Ferramentas de Busca da Informação na Web de Dados

Francisco Carlos Paletta¹, José Antonio Moreiro González²

¹ <https://orcid.org/0000-0002-4112-5198>, Universidade de São Paulo, Brasil. fcpaletta@usp.br

² <https://orcid.org/0000-0002-8827-158x>, UC3M, Espanha. jamore@bib.uc3m.es

Palavras-chave: Informação; Web de Dados; Ciência Aberta; Ferramentas de Busca da Informação.

1 Introdução

A Web é uma rede cujos conteúdos estão interligados através de documentos de hipertexto. Seu estudo é possível por um processo de análise e coleta sucessiva das páginas, a partir de um conjunto de sites previamente conhecidos. Essa busca é feita de forma automática por um programa de computador normalmente chamado de crawler (um programa que sistematicamente browsers “procura” a World Wide Web) com objetivo de criar uma indexação de dados.

Nem toda a Web está interligada, contudo, embora a maior parte dela esteja, há “ilhas” de tamanhos variados sem ligação com o restante da rede. Isso significa que o conjunto inicial de sites a partir dos quais a pesquisa é feita influencia o resultado, e encontrar o conjunto adequado, geralmente o mais completo possível, é um passo importante. O primeiro princípio da Web, proposto pelo W3C Brasil (Consórcio World Wide Web), afirma que “o principal valor da Web é o social. Mais do que tecnológico, este é um ambiente de comunicação humana, de transações comerciais, de oportunidades para compartilhar conhecimentos e, para ser um ambiente universal, deve estar disponível para todas as pessoas, independentemente dos equipamentos e softwares que utilizem, principalmente da cultura em que se inserem, da localização geográfica, das habilidades físicas ou mentais, das condições socioeconômicas ou de instrução”. A universalidade da Web só pode ser garantida e aprofundada com um modelo de governança democrático e pluralista que tenha foco no acesso por todos, e na sua própria evolução tecnológica (CGI.BR, 2010).

2 Ciência Aberta e a Web de Dados

As constantes mudanças nas formas de organização tradicional da informação presentes em ambientes informacionais digitais são reflexos da incorporação das Tecnologias de Informação e Comunicação (TICs), como no caso das bibliotecas e repositórios digitais, que armazenam, preservam, disseminam e permitem o acesso a produção acadêmica e científica, visando contribuir para o aumento da visibilidade e do valor da instituição ao agregar recursos que possibilitam o processo de construção do conhecimento, a partir da participação colaborativa aplicada em diferentes ambientes.

A Web 2.0 segue uma filosofia com princípios de leitura e escrita de natureza participativa, em que cada usuário pode intervir diretamente na escolha e introdução de dados no âmbito de cada site; é cooperativa, uma vez que compartilha ideias, preferências, informações e conhecimento; é interativa, na medida em que, através de toda a gama de recursos multimídia, é possível um diálogo simultâneo com os usuários; é democrática, pois sob essa filosofia existe liberdade de expressão, de

pensamento, e, sobretudo, de trânsito de informações, independentemente dos interesses de cada um; é também socio-técnica, pois, através de todas as suas características, é possível um intercâmbio de culturas, religiões, etnias e outros.

No contexto da Biblioteca Digital a maior parte dos pesquisadores concordaria que muito do que as bibliotecas aprovaram na primeira revolução da Web são estáticos. Por exemplo, catálogos online de acesso público (OPAC) exigem que os usuários busquem a informação. Embora muitos estejam iniciando a incorporar técnicas da Web 2.0 relativas à pesquisa de dados, eles não respondem com recomendações, tal como a Amazon, que se apresenta com um maior dinamismo. Do mesmo modo, a primeira geração de biblioteca online foi elaborada através de textos tutoriais estáticos e que não respondiam às necessidades dos usuários, nem permitiam que interagissem uns com os outros.

As bibliotecas, porém, tem começado a evoluir numa estrutura mais interativa, meios de comunicação social e rico em tutoriais, programação e animação com o uso de banco de dados mais sofisticados. A Web Semântica nas bibliotecas pode ser uma ferramenta que possibilite a gênese de uma base de conhecimento a partir da inteligência coletiva, como também ferramenta para a gestão do conhecimento que facilite, de maneira interativa, a descoberta deles. Passamos de uma biblioteca para o usuário para uma biblioteca com o usuário.

Ao integrar e complementar estes dois paradigmas, devemos também considerar, conhecer, explorar e avaliar as novas ferramentas de comunicação, organização, participação e construção coletiva do conhecimento que estão disponíveis na Web. São aplicações, a nosso favor, na sua maioria intuitivas, gratuitas e que respondem as novas necessidades de informação e participação da comunidade. Deste modo, fica clara a transformação, ou seja, a nova abordagem da relação entre informação e conhecimento no contexto das bibliotecas e dos repositórios, ao contrário dos ambientes tradicionais que só permitia ao usuário uma única forma de se relacionar com o conteúdo armazenado fisicamente. Hoje, não só o usuário participa, como também adiciona conteúdo que ao serem compartilhados em outros recursos colaborativos passam a receber comentários dos membros vinculados à sua comunidade agregando valor à Informação compartilhada, além de criar vários caminhos para a localização do próprio recurso. Assim, temos uma nova relação e interação na qual aumenta o nível sofisticado de atitudes sociais digitais (CITILAB, 2009).

A Internet é uma rede que interliga várias redes. Seu serviço mais importante é a navegação nas páginas da rede mundial de computadores, a WWW (a Web de Dados). A evolução da Internet deu-se de maneira espontânea, ou seja, não houve uma trajetória pré-determinada ou um projeto definido organizado por algumas pessoas ou instituições. As inovações abriram um campo de ação para o surgimento de inovações complementares. Foi criado um ambiente propício, principalmente nos Estados Unidos a princípio por meio da agência de fomento ARPA. Segundo Castells (2003, p. 24) a "ARPA foi uma agência financiadora e orientadora de pesquisas dotada de considerável autonomia que veio a se tornar uma das mais inovadoras instituições de política tecnológica do mundo, não apenas em torno da interconexão de computadores, mas em vários campos decisivos de desenvolvimento tecnológico." (Castells, 2003, pg. 24)

A ARPA tinha orientação militar, mas apoiava projetos de centros de pesquisas universitários. Sua principal divisão foi a ARPANET (1970) na qual surge claramente a orientação para desenvolvimento de redes de comunicação. A partir daí, as inovações que surgiram foram determinantes para o estabelecimento da Internet, a rede que interliga várias redes possibilitando acesso a WWW viabilizando a Ciência Aberta.

3 Ferramentas de Busca da Informação na Era Digital

O Buscador (Search Engine) permite recuperar informações na Web de Dados. Os buscadores são fundamentais no processo de navegação na Web, pois sem eles, seria necessário conhecer o endereço da página buscada com antecedência o que tornaria a navegação muito mais complicada. O software vasculha a rede em busca de informações relativas à consulta digitada na “chave” de busca e organiza uma lista de sites com os resultados obtidos (SERP - Search Engine Results Page) para avaliação. A extração da informação na Web de dados é realizada de modo automático e, normalmente, as primeiras páginas fornecem informações mais precisas e adequadas à consulta.

A Tabela 1 apresenta os buscadores mais utilizados em todas as plataformas (desktop, telefonia, tablet, console) no período de 09/ 2017 a 09/2018.

Tabela 1: Principais buscadores em todas as plataformas (desktop, celular, tablet e console)

Google	92,3%
Yahoo	2,51%
Bing	2,28%
Baidu	0,85%
YANDEX RU	0,61%
Duckduckgo	0,33%

Fonte: Stat Counter: <http://gs.statcounter.com/search-engine-market-share>

Os resultados obtidos por meio dos buscadores podem ser mais específicos e atender melhor às necessidades dos usuários com a utilização dos recursos e operadores de pesquisa. Atualmente, a Web de Dados é uma fonte inesgotável de informações com mais de 1.921.791.628 páginas na Web de Dados (fonte: <http://www.internetlivestats.com/>). Somente no buscador Google são realizadas mais de 2.738.900.450 (Fonte: <http://www.internetlivestats.com/>) de consultas diariamente. A busca na Web é uma atividade que faz parte do cotidiano das pessoas. Os processos de busca na Web podem ser simplificados pelos operadores e recursos que refinam os resultados da pesquisa. Desta forma é possível simplificar, limitar ou expandir os resultados durante o processo de busca na Web de Dados.

4 Resultados

Como principal resultado deste trabalho de pesquisa está a proposta de criação de uma atividade de cultura e extensão que apresente as principais “Ferramentas, Operadores e Recursos para Otimização de Resultados de Buscadores na Internet”.

Na Era da Informação e, mais precisamente, na conjuntura de rede onde estamos imersos, muitos são os desafios e os problemas que não podemos apenas enfrentar, mas devemos estudar com minúcia e profundidade a fim de irmos encontrando as respostas globais e específicas mais adequadas (Silva, 2008).

Além de uma fonte de pesquisa, a Internet é uma área de estudos multidisciplinar Silva (2008) destaca os principais focos deste campo de estudos sobre Internet:

- ❖ tecnologia (design e desenvolvimento);
- ❖ utilização;
- ❖ políticas e processos de governança da Internet.

Estes tópicos constituem a base sobre a qual fundamentam-se os estudos sobre Internet, uma área de estudos em expansão que exige pesquisas multi-inter-transdisciplinares. Os operadores de pesquisa podem facilitar a busca, acesso, uso e apropriação da informação na Web de dados. Torna-se fundamental entender os recursos disponibilizados por esses operadores com objetivos de:

- ❖ aprimorar as habilidades e as estratégias de pesquisa;
- ❖ elevar o nível de conhecimento e desempenho individual visando à melhoria na qualidade do desenvolvimento das atividades de pesquisa;
- ❖ complementar o aprendizado de técnicas de pesquisa para elevar o nível de eficácia e de eficiência em seus trabalhos.

Desenvolver competências informacionais é garantir autonomia ao usuário da Informação nos processos de pesquisa na Internet. O desenvolvimento das habilidades de pesquisa online facilita o aprendizado e a compreensão dos temas associados aos processos de busca, acesso e apropriação da Informação na Web de Dados. Dentre as principais habilidades destacamos:

- ❖ Pesquisa online em buscadores;
- ❖ Como funciona os buscadores;
- ❖ Características da página de resultados (SERP);
- ❖ Apresentação dos otimizadores de resultados de busca.

É fundamental que o usuário da informação tenha competências que permitam criar planos de pesquisa e elaborar estratégias para solucionar os desafios que existem em qualquer percurso profissional e/ou acadêmico no contexto de busca, acesso e apropriação da informação na Web de Dados.

5. Conclusões

A Era Digital demanda a necessidade de desenvolver novas competências a partir de uma nova forma de interatividade entre usuário e a internet. A Web Semântica ou Web Inteligente está a permitir a construção de uma internet mais inteligente que produza uma transformação digital: a revolução 4.0. Com o uso de novas tecnologias é imperativo o uso das TICs em tornar as coisas mais fáceis e agilizar os processos de busca de informação e geração de novo conhecimento. A WEB Semântica é uma Web com toda sua informação organizada de forma que não somente seres humanos possam entendê-la, mas principalmente máquinas. É neste ponto que surge um novo usuário da informação com novas demandas por recursos computacionais e novas capacidades em produzir novos conhecimentos.

5 Referencias

Castells, M. (2003). *A Galáxia Internet: reflexões sobre a Internet, negócios e a sociedade*. Rio de Janeiro: Jorge Zahar Editor Ltda.

Choo, C. W. (2003). *A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões*. São Paulo: Editora Senac São Paulo.

CITILAB – Observatório para a Cibersociedade (2009). *Ambientes informacionais: as bibliotecas digitais e os repositórios institucionais no contexto da web 2.0*. Retrieved from <http://www.cibersociedad.net/congres2009/es/coms/ambientes-informacionais-as-bibliotecas-digitais-e-os-repositorios-institucionais-no-contexto-da-web-20/1056/>

CGI.BR. (2010). *Censo da Web.br 2010 - Dimensões e características da Web brasileira: um estudo do .gov.br*. NIC.BR/CETIC.BR. Retrieved from <http://www.cgi.br/publicacao/censo-da-web-br-dimensoes-e-caracteristicas-da-web-brasil-um-estudo-do-gov-br/>

Dutton, W. H. (2013). *The Oxford Handbook of Internet Studies*. Oxford: University Press.

Eurocid - *Estratégia de Lisboa*. (2000). Retrieved from http://www.eurocid.pt/pls/wsd/wsdwcot0.detalhe?p_cot_id=1079&p_est_id=5266

Silva, A. M. (2008). **Inclusão Digital e Literacia Informacional em Ciência da Informação**. *PRISMA.COM*, 7, pp.16-43.
Retrieved from <http://revistas.ua.pt/index.php/prisma.com/article/view/683>

TIC DOMICÍLIOS. (2017). *Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros [livro eletrônico]*: TIC domicílios 2017. Núcleo de Informação e Coordenação do Ponto BR [editor]. São Paulo: Comitê Gestor da Internet no Brasil, 2016. Retrieved from <https://cetic.br/noticia/acesso-a-internet-por-banda-larga-volta-a-crescer-nos-domicilios-brasileiros> e <https://cetic.br/pesquisa/domicilios/>

ACKNOWLEDGMENT: FAPESP Research Project - Processo 2016 / 07358-6