

O sistema de busca das revistas científicas eletrônicas: implicações para o acesso aberto

PAULA CAROLINE SCHIFINO JARDIM PASSOS

SÔNIA ELISA CAREGNATO

Programa de Pós-Graduação em Comunicação e Informação

Universidade Federal do Rio Grande do Sul (UFRGS)



Resumo

Os sistemas de busca das revistas científicas eletrônicas consistem em aplicações de software com campos que o usuário pode preencher a fim de solicitar a recuperação de algum assunto. A presente pesquisa analisa a busca em três revistas de renome internacional Nature, Science e PLOS Biology. Para tanto, inicialmente apresenta-se breve revisão bibliográfica sobre o sistema de busca e seus os componentes. A seguir, demonstra-se os resultados da análise. O método escolhido foi a inspeção por lista de verificação, realizada mediante sentenças teste descritas para esta pesquisa com base nos autores de referência na área Rosenfeld, Morville e Arango (2015), Kalbach (2009) e Garrett (2011). A verificação sistemática e aprofundada permitiu elencar os recursos aplicados à busca em revistas científicas de alta visibilidade. Por outro lado, foi possível identificar pontos onde algumas modificações trariam benefícios à experiência dos usuários das revistas. Além disso, por meio da comparação entre as três revistas, evidenciou-se as melhores soluções utilizadas em cada um dos mecanismos de busca. Considerando a importância das revistas para a comunicação científica, entende-se como relevantes estudos voltados à constante atualização de seus sistemas. A análise da busca contribui para o aprimoramento desses veículos de informação e, conseqüentemente, promove a visibilidade dos textos publicados.

Palavras-chave

Sistema de busca; revista científica eletrônica; comunicação científica.

Abstract

The search engines of electronic scientific journals consist of software applications with fields that the user can fill in order to request the retrieval of some subject. The present research analyzes the search in three internationally renowned journals Nature, Science and PLOS Biology. In order to do so, we present a brief bibliographic review of the search system and its components. The results of the analysis are shown below. The method chosen was the inspection by checklist, carried out using test sentences described for this research based on the reference authors in the area Rosenfeld, Morville and Arango (2015), Kalbach (2009) and Garrett (2011). Systematic and in-depth verification enabled us to list the resources applied to the search of high visibility scientific journals. On the other hand, it was possible to identify points where some modifications would bring benefits to the users' experience of the magazines. In addition, through the comparison between the three journals, the best solutions used in each of the search engines were evidenced. Considering the importance of journals for scientific communication, it is understood as relevant studies aimed at the constant updating of their systems. The analysis of the search contributes to the improvement of these information vehicles and, consequently, promotes the visibility of the published texts.

Keywords

Search system; electronic scientific journal; scientific communication.

Introdução

Os Sistemas de Busca consistem em aplicações de software em campos que permitem ao usuário consultar o sistema a fim de recuperar conteúdos acerca de algum assunto. É possível utilizar linguagem natural ou operadores booleanos. As solicitações dos usuários são cruzadas com um índice, formado por todos os termos encontrados nos documentos ou por uma lista com títulos, autores, categorias e informação relacionada. Metadados também podem identificar os documentos armazenados (AGNER, 2009). A Figura 1 representa o sistema de busca descrito por Morville e Rosenfeld (2006).

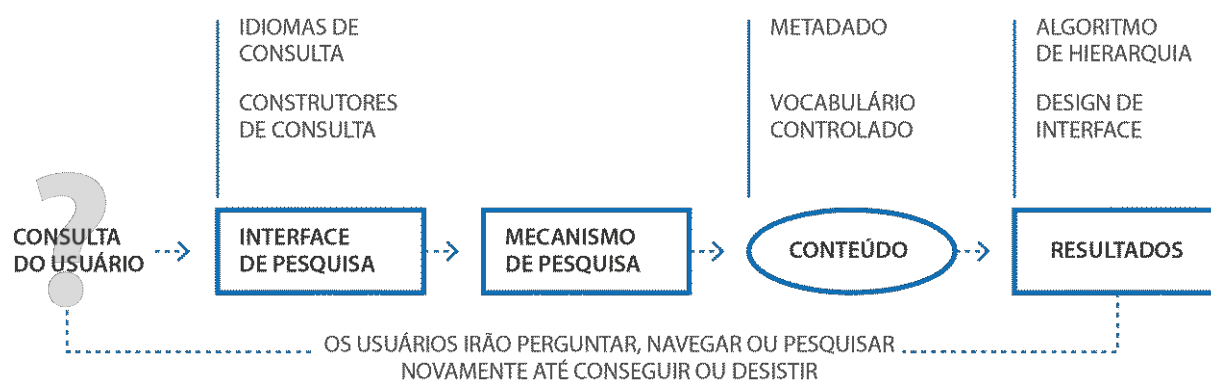


Figura 1. Sistemas de Busca

Fonte: Morville e Rosenfeld (2006, 14)

Conforme Rosenfeld, Morville e Arango (2015), mesmo que a literatura especializada apresente muitos estudos referentes ao sistema de busca, o grande número de variáveis envolvidas (nível de conhecimento e motivação do usuário, tipos e quantidade de informação pesquisada etc.) impede o desenvolvimento de uma interface de busca ideal. Contudo, a partir de Rosenfeld, Morville e Arango (2015) e de Kalbach (2009) é possível afirmar que um sistema de busca deve ser eficiente tanto em nível de granularidade grossa —zonas de pesquisa (como tipo de conteúdo, público, assunto, cronologia etc.)—, quanto em nível mais refinado de pesquisa, ou seja, em componentes do conteúdo dentro do próprio documento (como título do artigo, autor, afiliação, palavras-chave etc.).

Os usuários nem sempre sabem expressar aquilo que estão buscando, mas boas interfaces de busca podem auxiliá-los a encontrar aquilo que procuram. Para Rosenfeld, Morville e Arango (2015), a apresentação dos resultados de pesquisa também se configura como uma possibilidade de estreitar a busca. Nesse sentido, e ainda visando qualificar solicitações de busca dos usuários inexperientes, os autores recomendam o uso de construtores de consulta, ou seja, ferramentas que podem melhorar o desempenho de uma busca.

Os principais construtores, segundo Rosenfeld, Morville e Arango (2015), são: verificadores ortográficos, que recuperam documentos por proximidade, mesmo que o usuário tenha digitado a palavra errada (ex. sapati e sapato); ferramentas fonéticas, que recuperam documentos com a mesma percepção de som da fala (ex. Smith, Smyth); ferramentas resultantes ou decorrentes, que recuperam documentos que contenham termos variantes do mesmo radical (ex. pedra, pedreiro, pedregulho); ferramentas de processamento de linguagem natural, que examinam a natureza sintática da consulta (ex. “como” pergunta ou “quem é” questão?); vocabulários controlados e tesouro, que

avaliam a natureza semântica de uma consulta ao incluírem automaticamente sinônimos (ex. sistemas de busca, caixa de pesquisa).

Realizada a busca, é importante que o sistema permita que o usuário classifique os resultados obtidos conforme lhe for mais conveniente. Rosenfeld, Morville e Arango (2015) apresentam algumas possibilidades de classificação: classificação alfabética, que é familiar para a maioria dos usuários; classificação cronológica, que é muito útil aos pesquisadores, pois permite a seleção dos resultados mais recentes; classificação por relevância, que se baseia em quantas vezes o termo consultado ocorre no documento e na popularidade do documento onde o termo aparece, entre outros fatores; ranking de popularidade, ou o número de links externos que apontam para o documento; ranking por classificação dos pesquisadores, ou a classificação por número de visualizações, visitas, comentários, compartilhamentos, *likes* etc.; e ranking pago, espaço comprado.

Metodologia

Neste artigo, importa verificar como funcionam os atuais sistemas de buscas revistas científicas eletrônicas. O método escolhido foi a inspeção por lista de verificação, realizada a partir de sentenças teste descritas com base nos autores de referência na área Rosenfeld, Morville e Arango (2015), Kalbach (2009) e Garrett (2011). Foram selecionadas três revistas de relevância para a ciência e com elevados fatores de impacto na área multidisciplinar, que são as seguintes: *Nature*, *Science* e *PLOS Biology*. As duas primeiras são de acesso restrito, por meio de pagamento, enquanto a última é de acesso aberto. A coleta de dados ocorreu entre os meses de agosto e dezembro do ano de 2015.

Resultados

As revistas foram avaliadas quanto a suas caixas de busca, seus mecanismos de pesquisa e os resultados de pesquisa gerados. Nas revistas *Nature*, *Science* e *PLOS Biology*, as caixas de busca aparecem de forma integrada à navegação principal, próximas às demais opções de navegação. Esse posicionamento é adequado e o mais apropriado segundo Rosenfeld, Morville e Arango (2015). Entretanto, esses autores também recomendam que se considere, além da

localização, a maneira como a caixa é apresentada. Nesse sentido, a caixa de pesquisa da revista *Science* não está bem alocada, pois aparece cercada de grande quantidade de informação, o que pode prejudicar a visualização do usuário. A Figura 2 mostra que ao lado da caixa de busca há um menu *drop-down* com outras funcionalidades. O ideal é que a caixa de busca apareça em uma área isolada, o que lhe confere o destaque necessário. Como ponto positivo, porém, a caixa de busca aparece de forma consistente ao longo do site dessa revista.



Figura 2. Posicionamento de caixa de busca na revista *Science*

Fonte: *Science* (2015)

A revista *Nature*, por sua vez, apresenta um problema na caixa de busca relacionado. O rótulo "Go, ou "Ir, definido para o botão posicionado ao lado da caixa busca nessa revista não é satisfatoriamente expressivo. O indicado seria utilizar o termo "Search", como nas duas outras revistas. A Figura 3 mostra o cabeçalho da revista *Nature* com destaque para a área de busca.



Figura 3. Posicionamento de caixa de busca na revista *Nature*

Fonte: *Nature* (2015)

Na revista *PLOS Biology* (Figura 4), os rótulos textual e iconográfico estão aplicados adequadamente. A palavra "Search" e o ícone da lupa são internacionalmente reconhecidos para essa função. O local de

posicionamento da caixa é adequado, aproveitando a experiência do usuário em outros sites, e a área é limpa, conduzindo o olhar para o campo de busca. A caixa de busca também aparece de forma consistente ao longo do site. A comparação entre as três revistas permite verificar que a revista *PLOS Biology* apresenta a caixa de busca de forma mais adequada.

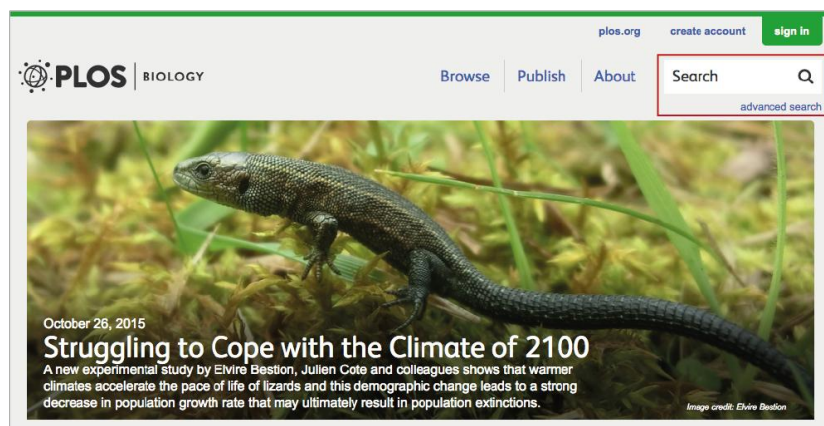


Figura 4. Posicionamento de caixa de busca na revista *PLOS Biology*

Fonte: *PLOS Biology* (2015)

É preciso ainda citar que, em relação à consistência, a revista Nature se destaca negativamente se comparada às revistas PLOS Biology e Science. Isso porque quando o usuário faz uma consulta o resultado da busca é apresentado em uma página que não mantém consistência visual com as demais páginas da revista.

Em relação aos mecanismos de pesquisa, considera-se as questões técnicas sobre como é realizada a busca por informações. É desejável que o sistema seja eficiente em buscas de nível de granularidade grossa, trazendo resultados sobre zonas de pesquisa e em buscas de granularidade fina, que tratam da pesquisa em componentes do conteúdo do documento. Essa inspeção também verifica se o site utiliza construtores de busca, como verificadores ortográficos, fonéticos, resultantes, processamento natural, vocabulário controlado; e apresenta recursos para melhorar o desempenho de busca, como o autocomplete, autossugestão e alerta de busca. Esses termos foram definidos a partir do estudo de Rosenfeld, Morville e Arango (2015) e de Kalbach (2009).

No que diz respeito à busca por zonas de pesquisa, ou granularidade grossa, as revistas Nature, Science e PLOS Biology apresentam boa eficiência se avaliarmos os resultados apresentados a partir de solicitações da caixa de

pesquisa simples. Isso fica visível pelo número de possibilidades de ajuste fino que as revistas apresentam junto aos resultados de busca, indicando que diversas zonas de pesquisa foram consultadas. A Figura 5 mostra como exemplo as zonas de pesquisa na página dos resultados de busca da revista *Science*.

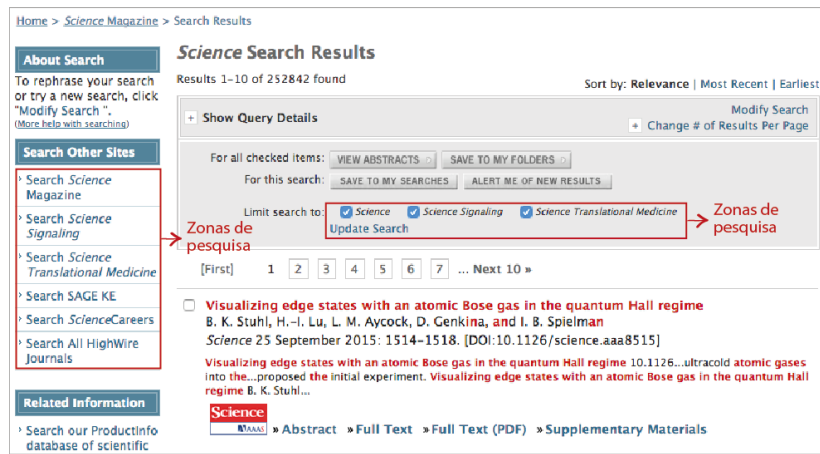


Figura 5. Zonas de pesquisa nos resultados de busca da revista *Science*

Fonte: *Science* (2015)

Todas as três revistas apresentam opção de pesquisas avançadas, porém as revistas *Science* e *PLOS Biology* tornam possível direcionar a pesquisa seja para granularidade grossa, seja para granularidade fina, por meio de diversos campos de configuração. Isso não acontece na pesquisa avançada da revista *Nature*, que apresenta uma página de busca avançada com número de opções reduzido, como mostra a Figura 6.

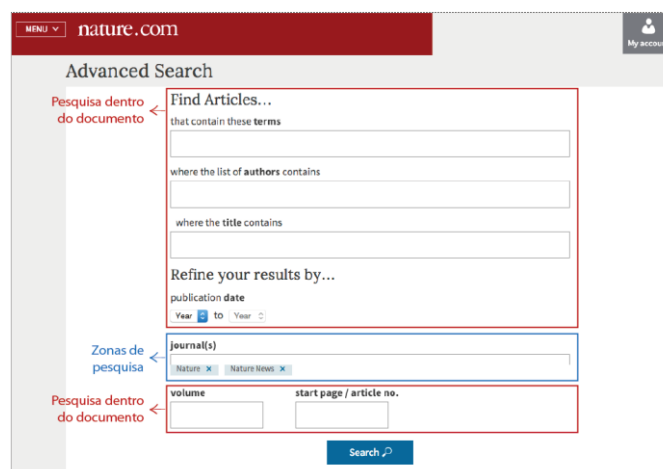


Figura 6. Busca avançada na revista *Nature*

Fonte: *Nature* (2015)

As revistas analisadas também apresentaram alguns recursos para melhorar os resultados das buscas. De acordo com Rosenfeld, Morville e Arango (2015), os sistemas de busca dos sites em geral, normalmente, são desenvolvidos com foco nos usuários inexperientes, ou que não estão solícitos a investir tempo aprendendo a utilizar a interface de pesquisa. Por isso, a regra de ouro é que se mantenha a interface de busca o mais simples possível. No entanto, para ajudar a aprimorar a busca, mesmo para esses usuários, passaram a ser incorporados à busca simples, alguns recursos que outrora eram disponíveis apenas a usuários avançados, como “auto complete” e o “auto sugestão”.

A inspeção permitiu verificar que as revistas *Nature*, *Science* e *PLOS Biology* armazenam antigas solicitações de busca e as apresentam no momento da digitação como sugestões para novas consultas, como exemplifica o destaque da Figura 7, retirada da revista *PLOS Biology*.



Figura 7. Sugestões do sistema de busca da revista *PLOS Biology*

Fonte: *PLOS Biology* (2015)

As três revistas, no entanto, não disponibilizam “auto complete” e “auto sugestão”. Atualmente, esses recursos são amplamente utilizados em sistemas de busca, pois representam ferramentas úteis para ajudar os usuários a identificarem potenciais conteúdos a partir de informações parciais ou incompletas digitadas no campo de busca. Conforme Rosenfeld, Morville e Arango (2015), estes recursos, em alguns casos, fornecem dicas sobre a forma como o sistema está organizado, permitindo assim sua exploração diretamente pela caixa de pesquisa, o que possibilita realização de buscas mais inteligentes. A Figura 8 mostra um exemplo desses recursos na busca do Google.

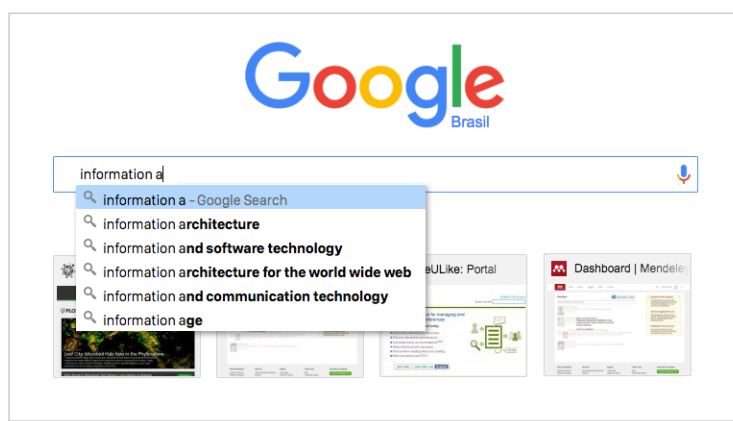


Figura 8. Auto complete e auto sugestão do Google

Fonte: Google (2015)

Em verificação quanto ao uso de construtores, percebe-se que o sistema de busca das revistas analisadas não aplica verificadores ortográficos, fonéticos, decorrentes, processamento natural ou vocabulário controlado. Aparentemente, os algoritmos empregam procedimentos necessários para a realização de uma tarefa de recuperação de correspondência, ou seja, o sistema de busca compara a consulta do usuário com um índice de textos, contidos no sistema, procurando a mesma sequência de texto. Quando uma *string* (conjunto de caracteres) correspondente for encontrada, o documento de origem é adicionado ao conjunto de resultados recuperados. Assim, no caso de uma palavra ser digitada incorretamente, o processo é realizado igualmente, trazendo resultado negativo.

Rosenfeld, Morville e Arango (2015) observam que recuperar conjunto de resultados de grandes dimensões é desmotivador para o usuário e recomendam o fornecimento de instruções de como restringir os resultados de busca. Por outro lado, eles consideram a pesquisa sem resultado algo frustrante para os usuários e recomendam a adoção políticas para solução do problema com: fornecer outra opção, mesmo que eles tenham recuperado zero resultados ou apresentar dicas ou conselhos que possibilitem melhorar sua pesquisa. Verificou-se que as três revistas inspecionadas adotam as medidas sugeridas.

Mesmo que o usuário padrão seja impaciente e inexperiente e que não reconheça a complexidade e a capacidade de um sistema de busca, é comum haver usuários experientes e altamente capacitados no que diz respeito à elaboração de buscas. Nielsen (2001) traz o princípio de flexibilidade e eficiência de uso afirmando que o sistema deve oferecer aceleradores que sejam invisíveis para os usuários inexperientes, mas que permitam aos

experientes realizarem tarefas com mais rapidez. Assim, boas interfaces de pesquisa permitem que os usuários experientes configurem suas buscas de forma avançada, através do uso de operadores booleanos, que podem ser combinados de diferentes maneiras no processo de elaboração de pesquisas. As três revistas analisadas efetuam buscas avançadas a partir de operadores booleanos.

A inspeção verificou se as revistas *Nature*, *Science* e *PLOS Biology* exibem os resultados de busca de forma organizada, se os separam hierarquicamente, se possibilitam classificação por critérios como alfabético, cronológico, relevância e popularidade e se permitem o refinamento por meio de filtros de ajustes. Existem muitas maneiras de exibir resultados de busca. Rosenfeld, Morville e Arango (2015), argumentam as questões a considerar nesse caso, são duas: quais componentes do documento recuperado convém apresentar?; e como listar os resultados de uma busca?

A resposta da primeira questão não é conclusiva, ela está relacionada ao tipo de conteúdo do documento (um texto, imagem, vídeos etc.) e ao público. No contexto dos periódicos científicos os arquivos são predominantemente textuais, mesmo que a grande maioria contenha imagens (figuras, tabelas e quadros). Em relação ao público, a grande maioria é constituída por usuários do âmbito acadêmico – alunos do ensino superior (graduação e pós-graduação), professores, pesquisadores, membros da comunidade científica em geral.

No que diz respeito ao público e às suas necessidades como usuário, a recomendação de Rosenfeld, Morville e Arango (2015) é que se ofereça um número menor de informações para usuários que sabem o que estão procurando e maior para usuários que não têm certeza do que querem. No caso de usuários de campos científicos, esses autores observam que há maior interesse em resultados de pesquisa com alta taxa de recuperação de um sistema de busca, do que em altas taxas de precisão. Esse indicador pode estar ligado à atividade do pesquisador, que busca em um primeiro momento abarcar todo conhecimento registrado em relação a um determinado assunto.

Já a segunda questão, em relação a organização e apresentação dos resultados de busca, Rosenfeld, Morville e Arango (2015) salientam que uma boa organização e hierarquia ajudam o usuário a visualizar rapidamente a página em busca de partes importantes de cada resultado. Na avaliação, de maneira geral, as três revistas analisadas apresentam organização satisfatória e hierarquia bem definida na apresentação dos resultados do sistema de busca. Em todas as três revistas inspecionadas é possível identificar o número

total de resultados encontrados, mas na revista Nature não é possível configurar o número de resultados que serão exibidos por página.

Conclusões

Quando se fala em sistema de busca, o pensamento é facilmente capturado pela imagem da caixa de busca, posicionada no canto superior direito da página digital. No entanto, a pesquisa em um site exibido na internet envolve mais do que isso. Neste artigo, tratou-se dos mecanismos e resultados de pesquisa, mas poderíamos considerar que a busca envolve todo o contato do usuário com o site.

Ao acessar um site, o usuário está à procura de informações, seu olhar percorre a página, parando nos pontos que lhe interessam. Ele navega no hipertexto e acessa os links que lhe chamam atenção. Nessas ações, estão presentes os quatro sistemas da Arquitetura da Informação, todos servindo ao exercício da pesquisa. Os sistemas de navegação, rotulação, organização e busca fazem parte de um todo. Assim, essa complexa arquitetura funciona unida para atender aos objetivos do usuário.

Portanto, os componentes reconhecidos como sistema de busca, nada mais são do que complementos para o aprimoramento da experiência do usuário. Nesse sentido, trabalham conjuntamente para lhe oferecer informação e oportunidades de interação. Assim, esse sistema auxilia na recuperação da informação e confere credibilidade ao site, fatores essenciais para uma revista científica.

Percebeu-se, nesta pesquisa que as revistas de alta visibilidade, sejam abertas, sejam pagas, seguem um padrão de apresentação similar. Isso pode ocorrer porque ambas contam com recursos financeiros e humanos que as possibilitam implementar soluções adequadas. Quando isso não ocorre, como no caso das revistas menores de acesso aberto, é preciso otimizar os recursos disponíveis. Nesse sentido, os resultados desta pesquisa podem contribuir como base para melhorias nas revistas científicas, que beneficiem a experiência de seus usuários.

Referências

- AGNER, LUIZ (2009). *Ergodesign e arquitetura de informação: trabalhando com o usuário*. 2. ed. Rio de Janeiro, RJ: Quartet.
- GARRETT, J. J. (2011). *The elements of user experience: user centered design for the Web and beyond*. 2nd. ed. Berkeley: New Riders.
- GOOGLE (2015). [Web page]. Busca. Disponível em: <<http://www.google.com.br>>
- KALBACH, J. (2009). *Design de navegação web*. Porto Alegre: Bookman.
- MORVILLE, P.; ROSENFELD, L. (2006). *Information Architecture for the World Wide Web*. 3rd. d. Sebastopol: O'Reilly.
- NATURE (2015). [Web page]. Disponível em: <<http://www.nature.com>>
- PLOS BIOLOGY (2015). [Web page]. Disponível em: <<http://journals.plos.org/plosbiology>>
- ROSENFELD, L.; MORVILLE, P.; ARANGO, J. (2015). *Information Architecture: for the web and beyond*. North, Sebastopol: O'Reilly.
- SCIENCE (2015). [Web page]. Disponível em: <<http://www.sciencemag.org>>