta modalidade.

Terminologias à parte, Pavlik (2001) [13], Mielniczuk (2003) [9] e Schwingel (2005) [15] reconhecem a existência de um estágio no qual, em última análise, o jornalismo digital deixa de ser uma versão do modelo impresso, e passa a ser uma nova e distinta forma de se apresentar informações, momento este de extrema relevância para o trabalho. Assim sendo, como referência a essa fase na evolução do jornalismo *online*, será adotado o termo "jornalismo digital de quarta geração", assim definido por Schwingel (2005) [15], ou simplesmente "jornalismo digital".

Especificamente sobre o jornalismo digital no Brasil, Palacios (2002) [11], ao realizar um estudo que avalia as características anteriormente apontadas – que são *multimidialidade*, *interatividade*, *hipertextualidade*, *personalização*, *memória* e *atualização contínua* – conclui que poucos são os recursos oferecidos pela Web que são utilizados pelos jornais *online* brasileiros, apresentando fortes características das publicações em papel, não explorando de forma satisfatória as possibilidades oferecidas pelo ambiente digital para o desenvolvimento de produtos jornalísticos. Entre outras coisas, percebe-se que a *multimidialidade* não aparece como uma característica explorada e que a *hipertextualidade* ainda é utilizada apenas para organizar a publicação, não sendo explorada na própria narrativa do fato jornalístico.

Apesar da análise – que pode ser considerada ultrapassada, já que foi realizada há nove anos – é possível afirmar que, no Brasil, estamos nos aproximando cada vez mais do jornalismo digital de quarta geração. Podemos observar, por exemplo, que algumas publicações organizam a narrativa em blocos de textos com *links* que levam a outros *conteúdos* ou extensões desta narrativa. Os *links* disponibilizam também versões e complementos em imagem, áudio, vídeo e outros formatos de informação, que vêm a constituir a narrativa jornalística. Em um primeiro momento, esses recursos são apenas informações a mais. Num passo seguinte, os recursos multimídia tornam-se parte da narrativa. Nesse estágio, a informação a mais aparece também como uma matéria sobre o mesmo assunto que já havia sido publicada pela mesma empresa. São explorados níveis de interatividade como o uso jornalístico do chat, enquetes e, principalmente, possibilidades do usuário comentar notícias e ler comentários de outros usuários. Finalmente, observa-se a valorização do arquivo, apesar dele já existir anteriormente. A divisão em *editorias* ainda é visível, porém algumas transformações já são praticadas, principalmente pelos portais de serviços, que fazem uso das *breaking news* – que podem ser definidas como reportagens ou boletins especiais, que cobrem acontecimentos que acabaram de ocorrer ou que estão em andamento, cuja veiculação interrompe (ou recebe mais destaque que) as notícias do momento ou aquelas que estão previstas na programação normal, por decisão editorial – sendo esta a situação que melhor retrata o jornalismo *online* brasileiro na atualidade. Nesse ponto, é válida uma análise mais apurada e aprofundada do assunto para compreensão do cenário e posterior identificação dos problemas.

Hoje, as *editorias* utilizadas pelas versões *online* dos jornais impressos, a exemplo de O Globo, ainda lançam mão da solução adotada por seus irmãos mais velhos, e mesmo os jornais puramente *online*, a exemplo do G1 (vide Figura 1), fazem referência a este modelo. Desse modo, evidenciamos *conteúdos* de toda sorte sendo classificados num espectro limitado de assuntos como: *Blogs e Colunas*, *Brasil*, *Carros*, *Ciência e Saúde*, *Concursos e Emprego*, *Economia*, *Especiais*, *Esporte*, *Loteria*, *Mundo*, *Natureza*, *Planeta Bizarro*, *Política*, *Pop & Arte*, *Tecnologia e Games*, *Vestibular e Educação*, sendo estas as *editorias* que

compõem o G1. Alternativamente, para atender a demanda do momento, algumas *categorias* – muitas vezes de caráter temporário – são disponibilizadas conjuntamente às *editorias*, porém com exposição diferenciada, nos jornais. Eis alguns exemplos: *Caso Isabella*, *Tsunami no Pacífico*, *Tragédia em Realengo*. Ou seja, dependendo da emergência ou da procura, um assunto até então inexistente pode receber destaque nos jornais *online*, sendo incorporado à sua estrutura de navegação, o que é característico das *breaking news*.

para apuração de irregularidades na emissão de passagens aéreas para senadores – nota-se que a mesma faz parte da *editoria Política*. No entanto, não há qualquer referência a assuntos que talvez sejam mais pertinentes ao leitor como, por exemplo, *Escândalo*, *Senado Federal* ou *Congresso Nacional*. São relacionadas apenas duas matérias para aprofundamento da leitura, enquanto é sabido, após realização de busca, que existem ao menos oito matérias sobre o tema. Além disso, no corpo da matéria há uma imagem de referência integrada ao texto e somente um *link*,

Figura 1. *Screenshot* da página inicial do jornal *online* G1, com o menu de *editorias* aberto.

A questão das *editorias* é um velho dilema que se estende até hoje, não apenas por causa do modelo já estabelecido pelos jornais tradicionais, mas pelo simples fato da maioria dos editores-chefe, e demais responsáveis, dos jornais *online* vir da área impressa, o que contribui à perpetuação de uma estrutura de *conteúdo* inadequada e incapaz de tirar todo proveito da dinâmica da Internet.

O problema das *editorias* se reflete tanto na estrutura de navegação quanto na organização dos *conteúdos* e suas inter-relações dentro do jornal, a saber: no G1, ao visualizar uma matéria cujo título é *Senado abre sindicância para apurar possível nova 'farra de passagens'* (vide Figura 2) – que, resumidamente, relata uma sindicância interna no Senado, instalada a pedido do senador Heráclito Fortes, que aponta para uma das matérias já relacionadas. Com base nesse exemplo, é possível identificar as seguintes restrições técnicas e operacionais:

- Existe uma gama extremamente restrita de *editorias* ao qual o *conteúdo* pode se relacionar, fazendo com que o mesmo seja associado a assuntos muito genéricos e, às vezes, estranhos à natureza do tema abordado;
- Por causa dessa restrição, não é possível estabelecer um relacionamento entre *conteúdos* – que tratam de um tema específico – de modo automatizado, tendo de fazê-lo manualmente;
- Da mesma forma, é inviável estabelecer novos eixos de navegação, automática ou semi-automaticamente, a partir de termos que são relevantes ao *conteúdo* ou que se

Figura 2. *Screenshot* da matéria *Senado abre sindicância para apurar possível nova 'farra de passagens'*, do jornal *online* G1.

repetem ao longo do texto. Exemplo: *Senado Federal*, *Mesa Diretora do Senado*, *Câmara dos Deputados* e *Heráclito Fortes*.

- A busca e recuperação da informação – seja ela no formato de texto, imagem, áudio ou vídeo – a integrar o *conteúdo* publicado é limitada e demorada, fazendo com que os esforços dispendidos na operação fiquem concentrados nos temas de maior importância. Consequentemente, a narrativa jornalística acaba por ficar prejudicada na maior parte dos *conteúdos* ofertados.
- Nenhuma das *editorias* existentes possui relação funcional, seja de hierarquia, dependência etc, entre si. Desse modo, não é possível fazer com que os *conteúdos* associados a uma *editoria* também sejam apresentados automaticamente em outra. Seria interessante, por exemplo, que a região (*editoria*) *DF (Distrito Federal)* fosse, na prática, um subitem da *editoria Brasil*; desse modo, todo *conteúdo* de *DF* seria exibido em *Brasil*.

A partir do exemplo dado e do entendimento que, teoricamente, as *editorias* nada mais são que *categorias* generalistas de classificação, é possível perceber que a adoção de uma estrutura que visa organizar seu *conteúdo* em *categorias* – que, além de poucas, são isoladas e não fazem alusão à particularidade do tema abordado – não se aproveita das potencialidades da Web e, sendo assim, deixa de aperfeiçoar a experiência de uso dos produtos jornalísticos desenvolvidos para este meio, tanto do lado do *produtor de conteúdo* quanto do leitor. Ao se verificar tais consequências, e levando em conta os conceitos envolvidos na classificação, algumas questões essenciais à superação dos problemas podem ser apontadas, fazendo-se antes, porém, as seguintes considerações:

- Por *conteúdo digital*, ou simplesmente *conteúdo*, entende-se toda e qualquer representação digital da informação, capaz de ser publicada na Web. Um *conteúdo* pode ser apresentado na forma de texto, imagem, vídeo etc;
- A *categoria* é um recurso de classificação que agrupa *conteúdos* similares – estes, por sua vez, abordam, retratam ou citam um tema em comum, seja ele qual for – recebendo o nome daquilo a que faz referência. Exemplos: *Escândalo*, *Política*, *Heráclito Fortes*, *iPhone*, *Senado Federal*, *Brasil* e *Carnaval 2011*;
- A *entidade* é uma forma particular de categorização que envolve a individualização de *categorias* do tipo *evento*, *lugar*, *organização*, *pessoa*, *produto* e congêneres. Exemplos: *Heráclito Fortes*, *iPhone*, *Senado Federal*, *Brasil* e *Carnaval 2011*. Nota-se, com base nos exemplos, que qualquer *entidade* pode ser tratada como *categoria*.

Uma vez esclarecidos os conceitos básicos de classificação, surgem questões que dizem respeito a:

1. Quantidade de *categorias*;
2. Especialização das *categorias*;
3. *Entidades* como *categorias*;
4. Associação entre *categorias*;
5. Associação entre *conteúdos* e *categorias*.

É interessante observar que a associação entre *conteúdos* e *categorias* está estreitamente relacionada ao processo de rotulação/anotação do *conteúdo*, e será explicado em breve.

Com base nos fatores expostos, uma orientação é claramente estabelecida, no que diz respeito tanto ao levantamento das iniciativas quanto à proposição de uma solução, conforme será possível verificar adiante.

NOVAS INICIATIVAS

No âmbito internacional, é possível destacar, dentre uma grande variedade de jornais *online* de alto padrão editorial, nomes como The New York Times, BBC News, MSNBC.COM e CNN (vide Figura 3), sendo estes os produtos jornalísticos que foram utilizados como referência analítica. Todos são reconhecidos pela busca incessante por inovações, por fazerem uso do estado da arte em tecnologia, além de investirem constantemente em design, conduzindo pesquisas de qualidade e usabilidade, entre outras coisas.

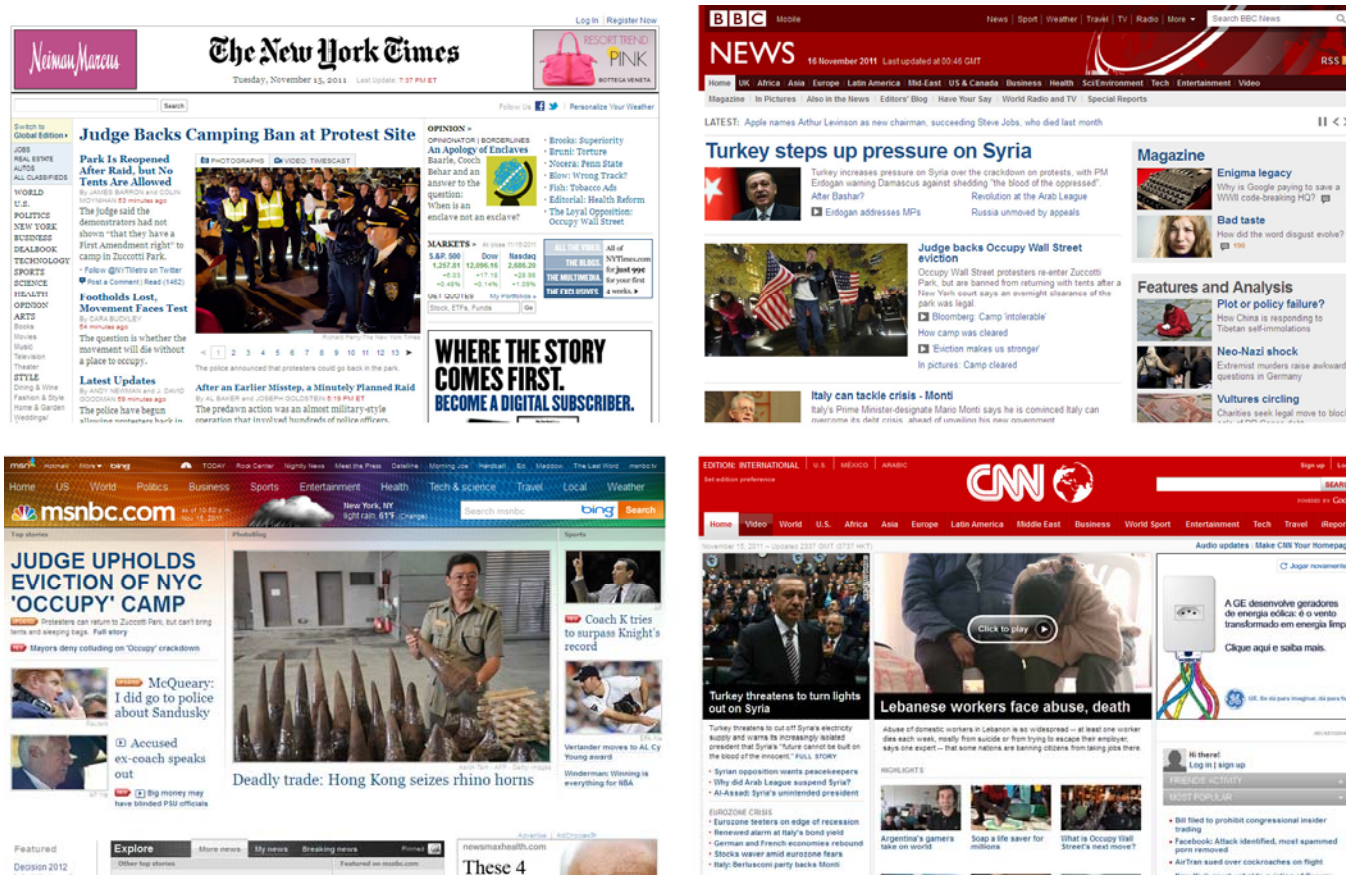

Figura 3. *Screenshots* da página inicial dos jornais *online* The New York Times, BBC News, MSNBC.COM e CNN.

Antes de nos aprofundarmos na questão da classificação do *conteúdo*, porém, é importante esclarecer que os jornais avaliados possuem um fluxo de trabalho que conta com o recurso de rotulação/anotação do *conteúdo* – que pode ser entendido como a atividade por trás da associação entre *conteúdos* e *categorias* – que é realizado de forma detalhada e meticulosa, demonstrando enorme preocupação com os metadados, que são a peça-chave na gestão das informações, responsáveis pelo modo como o *conteúdo* será indexado dentro do banco de dados. E mais: de modo geral, os metadados estão diretamente relacionados à narrativa jornalística, sendo muitas vezes o equivalente às respostas dadas para as clássicas perguntas que compõem o *lead* da notícia.

Segundo Colle (2002) [3], os metadados são dados sobre outros dados e se constituem em ferramentas que devem guiar os usuários tanto na consulta quanto na busca de novas informações sobre os mesmos e sobre o sistema em si.

Preocupada com o correto entendimento do que é metadado, Schwingel (2005) [15] esclarece:

> Para uma maior elucidação, podemos usar como exemplo hipotético a cobertura jornalística de algum escândalo financeiro ocorrido na década passada, no qual diversas pessoas tiveram seus nomes associados às fraudes constatadas. Caso alguma delas esteja vinculada atualmente a outro tipo de negócio que contenha irregularidades, a cobertura atual deveria contextualizar a questão com a trajetória de vida dos envolvidos (Schwingel, 2005, p:4) [15].

No que diz respeito ao banco de dados, e seus efeitos sobre a estruturação e a organização das informações, Fidalgo (2004) [4] destaca a sua flexibilidade para a forma de apresentação das notícias, assim como suas vantagens para garantir objetividade às mesmas. O autor introduz o conceito de *resolução semântica*, que é explicada da seguinte forma:

> Uma primeira notícia sobre um acontecimento, que à partida surge com um determinado sentido, pode ser complementada, alterada, corrigida, à medida que outras notícias sobre o mesmo acontecimento se lhe seguem. O que de início tinha contornos indefinidos, deixando múltiplas hipóteses em aberto, vai ganhando sucessivamente formas cada vez mais definidas. Um exemplo ajudará a perceber intuitivamente o que foi dito. Uma notícia de última hora pode ter a seguinte forma: "Houve uma forte explosão em Bagdad". É evidente que se trata de uma notícia de muito baixa resolução semântica. Não se sabe que tipo de explosão foi, se acidental ou provocada, se provocada por um carro armadilhado ou se por um ataque de aviação, se houve ou não houve vítimas, em que local da cidade, etc., etc. Porém, à medida que forem chegando notícias subsequentes a notícia do que ocorreu vai ganhando forma, ou seja, aumenta a sua resolução semântica (Fidalgo, 2004, p:3) [4].

Ademais, vale notar que Manovich (2001) [8] vê o banco de dados como a forma estruturadora dos produtos digitais na contemporaneidade, pois é um complexo de armazenagem de formas culturais que pode servir, inclusive, para criar novos gêneros e narrativas nas mais distintas áreas, enquanto que Gonçalves (2004) [5] o encara como um formato para o jornalismo digital.

Voltando às iniciativas internacionais, uma das soluções mais recentes adotadas pelo NYT, BBC News, MSNBC.COM e CNN é a ideia de *tópico*, que corresponde ao conceito de *categoria*, tal qual a definição dada no presente artigo. Desse modo, entende-se por *tópico* toda e qualquer *categoria* que, independente da forma como é representada em termos de desenho de interface – ou seja, da aparência visual de sua *webpage* – agrega *conteúdos* de todo tipo, relacionados ao tema a que se faz alusão. A diversidade de *tópicos* disponíveis é enorme, sendo permanentemente utilizado como um dos eixos da narrativa jornalística. Os *tópicos* são associados entre si, havendo a possibilidade de navegar por entre os *conteúdos*, a partir de uma cadeia de *tópicos* inter-relacionados.

Cada *tópico* tem definição e significado específicos – o que garante a relevância dos *conteúdos* a ele vinculados – sendo representado por uma *webpage* individualizada. Assim sendo, o *tópico* funciona como um importante nó de navegação e, portanto, de consumo dos *conteúdos* que ele tem por finalidade agregar.

As *editorias* continuam sendo usadas, mas desta vez com algumas mudanças. Na prática, elas correspondem a *tópicos* que se manifestam de modo diferenciado na interface do produto jornalístico, fazendo parte da estrutura de navegação básica do *website*. Nos demais aspectos, o seu comportamento é exatamente igual ao do *tópico*.

Além das *editorias*, que podem ser consideradas parte integrante dos *tópicos*, há outros tipos de *tópico* presentes na estrutura estabelecida. Observa-se, no entanto, abordagens distintas entre os jornais estudados. A solução mais interessante é a adotada pelo NYT, que divide os *tópicos* em *assunto*, *lugar*, *organização* e *pessoa* (vide Figura 4).

Figura 4. *Screenshot* da página principal de *tópicos*, do jornal *online* The New York Times.

Existe também o conceito de *cobertura especial*, que se pode chamar simplesmente de *história*. Trata-se de uma solução utilizada pelos jornais estudados, cujo desenho da interface se distingue dos demais e que, da mesma forma que o *tópico*, tem por objetivo reunir *conteúdos* diversos sobre um determinado tema, sendo este um acontecimento específico como, por exemplo, *Posse de Dilma*, *Réveillon 2011* e *Carnaval 2011*. Necessariamente, o tema abordado tem de ser um evento de caráter temporário, com início, meio e fim, ou seja: sua existência deve estar fortemente

vinculada a um período, ainda que indeterminado, de tempo.

Embora as *histórias* e as *breaking news* possuam algumas características em comum – como, por exemplo, o fato de retratarem um evento e receberem destaque nos jornais *online* – os seus conceitos são distintos: resumidamente, enquanto que o primeiro diz respeito a um acontecimento cuja veiculação está programada pelo jornal, o segundo, ao contrário, está associado à interrupção das notícias previstas na programação normal. Isso não significa, porém, que as *breaking news* não possam ser consideradas *histórias*, findo o período emergencial; de certo modo, elas podem ser encaradas como uma especialização do conceito de *história*.

Apesar da diferenciação conceitual entre *história* e *tópico*, nos jornais avaliados os dois tipos de *categoria* se comportam de modo semelhante, havendo associações entre si, inclusive. Na MSNBC.COM, por exemplo, há uma estreita relação entre *história* e *assunto* e vice-versa, explorando um pouco mais as dimensões da estrutura narrativa.

Considerando-se os cinco fatores levantados ao final da seção anterior, é possível constatar que todos os jornais estrangeiros estudados incorporam cada um deles, sem exceção, devido à lógica de funcionamento dos *tópicos*. O conceito de *história*, por sua vez, serve de apoio aos *tópicos* e aprofunda ainda mais a narrativa do fato jornalístico.

Em relação à tecnologia, é válido dizer que no NYT, BBC News e MSNBC.COM a plataforma que abarca tais inovações é baseada na Web Semântica, e será explicada mais adiante.

É importante notar que, no Brasil, embora o jornal *online* Estadão se utilize do termo "tópico", este nada mais é do que o resultado do uso de *tags* sem tratamento semântico, ou seja, da rotulação do *conteúdo* com termos desprovidos de significado, ao contrário do que acontece na Web Semântica – não sendo possível, consequentemente, estabelecer qualquer tipo de relação entre uma *tag* e outra; a relação é estabelecida tão somente entre *conteúdos* que são marcados com uma mesma *tag* – distanciando-se consideravelmente da solução que neste artigo se pretende propor.

SOLUÇÃO PROPOSTA

Considerando-se as iniciativas observadas, bem como os fatores que devem nortear as mudanças necessárias à suplantação dos problemas identificados anteriormente, finalmente uma solução pode ser proposta aos jornais *online* brasileiros.

A criação dos *tópicos*, que atualmente integram os jornais estrangeiros estudados, deve ser o ponto de partida dos produtos jornalísticos *online* brasileiros, de modo que os resultados obtidos por aqueles veículos também se façam presentes no cenário nacional.

Inicialmente, os *tópicos* devem ser divididos em seis tipos: *assunto*, *evento*, *lugar*, *organização*, *pessoa* e *produto*, que, além de serem *categorias* de fácil distinção, se assemelham aos modelos taxonômicos de cunho generalista mais comumente utilizados na Web. Sendo assim, teríamos, a título de exemplo, o seguinte resultado: *Política* como *assunto*, *Sindicância sobre nova 'farra de passagens'* como *evento*, *Distrito Federal* como *lugar*, *Senado Federal* como *organização*, *Heráclito Fortes* como *pessoa* e *iPhone* como *produto*. Posteriormente, novos tipos poderão ser estabelecidos, conforme a necessidade dos produtos jornalísticos, de modo a tornar a classificação dos *tópicos* mais específica.

Por definição, todo e qualquer *tópico* poderá ser tratado como *editoria*, desde que o objetivo seja destacá-lo, sendo incorporado à estrutura de navegação básica do jornal, e aplicando-se um modelo de interface diferenciado. Ou seja, a *editoria* não será encarada apenas como um tipo de *categoria*, mas como uma possível forma de apresentação dos *tópicos*.

Os *tópicos* do tipo *evento* assumirão o mesmo papel que tem as *histórias* nos produtos de jornalismo digital estrangeiros. A única diferença reside no fato de que o *evento* será considerado, por si só, um *tópico*, e não um tipo de *categoria* à parte.

Dado que, de modo geral, os tipos de *tópico* estabelecidos nesta solução têm especificidades completamente diferentes entre si, é possível que, assim como as *editorias*, cada tipo de *tópico* possua um desenho de interface diferenciado e, provavelmente, mais adequado às suas necessidades. Por exemplo: enquanto que o *tópico Senado Federal* pode exibir vídeos, galeria de fotos, *podcasts*, matérias relacionadas, enquete e uma lista com os nomes dos senadores, o *tópico Distrito Federal* pode se limitar a apresentar algumas informações sobre a região, sua localização no mapa político do Brasil, bem como matérias e fotos associadas.

Em síntese, para os leitores a oferta de *tópicos*, que abrangem conceitos como *assunto*, *história* (*evento*) e *entidade* (*lugar*, *organização*, *pessoa* etc), se traduzirá em novos eixos dentro da narrativa, enquanto que para os *produtores de conteúdo* uma nova estrutura será disponibilizada, levando-se em consideração os cinco fatores já mencionados, onde o *conteúdo* será tratado, categorizado e inter-relacionado de modo consistente, flexível e sistemático.

Para implantação desta proposta, a Web Semântica deverá ser a base tecnológica utilizada, assim como acontece no NYT, BBC News e MSNBC.COM, sendo possível associar qualquer tipo de *conteúdo* a uma ou mais *categorias* – processo este conhecido como anotação semântica – de modo simples e rápido, durante as etapas de criação e edição do *conteúdo*, proporcionando não apenas o relacionamento entre *conteúdo* e *categoria*, mas entre os

próprios *conteúdos* e *categorias* em si, o que vai ao encontro dos fatores críticos levantados anteriormente.

Idealizada por Berners-Lee, Hendler & Lassila (2001) [2], a Web Semântica é uma extensão da Web atual, que permitirá aos computadores e seres humanos trabalharem em cooperação. Ela interliga significados de palavras e, neste âmbito, tem como finalidade conseguir atribuir sentido aos *conteúdos* publicados na Internet de modo que seja perceptível tanto pelo homem como pelo computador. A proposta da Web Semântica é estender os princípios dos documentos da Web para os dados. Os dados podem ser acessados usando a arquitetura Web, sendo relacionados uns com os outros da mesma forma que os documentos. Isso também implica a criação de uma plataforma comum que permita o compartilhamento e a reutilização dos dados por meio das fronteiras das aplicações, empresas e comunidades, podendo ser processados automática e manualmente, e revelando novos relacionamentos possíveis entre porções de dados.

A estrutura de *categorias* e suas inter-relações serão definidas num modelo formal e explícito chamado *ontologia*, que será armazenada num banco de triplas (*Virtuoso*, *BigOWLIM* etc) – que equivale a uma espécie de banco de dados, porém mais flexível e escalável que os modelos relacionais (*MySQL*, *SQL Server* etc) – capaz de armazenar, processar e extrair informações em larga escala, sendo este o pilar da tecnologia. Um dos maiores diferenciais da *ontologia* é o modo como as informações são rotuladas e interligadas, pois o significado das palavras é preservado, tornando sua manipulação mais amigável e suas respostas compreensíveis ao ser humano.

Da mesma forma que um banco de dados relacional, uma vez concebida a *ontologia* inicial, esta evolui gradualmente, conforme as necessidades de organização e manipulação das informações forem se configurando, tornando-se maior e mais detalhada. Na *ontologia*, porém, esse processo evolutivo é mais simples e natural, sendo a sua manutenção menos onerosa às empresas.

No tocante ao uso da Web Semântica na classificação de *conteúdo*, o NYT é o melhor exemplo de integração da tecnologia ao fluxo de trabalho. Segundo Sandhaus E. & Larson R. (2009) [14], a empresa possui um complexo sistema de anotação semântica, onde o próprio *Content Management System - CMS* sugere *categorias* a serem relacionadas ao *conteúdo*, com base na análise dos dados nele contidos, a partir de uma ferramenta integrada de *Natural Language Processing - NLP*.

Além disso, tanto no NYT quanto na BBC News, observa-se que cada conceito representado na *ontologia* possui um identificador único chamado *Uniform Resource Identifier - URI* – que é uma cadeia de caracteres compacta usada para identificar ou denominar um recurso na Internet – sendo este responsável pela definição do nome de sua *webpage*, o que é considerado uma boa prática do ponto de vista da Web Semântica. Isso vai ao encontro da ideia de *tópico*, dado que o sistema de classificação utilizado em ambos os jornais tem a capacidade de gerar automaticamente, sempre que necessário, uma *webpage* para cada novo conceito criado.

CONCLUSÃO

Embora a solução proposta pareça ser simples, uma série de mudanças precisará ser realizada no âmbito tecnológico (banco de dados, plataforma de publicação etc), operacional (rotulação/anotação do *conteúdo*, extração de dados etc) e, consequentemente, humano, dado que há uma sensível alteração no fluxo de trabalho dos profissionais responsáveis pela gestão de *conteúdo*. Desse modo, aconselha-se que o processo de mudança seja gradativo, possibilitando a adequação do modelo mental, até então em vigor, à nova realidade estabelecida. Ainda assim, é de se esperar que focos de resistência na cadeia produtiva sejam encontrados, até que, finalmente, todos os profissionais envolvidos tenham condições de perceber as vantagens do novo modelo de trabalho, beneficiando-se do mesmo.

Para o sucesso do que aqui se propõe, é preciso que os *produtores de conteúdo digital* reconheçam o valor da mudança, pois, tomando como base as iniciativas levantadas, os ganhos são altos tanto para estes profissionais quanto para os leitores.

Considerando-se a proposta em si, parece claro que os fatores que garantirão o seu sucesso estão sendo atendidos. A implantação deste novo modelo de classificação do *conteúdo* permitirá, e ensejará, a criação de uma grande quantidade de *categorias* cujo objetivo é reunir *conteúdos* afins – com um nível considerável de especialização, onde também serão abrangidas as definições que dizem respeito à *pessoa*, *lugar*, *organização* etc, chamadas aqui de *entidades* – sendo todos os tipos de *conteúdo* associados a inúmeras *categorias*, além da possibilidade das *categorias* serem relacionadas entre si.

Com as *categorias* já estabelecidas e relacionadas, a operação dos *produtores* e *gestores de conteúdo* ganhará agilidade, tornando os processos de manutenção, armazenamento e recuperação das informações mais eficientes. Do lado do leitor, novos eixos na estrutura narrativa poderão ser apresentados, o que proporcionará uma experiência de imersão mais diversificada e profunda durante o consumo dos fatos jornalísticos, levando-o a requisitar mais *conteúdos*. Consequentemente, o tempo de permanência do leitor aumentará, garantindo mais lucros aos jornais *online* brasileiros e, muito provavelmente, também se traduzirá num aumento de valor em suas marcas, devido à percepção positiva por parte do público consumidor.

Finalmente, é importante observar que, ao se colocar em prática a proposta em questão, os jornais *online* brasileiros terão condições de alcançar o que se entende por jornalismo digital de quarta geração – ao se extrapolar a ideia de uma versão para Web de um jornal impresso previamente

existente – constituindo-se em produtos totalmente exclusivos para a Internet, onde o uso de bancos de dados complexos se consolida através da utilização de ferramentas automatizadas e diferenciadas na sua criação e manutenção, sem nada terem a dever aos seus concorrentes estrangeiros, nos termos aqui expostos.

REFERÊNCIAS

1. Bardoel, J. & Deuze, M. (2001). Network journalism: converging competences of media professionals and professionalism. In: *Australian Journalism Review*, S. L., v. 2, n. 23, pp. 91-103.
2. Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. *Scientific American Magazine*, S. L., pp. 35-43, maio 2001.
3. Colle, R. (2002). *Explotar la información noticiosa*. Data minino aplicado a la documentación periodística. Madri: Universidad Complutense de Madrid.
4. Fidalgo, A. (2004). Do poliedro à esfera: os campos de classificação. A resolução semântica no jornalismo online. In: *2º Encontro Nacional de Pesquisadores em Jornalismo*, 2004, Salvador. Anais... Brasília: SBPJor. 1 CD-ROM.
5. Gonçalves, E. (2004). O banco de dados como formato no jornalismo digital. In: *Congresso da LUSOCOM*, 6., 2004, Covilhã. Anais... Covilhã: LUSOCOM. 1 CD-ROM.
6. Machado, A. (1997). Hipermídia: o labirinto como metáfora. In: Domingues, D. (Org.) A arte no século XXI: a humanização das tecnologias. São Paulo: Fundação Editora da UNESP. pp. 144-154.
7. Machado, E. (2004). Banco de dados como formato no jornalismo digital. In: Congresso Lusófono de Ciências da Comunicação, 6., 2004, Covilhã. [Anais...] Covilhã: Lusocom.
8. Manovich, L (2001). *The language of new media*. Cambridge: MIT Press.
9. Mielniczuk, L. (2003). *Jornalismo na web*: uma contribuição para o estudo do formato da notícia na escrita hipertextual. Tese de Doutorado. FACOM/UFBA.
10. Palacios, M. (2002). *Jornalismo online, informação e memória*: apontamentos para debate. Covilhã: Editora Universidade da Beira Interior.
11. Palacios, M., Mielniczuk, L., Barbosa, S., Ribas B. & Narita, S. (2002). *Um mapeamento de características e tendências no jornalismo online brasileiro e português*. Comunicarte, Aveiro, v. 1, n. 2.
12. Palacios, M. (2003). Ruptura, continuidade e potencialização no jornalismo online: o lugar da memória. In: Machado, E. & Palacios, M. (Orgs.). *Modelos do jornalismo digital*. Salvador: Editora Calandra.
13. Pavlik, J. (2001). *Journalism and new media*. Nova York: Columbia University Press.
14. Sandhaus E. & Larson R. (2009). *A century of semantic technology*: semantics at The New York Times.
15. Schwingel, C. (2005). Jornalismo digital de quarta geração: a emergência de sistemas automatizados para o processo de produção industrial no jornalismo digital. In: *Compós*, 2005, Niterói. Anais... Niterói: Compós. 1 CD-ROM.
16. Zuin, A. L. A. & Correia, C. M. de C. (2008). Jornalismo impresso e jornalismo online: a linguagem híbrida da informação. In: Simpósio Nacional de Pesquisadores em Comunicação e Cibercultura - ABCIBER, 2., 2008, São Paulo. [Anais...] São Paulo: PUC-SP.