

14. Recuperação de informação em jornais on-line: atributos de pesquisa, mecanismo de busca e percepção profissional

Antonio Paulo Carretta, Vânia Mara Alves Lima

Estudo analisa questões de organização e recuperação de informação em repositórios de jornais on-line. Destaca aspectos do suporte hipermídia, estrutura informativa do documento digital e gênero do conteúdo da informação jornalística on-line; aborda a noção de memória como atributo de ativação e conexão de informações no contexto da Web; descreve a estrutura básica de mecanismos de busca e traça o perfil de jornalistas no âmbito da convergência digital. Para investigar potenciais dificuldades de pesquisa e recuperação de informação, adota pesquisa exploratória para inspeção das interfaces similares de mecanismos de busca de jornais selecionados, nacionais e estrangeiros, e aponta indicadores da percepção de usuários especialistas, jornalistas e profissionais da área de comunicação, no processo de recuperação de informação por meio destas ferramentas de busca interna. Como resultado, discute-se sensibilidades dos padrões técnicos de tratamento da informação, carências do processo de pesquisa e fatores de satisfação para recuperação de informação em ambiente digital.

Recuperação de Informação. Mecanismo de Busca. Jornalista.
Jornal Online. Informação Jornalística.

Introdução

Desde as primeiras gerações de jornais digitais, uma significativa mudança tem ocorrido na produção, distribuição e consumo da informação jornalística disponível na web. Segundo Pariser (2012), está ocorrendo uma queda constante no valor para produção e a distribuição de qualquer tipo de mídia (palavras, imagens, vídeos e áudio) se aproximará do custo zero, conseqüentemente, provocará uma constante necessidade de escolhas e seleções para filtrar um grande volume de informação. Além disso, uma maior apropriação de ferramentas de busca é observada pela dimensão de sua aplicação e uso: cerca de 40%¹ da população mundial é usuária da internet, mais de 90% e utiliza frequentemente mecanismo de busca para obter informações/notícias².

Neste atual panorama, surgem novas tendências de organização da informação (divulgação e classificação de conteúdo), projetos de digitalização de documentos, incorporação de arquivos retrospectivos, personalização da disseminação de notícias e, como resultado dos esforços para melhorar e agilizar a capacidade de tecnologia de pesquisa: a inovação dos mecanismos de busca e modificações na experiência de pesquisa para recuperação e divulgação de acervos digitais. Estes indicadores informais de mudança (técnica ou estratégica) são percebidos em jornais nacionais ou

¹INTERNET LIVE STATS. Disponível em: <www.internetstats.com> Acesso em: 20 out. 2014

²MÍDIA DADOS BRASIL 2014. São Paulo: Grupo de Mídia São Paulo, 2014. Disponível em: <www.gm.org.br> Acesso em: 20 out. 2014

estrangeiros com produção de conteúdo digital. Entretanto, mesmo observando que as ferramentas de busca são frequentes em sites com permanente produção e arquivamento de informação, estes serviços para recuperação de informação possuem reduzidos, irregulares e insatisfatórios atributos de pesquisa. Conseqüentemente, profissionais que atuam em redações on-line lidam diariamente com este problema que, carente de uma análise mais profunda dentro do contexto profissional, necessita de uma avaliação sobre seu impacto na rotina de trabalho. Portanto, observada pela relação entre mídia impressa e digital, que provoca uma mudança de cultura do saber profissional, o principal argumento deste projeto é ampliar estudos em tecnologia de pesquisa e acesso à informação. Para isso, a proposta de estudo está delimitada na avaliação dos atributos de pesquisa da informação jornalística on-line, comparação dos mecanismos de busca de sites jornalísticos e perfil dos profissionais-pesquisadores (relação ferramenta-usuário), inseridos no atual ambiente de web jornalismo e durante a prática de pesquisa de conteúdo e recuperação de notícias.

Objetivos

A condução de explorações e experimentações propostas neste estudo tem por objetivo geral analisar o comportamento de pesquisa de usuários especialistas para recuperação de informação em jornais on-line. Para isso, as características específicas de estudo são: analisar atributos de pesquisa da ferramenta de busca; identificar necessidades de informação e tarefas de pesquisa do usuário; reconhecer dificuldades no

processo de pesquisa; identificar características de usabilidade da ferramenta e satisfação do usuário. Todas as características têm como foco a percepção e experiência de jornalistas sobre o uso desta ferramenta em sua prática de pesquisa. Neste estudo o mecanismo de busca será entendido como uma ferramenta (interface) para comunicação entre objeto (informação) e indivíduo (usuários que fazem parte de um grupo social formado por profissionais atuando em redações de jornais). Como resultado final deste estudo pretende-se apresentar uma avaliação dos atuais atributos de informação disponíveis para pesquisa e recuperação de conteúdo jornalístico e uma síntese do comportamento de pesquisa dos usuários especialistas. Estas propostas visam à identificação de parâmetros de avaliação da ferramenta e indicadores (da satisfação, limitação ou modificação) das práticas de uso, que permitam reduzir problemas nas estratégias de recuperação de informação e propor recomendações para melhorar a usabilidade da ferramenta de trabalho.

Justificativa

Diariamente, sites de notícias produzem milhares de informações que são publicadas em suas diversas áreas editoriais. Para recuperar uma informação veiculada exclusivamente no ambiente digital e publicada na semana anterior, ou no ano passado, em princípio, um profissional da produção on-line utiliza apenas o mecanismo de busca oferecido pelo site. Sendo assim, é fato que para um jornalista que lida com a busca de notícias retrospectivas, cuja origem de produção é apenas on-line, o ambiente de navegação e seus recursos são os mesmos dos

leitores de portais de notícias que procuram uma informação qualquer. Para este profissional recuperar algum dado que o ajude, por exemplo, compor um texto de memória, é necessário iniciar uma busca no conteúdo on-line do jornal. É neste momento que surgem imprecisões e falhas de recuperação. Em meio a milhões de páginas com atualizações sistemáticas, assim como diferentes formatos de informação, tanto o leitor comum, quanto o profissional acostumado com a estrutura do site e do fluxo de informações, encontram dificuldades em explorar, localizar e determinar a relevância das matérias apresentadas nos resultados de busca. Observando este quadro, levantamos duas indagações:

- 1) Os mecanismos de busca utilizados para recuperação de informações jornalísticas possuem estruturas e atributos de pesquisa adequados?
- 2) A ferramenta para recuperação de informação satisfaz o jornalista em sua rotina de trabalho e sua dinâmica de pesquisa?

Estas questões justificam e norteiam os experimentos pretendidos neste trabalho e pontuam o embasamento teórico, assim como observações, análises, métodos de pesquisa e tópicos selecionados.

Procedimentos metodológicos

O processo de construção teórico-empírico da pesquisa está dividido em cinco grandes etapas metodológicas, que são apresentadas e detalhadas da seguinte forma:

BASE TEÓRICA

Etapa 1: Análise de conceitos para fundamentação teórica da relação entre processo de pesquisa, comportamento e experiência no uso de informação por meio de uma ferramenta de busca. Levantamento com enfoque multidisciplinar (Comunicação, Arquitetura da Informação e Computação) para apoio a linha de pesquisa da Ciência da Informação.

BASE EMPÍRICA

Etapa 2: Análise de interfaces Similares - Elaboração de quadro com características de buscadores dos jornais que são referência para pesquisa dos usuários estudados. Inspeção baseada nos principais elementos de identificação da informação, considerados aqui como essenciais para organização e recuperação da informação jornalística. Estes elementos estão presentes em cada unidade de informação produzida: data e período (de publicação da notícia), título, autoria da matéria, área editorial (editoria que produz e agrupa texto de mesma linha temática).

Etapa 3: Análise do usuário especialista (produção, organização e consumo de informação) por meio de literatura: bibliografia da área de comunicação, avaliação e descrição de informações extraídas de pesquisas de perfil profissional, manuais de práticas jornalísticas, sites de associações profissionais e empresas jornalísticas (ex. ANJ).

Etapa 4: Elaboração de questionário (online) para identificação do comportamento de pesquisa, de jornalistas, assim como as impressões sobre o uso de buscadores como ferramentas de trabalho. Formato de aplicação com questões fechadas: uso justificado pela capacidade e agilidade da

ferramenta para aplicação e levantamento de dados sobre hábitos, comportamento e opinião.

Etapa final: Análise dos questionários pontuando perfis de usuários, tarefas, tendências de pesquisa, recuperação de informação e usabilidade dos mecanismos de busca.

Fundamentação

Com intenção de promover associações com o pensamento da chamada “era da informação” e a Ciência da Informação (CI), o contexto que abordamos neste estudo (usuário/informação/recuperação) tange o tradicional tratamento de informação e expande para o contexto digital (de rede, hiperlinks e navegação) que conecta informações e pessoas. Tratamos das percepções que envolvem tecnologia, informação e os paradigmas de referência relacionados à complexidade estrutural da web e da informação digital de conteúdo jornalístico, que exige enfoques multidisciplinares. Portanto, para facilitar o entendimento do quadro teórico, conceitos operacionais são extraídos das áreas de Comunicação, Arquitetura da Informação e Computação. Muito embora estes conceitos sejam conteúdo de apoio para CI, exploramos sua aplicação de forma a criar dimensões de análise e fundamentar observações e relações da investigação com: aspectos da informação jornalística on-line; memória digital; recuperação da informação (por meio de um mecanismo de busca); noções de usabilidade de busca e do sistema de informação jornalística; perfil do jornalista em ambientes digitais.

Na discussão destes aspectos, pontuamos atributos das informações em sites de notícias e como se apresentam, organizam e recuperam seu conteúdo. Visando complementar esta análise e indicar etapas de desenvolvimento do jornalismo na web, consideramos a classificação de webjornalismo proposta por Machado, Borges e Miranda (2003) na qual se diferenciam as gerações de webjornais e seus modelos de produção de conteúdo. Além disso, analisamos o meio digital como fonte de pesquisa do usuário profissional, o jornalista (MACHADO, 2003), e papel do veículo de comunicação como ferramenta (ou sistema) de recuperação de informação (MANNARINO, 2000).

Ao tratar da estrutura do mecanismo de busca e recuperação de Informação na Web, inserimos discussão sobre expressão de busca e uso de operadores booleanos, a partir da abordagem de Gutiérrez (2000), que considera a formação de linguagem de pesquisa apoiada em uma gramática formal para recuperação de informações. Sobre usabilidade, além das noções gerais, consideramos também os fatores (frequência, impacto e persistência) que afetam uso de mecanismos de busca (NIELSEN; LORANGER, 2007).

Para avaliar o comportamento de pesquisa dos jornalistas, que muito recentemente lidam com a fusão ou aproximação que os jornais têm provocado entre redações da versão impressa e online, coletamos dados de pesquisas (FIGARO, 2012) que esclarecem a rotina de tarefas e traçam perfil profissional. Dentro deste contexto, são consideradas as novas habilidades experimentadas no jornalismo digital, indicadas pelo jornalista Mark Briggs (2007).

Por fim, introduzimos uma abordagem focada no usuário e no processo de pesquisa (informação, consumo e uso), com base teórico-conceitual e análises fundamentadas pelos estudos da CI, que definem modelos de processo de pesquisa, comportamento e experiência do usuário, a exemplo dos estudos de Kuhlthau (2004) e Ingwersen e Järvelin (2005) que, muito embora estejam focados numa ampliação do ponto de vista cognitivo, propõem uma integração de análises dos estudos sobre o processo de busca (proposto em uma dimensão humana pela Ciência da Informação) e recuperação de informação (com suas bases próximas da Ciência da Computação).

Resultados esperados

Norteados por duas indagações, a primeira sobre a estrutura dos atributos de pesquisa e a segunda na satisfação de uso da interface de busca, ambas focadas na recuperação de informação. Os resultados esperados, durante o período de investigação, consistem em demonstrar que:

1) Baseados na estrutura da informação jornalística, atributos de pesquisa dos mecanismos de busca de jornais online apresentam recursos não padronizados, reduzidos e pouco explorados. Em conflito com a dinâmica de evolução tecnológica aplicada no ambiente Web, construídos em narrativa de hipertexto e hiperlinks, buscadores de notícias possuem capacidade limitada para filtrar informação relevante, ou mesmo criar relações semânticas por meio de conexão entre todas as unidades de informação (texto, imagem, vídeo, infográfico etc.). Esta condição implica em desagregação do conteúdo gerado em

crescimento constante e, conseqüentemente, invisibilidade, ruído na recuperação e uso desarticulado da linguagem jornalística. Além de impedimentos para inovação na representação e organização de conhecimento, por exemplo, uso de ontologias, ou construção de modelos elaborados para visualização de resultados de pesquisa.

2) Ocorre insatisfação do profissional jornalista ao utilizar a ferramenta de busca em sua rotina de pesquisa, derivada de fatores estruturais que afetam a severidade do problema de busca, frequência de usuários que reconhecem comprometimento da pesquisa, indicadores de impacto (tempo gasto) e persistência (dificuldade contínua). Em relação direta com a geração de atributos de pesquisa, o grau de satisfação apresentado pode justificar investimentos no processo de recuperação de informação, uso avançado e consistente de padrões de metadados, assim como maior conexão entre ciclo de produção de notícias, aplicação de meta tags, interfase de busca, necessidade do usuário e resultados da pesquisa.

No resultado final deste estudo, em sintonia com a necessidade do usuário profissional, pretende-se ainda apresentar indicações de boas práticas para otimização da interface de busca e recomendações de estruturação da informação jornalística, referenciando uma proposta de taxonomia para metadados de notícias, criada e mantida pelo International Press Telecommunications Council (IPTC), apoiada em modelo SKOS (Simple Knowledge Organization System) para padronização e compartilhamento de dados na Web.

Considerações preliminares

Empresas de informação, que geram e utilizam documentos em ambientes on-line, começam a formar grandes repositórios de informação híbrida, de estrutura complexa e conteúdo com dinâmica associativa e modelado para atingir redes sociais.

O documento digital adquiriu nova forma (formatos) e é mais denso, orgânico, com simultâneas e múltiplas conexões. Considerando que as tecnologias de informação e comunicação, geradas pela Internet, trouxeram novas dimensões físicas e temporais (espaço e tempo) no tratamento da informação, assim como novas perspectivas de ordem digital, fica evidente que os paradigmas em ambientes digitais começam a sofrer novas transformações (ou interpretações): volatilidade de dados, velocidade de ações e interatividade (de pessoas e ideias) são novos paradigmas.

Os resultados iniciais, da inspeção de mecanismos de busca desenvolvida nesta pesquisa, indicam que os modelos de mecanismo de busca selecionados não operam com padrões comuns e absorvem as mudanças provocadas pela tecnologia disponível e também pela lógica comercial de empresas na Web. O volume de informação gerado pelo meio é outro fator que implica em uma série de indagações: Toda a informação contida no repositório de textos produzidos é recuperável? O mecanismo de busca (interno) dos sites selecionados oferece soluções ideais para pesquisa? Como recuperar informações de áreas restritas (páginas especiais, blogs de comentaristas, infográficos interativos, serviços de vídeo e podcast)? Como facilitar a recuperação de informações perdidas em sistemas

desestruturados? Quais linguagens de recuperação podem ser aplicadas? Que padrões de representação deste conhecimento precisam ser padronizadas ou desenvolvidas?

Diante de tantas questões, percebe-se que a necessidade de ampliar estudos em tecnologia de pesquisa e acesso à informação, aplicados na avaliação do tráfego de informação jornalística em rede, exigirão do profissional em CI maior compreensão sobre a mudança da modelagem conceitual e do ambiente digital. Para isso, novas abordagens precisam fornecer reflexões sobre parâmetros fundamentais na avaliação de sistemas de informação em hipermídia, linguagem documentária de hipertextos, bem como análises do comportamento de pesquisa dos usuários on-line e indicadores de limitações ou modificações das práticas utilizadas na organização e recuperação de informação digital.

Principais referências

BRIGGS, Mark. **Jornalismo 2.0**: como sobreviver e prosperar: um guia de cultura digital na era da informação. Knight Center for Journalism in the Americas, 2007.

FIGARO, Cláudia. (Coord.). **O perfil do jornalista e os discursos sobre o jornalismo**: um estudo das mudanças no mundo do trabalho do jornalista profissional em São Paulo [Relatório Final da Pesquisa]. São Paulo: Centro de Pesquisa em Comunicação e Trabalho; ECA/USP, 2012. Disponível em: <http://www.eca.usp.br/Comunicacaoetrabalho/wp/wp-content/uploads/relatorio_final_2012.pdf> Acesso em: 20 ago. 2013.

GUTIÉRREZ, Mario Pérez. El lenguaje de interrogación: una gramática formal para recuperación de información. **Revista Española de Documentación Científica**, v. 23, n. 3, p. 246-266, 2000.

INGWERSEN, Peter; JÄRVELIN, Kalervo. **The turn**: integration of information seeking and retrieval in context. Dordrecht: Springer, 2005.

KUHLTHAU, Carol Collier. **Seeking meaning**: a process approach to library and information services. Westport: Libraries Unlimited, 2004.

MACHADO, Elias. **O ciberespaço como fonte para jornalistas**. Salvador, BA: Calandra, 2003. (Coleção Biblioteca J)

MACHADO, Elias. BORGES, Clarissa. MIRANDA, Milena. Modelos de produção de conteúdo no jornalismo digital baiano.

In: MACHADO, Elias. PALACIOS, Marcos. (Org.) **Modelos de jornalismo digital**. Salvador: Edições GJOL, Calandra, 2003.

MANNARINO, Marcus Vinicius Rodrigues. **O papel do web jornal**: veículo de comunicação e sistema de informação. Porto Alegre: EDIPUCRS, 2000. (Coleção Comunicação, 5)

NIELSEN, Jakob; LORANGER, Hoa. **Usabilidade na web**: projetando websites com qualidade. Rio de Janeiro: Elsevier, 2007.

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