ARTICLE

RDBCI: Revista Digital Biblioteconomia e Ciência da Informação
RDBCI : Digital Journal of Library and Information Science

DOI 10.20396/rdbci.v16i2.8651090

ANALYSIS OF SEARCH SYSTEMS FOR ELECTRONIC JOURNALS*

ANÁLISE DOS SISTEMAS DE BUSCA DE REVISTAS CIENTÍFICAS ELETRÔNICAS

ANÁLISIS DE LOS SISTEMAS DE BÚSQUEDA DE REVISTAS CIENTÍFICAS ELECTRÓNICAS

[1]Paula Caroline Schifino Jardim Passos
[1]Sônia Elisa Caregnato
Universidade Federal do Rio Grande do Sul[1]

Correspondence to Author

[1]Paula Caroline Schifino Jardim Passos
Universidade Federal do Rio Grande do Sul
Porto Alegre, RS - Brazil
Email: paulacarolinejardim@gmail.com
ORCID: http://orcid.org/0000-0002-2354-847X

Submitted: 19/02/2018
Accepted: 15/03/2018
Published: 03/05/2018

JITA: HN. e-journals.

*Text modified and approved in BIREDIAL (2017) for the Dossier form to publication in the RDBCI.

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 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 em autores de referência na área como 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, consequentemente, promove a visibilidade dos textos publicados.

PALAVRAS-CHAVE: Sistema de busca. Revista científica eletrônica. Comunicação científica. Acesso aberto.

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 reference authors in the area such as 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 journals. 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. Open access.

RESUMEN: Los motores de búsqueda de revistas científicas electrónicas consisten en aplicaciones de software con campos que el usuario puede llenar para solicitar la recuperación de algún tema. La presente investigación analiza la búsqueda en tres revistas de renombre internacional, Nature, Science y PLOS Biology. Para ello, presentamos una breve reseña bibliográfica del sistema de búsqueda y sus componentes. Los resultados del análisis se muestran a continuación. El método elegido fue la inspección por lista de verificación, llevada a cabo utilizando oraciones de prueba descritas para esta investigación basadas en autores de referencia en el área como Rosenfeld, Morville y Arango (2015), Kalbach (2009) y Garrett (2011). La verificación sistemática y en profundidad nos permitió enumerar los recursos aplicados a la búsqueda de revistas científicas de alta visibilidad. Por otro lado, fue posible identificar puntos donde algunas modificaciones aportarían beneficios a la experiencia de los usuarios de las revistas. Además, a través de la comparación entre las tres revistas, se evidenciaron las mejores soluciones utilizadas en cada uno de los motores de búsqueda. Considerando la importancia de las revistas para la comunicación científica, se entiende como estudios relevantes orientados a la actualización constante de sus sistemas. El análisis de la búsqueda contribuye a la mejora de estos vehículos de información y, en consecuencia, promueve la visibilidad de los textos publicados.

PALABRAS CLAVE: Sistema de búsqueda. Revista científica electrónica. Comunicación científica. Acceso abierto.

1 INTRODUCTION

This paper presents a study of the search engines of Nature, Science and PLOS Biology journals. The analysis of the search system of internationally renowned electronic journals aims to propose solutions that can be applied to the improvement of other journals. The specific objectives of the investigation are to observe the characteristics of the interface, of the search mechanisms and of the presentation of the results in each of the journals. The study also aims to compare aspects of searching in paid subscription journals with their corresponding open access journals.

In relation to scientific literature, the Budapest Open Access Initiative (BOAI15, 2017) defines open access as the free availability of full texts of articles on the Internet, allowing reading, copying and distribution without financial, legal or technical cost. The BOA document states that "The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited".

The search is the component of Information Architecture that allows user interact with the system and retrieve content. It is important to note that a website is a complex structure, replete with interconnected and interdependent systems. Although this study treats the search system separately from other systems (navigation, labeling and organization), it is understood that they function in an integrated way to construct the search results. Morville and Rosenfeld (2006) point out that a link on a page can be part of several systems simultaneously. Thus, the separation into different types of systems is only didactic and operational.

2 SEARCH SYSTEMS

Search Systems consist of software applications in fields that allow users to query the system to retrieve content about any subject. They allow the use of natural language and Boolean operators. User requests are cross-referenced with an index consisting of all terms found in documents or a list of titles, authors, categories, and related information. Metadata allows to identify stored documents (AGNER, 2009).

According to Rosenfeld, Morville and Arango (2015), even though the specialized literature presents many studies related to search system, the large number of variables involved (level of knowledge and motivation of the user, types and amount of information searched, etc.) hinders the development of an ideal search interface. However, from Rosenfeld, Morville and Arango (2015) and Kalbach (2009) it is possible to state that a search system must be efficient in both coarse level of granularity (research areas, such as content type, audience, subject, timeline, etc.), and in a finer-grained level (content components within the document itself, such as article title, author, affiliation, keywords, etc.).

Users do not always know how to express what they are seeking, but good search interfaces can help them find what they're looking for. To Rosenfeld, Morville and Arango (2015), the presentation of search results also sets up as a possibility to narrow the search. In this sense, and still aiming to qualify search requests from inexperienced users, authors recommend the use of query constructors, that is, tools that can improve the performance of a search.

The main constructors, according to Rosenfeld, Morville and Arango (2015), are: **spell checkers**, which retrieve documents by proximity, even if the user has typed the wrong word (e.g. sheo and shoe); **phonetic tools** that retrieve documents with the perception of speech sound (e.g. knees, niece); **resulting or derived tools** that retrieve documents containing variant terms of the same radical (e.g. therapeutic, therapist, therapy); **natural language processing tools**, examining the syntactic nature of the query (e.g. "how" question or "who is" questioning?); **controlled vocabularies and thesaurus**, which evaluate the semantic nature of a query by automatically including synonyms (e.g. search systems, search box).

There are also features to improve search performance such as autocomplete, autosuggestion and search alert. According to Rosenfeld, Morville, and Arango (2015), site search systems generally are developed with a focus on inexperienced users, or who are not willing to spend time learning how to use the search interface. Thus, the rule of thumb is to keep the search interface as simple as possible. However, to help enhance the search, even for inexperienced users, some features that were once available only to advanced users, such as autocomplete and autosuggestion, were incorporated into the simple search.

Even if the standard user is impatient and inexperienced and does not recognize the complexity and capacity of a search engine, it is common to have experienced and highly skilled search users. Nielsen (2001) brings the principle of flexibility and efficiency of use, stating that the system should offer accelerators that are invisible to inexperienced users but that allow experienced users to perform tasks faster. Thus, good search interfaces allow experienced users to configure their searches in an advanced way, using Boolean operators, which can be combined in different ways in the search process.

Completed the search, it is important that the system allows the user to sort results as it is fits. Rosenfeld, Morville e Arango (2015) present some possibilities of classification: **alphabetical classification**, which is familiar to most users; **chronological classification**, which is very useful for selection of the most recent results; **classification by relevance**, which is based on how many times the term consulted occurs in the document and on the popularity of the document where the term appears, among other factors; **popularity ranking,** or the number of external links pointing to the document; **researchers ranking,** or ranking by number of views, visits, comments, shares, likes etc.; and **paid ranking**, which is based on space purchased.

DOI 10.20396/rdbci.v16i2.8651090

Figure 1 represents the search system described by Morville and Rosenfeld (2006). The diagram demonstrates the search path from the user's query to the presentation of the results in the system interface. The intermediate steps take into account the interface where the search is performed, the mechanisms that aid in the search and the content, which involves metadata and controlled vocabulary. The search results are presented according to the interface design and using a hierarchy algorithm.

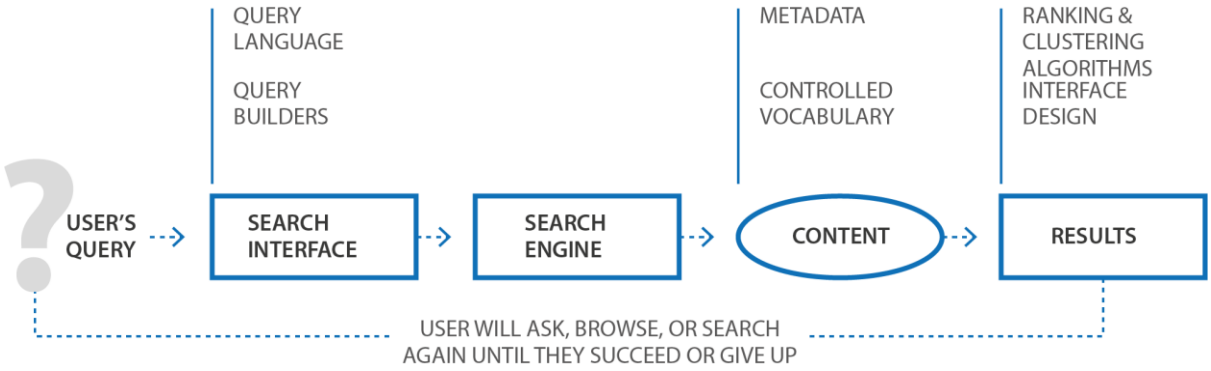

Figure 1. Search System

Source: Morville and Rosenfeld (2006, p. 14).

The evaluation of the search system reported in this paper employs three distinct but interrelated variables. The first one evaluates the search interface, considering the presence of the search box, as well as its positioning and consistency. The next variable refers to the search engine, which involve the technical questions about how the search for information is performed. Finally, the third evaluation variable is about the search results, related to the presentation of the retrieved information to the user.

3 METHODOLOGY

This investigation verifies how current search engines of electronic journals work. The method chosen was the checklist inspection, performed from the test sentences described by Rosenfeld, Morville and Arango (2015), Kalbach (2009) and Garrett (2011), Downey and Banerjee (2011) and Pressman 2011), experts on Information Architecture and interface development (Table 1).

Variable	Condition	Sources
4.1 Search box	• Are the box and the search button consistently integrated to the main navigation area across all pages?	Kalbach (2009); Rosenfeld, Morville e Arango (2015)
4.2 Search engines	• Is the system efficient for searches at fine and coarse granularity levels and uses constructors and features that can improve the performance of a search?	Kalbach (2009); Rosenfeld, Morville e Arango (2015); Downey e Banerjee (2011)
4.3 Results presentation	• In addition to being displayed as a hierarchy and organized, the search results can be sorted by several criteria (alphabetical, chronological, relevance, popularity, etc.) and refined using filters and adjustments?	Kalbach (2009); Rosenfeld, Morville e Arango (2015); Pressman (2011)

Tabela 1. Sentences for evaluation of the Search System of journals

Source: Elaborated by the authors based on the research data

Three journals of relevance to science and with high impact factors in the multidisciplinary area were selected, which are the following: Nature, Science and PLOS Biology. The first two are restricted by means of payment, while the latter is open access. Data collection occurred between August and December of 2015.

4 RESULTS

The journals were evaluated according to search boxes, search engines and results generated. Nature, Science and PLOS Biology have search boxes integrated to the main navigation, near to the other navigation options. This position is most appropriate according to Rosenfeld, Morville and Arango (2015). However, the authors also recommend considering, in addition to location, the way the box is presented. In this sense, the search box of the journal Science is not well placed, because it appears surrounded by a great amount of information, which can impair the visualization. Figure 2 shows that next to the search box there is a drop-down menu with other features. Ideally, the search box should appear in an isolated area, which gives it the necessary highlighting. On the positive side, however, the search box appears consistently throughout the journal's website.

 DOI 10.20396/rdbci.v16i2.8651090

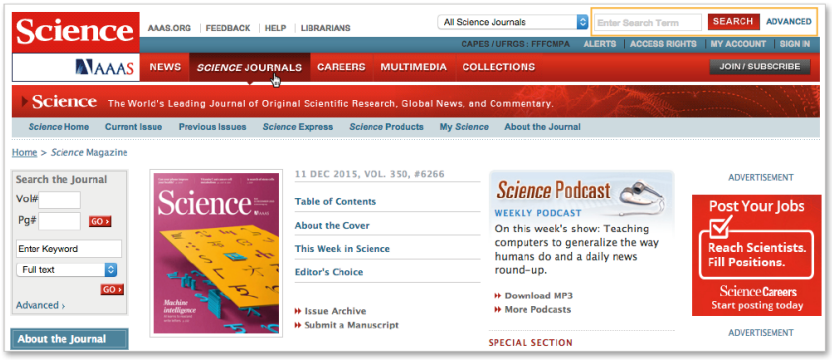

Figure 2. Search box positioning in the journal Science

Source: Science (2015).

The Nature journal, in turn, presents a problem in its search box. The "Go" label, next to the search box, is not satisfactorily meaningful. The suggestion is to use the term "Search", as the two other journals do. Figure 3 shows the heading of Nature highlighting the search area.

Figure 3. Search box positioning in the journal Nature

Source: Nature (2015).

In PLOS Biology (Figure 4), textual and iconographic labels are applied appropriately. The word "Search" and the magnifying glass icon are internationally recognized for this function. The positioning of the box is appropriate, taking advantage of the user experience on other sites, and the area is clean, leading the eyes to the search field. The search box also appears consistently throughout the site. The comparison between the three journals shows that PLOS Biology presents the search box in a more appropriate way.

DOI 10.20396/rdbci.v16i2.8651090

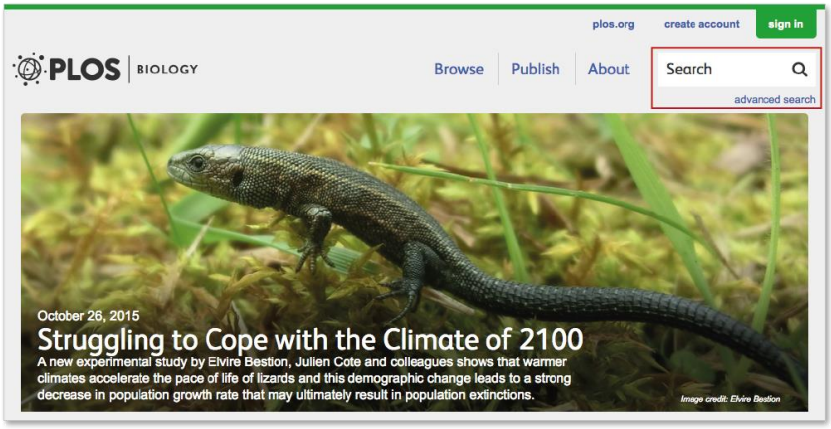

Figure 4. Position of the search box in the journal PLOS Biology

Source: PLOS Biology (2015).

It should also be mentioned that, in relation to consistency, the Nature journal stands out negatively compared to PLOS Biology and Science. When the user makes an inquiry the search result is displayed on a page that does not maintain visual consistency with the other pages of the journal.

Regarding the search mechanisms, technical questions about how the search for information is carried out is considered. It is desirable that the system be efficient in coarse granularity level searching, bringing results on search areas, and fine-grained searches, which deal with searching in components of document content. This inspection also checks whether the site uses search builders, such as spelling checkers, phonetic, natural language processing, controlled vocabulary; and provides resources to improve search performance, such as autocomplete, autosuggestion and search alert. These terms were gathered from the studies of Rosenfeld, Morville and Arango (2015) and Kalbach (2009).

When searched for areas of research, that is, applying coarse granularity, Nature, Science and PLOS Biology present good efficiency if we evaluate the results presented from requests made in the simple search box. This is confirmed by the presence of fine-tuning possibilities that journals display along the search results, indicating that several search areas have been queried. Figure 5 shows the search zones on the search results page of the journal Science.

DOI 10.20396/rdbci.v16i2.8651090

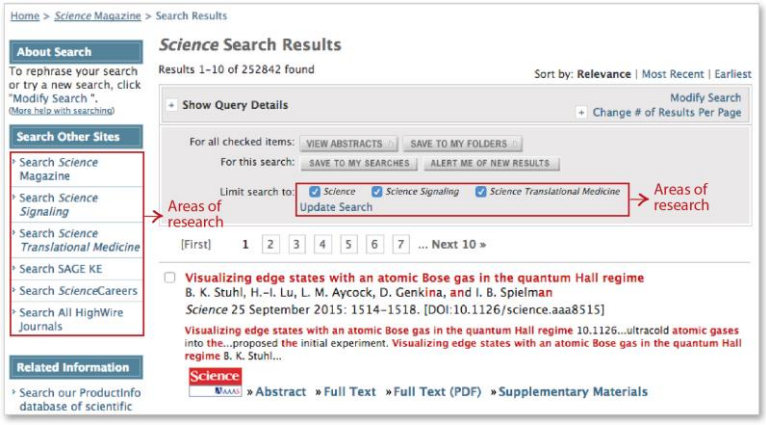

Figure 5. Research areas in the search results page of Science journal

Source: Prepared by the authors from Science (2015).

All three journals have an advanced search option, but Science and PLOS Biology make it possible to direct the search either for course granularity or for fine granularity through various configuration fields. This does not happen in Nature's advanced search, which features an advanced search page with a reduced number of options, as shown in Figure 6.

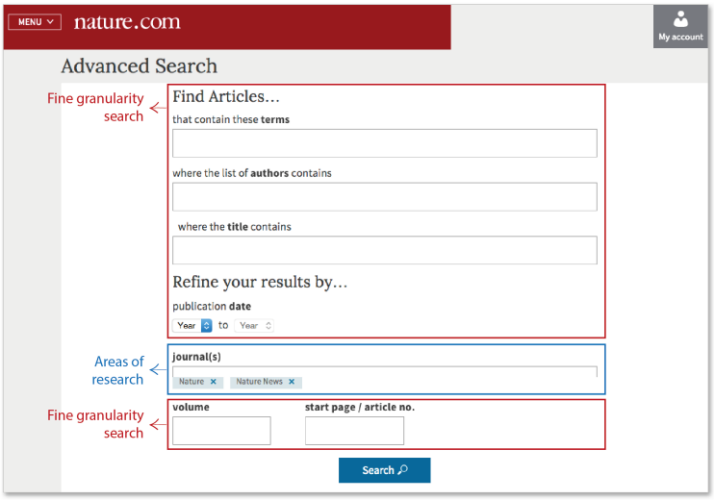

Figure 6. Advanced Search page of Nature journal

Source: Prepared by the authors from Nature (2015).

The journals analyzed also presented some resources to improve search results. The inspection has shown that Nature, Science and PLOS Biology store old search requests and suggest them when the user is typing new queries, as exemplified in Figure 7, a page taken from PLOS Biology.

DOI 10.20396/rdbci.v16i2.8651090

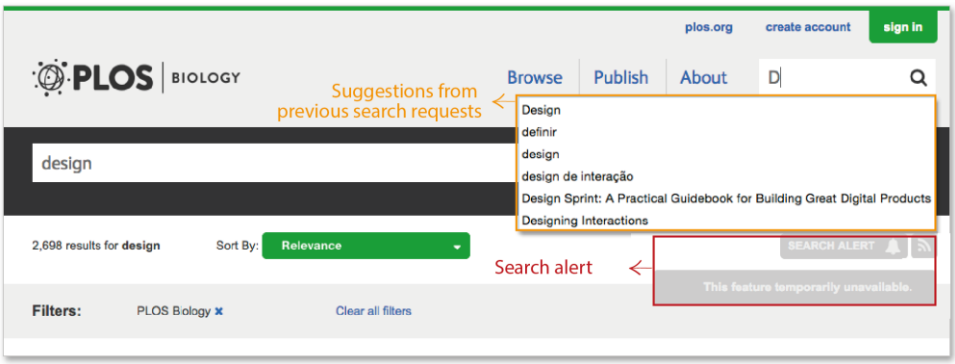

Figure 7. Suggestions from the search system of PLOS Biology journal

Source: Prepared by the authors from PLOS Biology (2015).

The three journals, however, do not provide autocomplete and autosuggestion features. Currently, these are widely used in search engines because they represent useful tools to help users identify potential content from partial or incomplete information typed in the search field. According to Rosenfeld, Morville and Arango (2015), these settings, in some cases, provide tips on how the system is organized, thus allowing its exploration directly from the search box, which allows for smarter searches. Figure 8 shows an example of these features in Google's search.

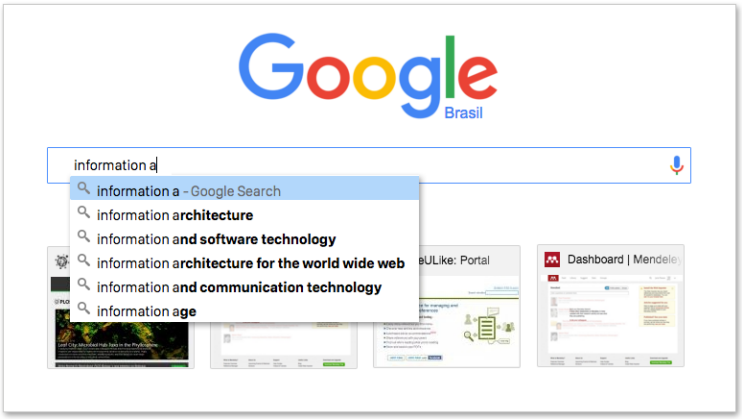

Figure 8. Google self-completion and self-suggestion features

Source: Google (2015).

During the verification of the use of constructors, it was perceived that the search systems of the analyzed journals do not apply spelling and phonetic verification, derivation, natural language processing or controlled vocabulary. Apparently, the algorithms employ procedures necessary to perform a matching retrieval task, that is, the search system compares the user's query with an index of texts, contained in the system, searching for the same text string. When a matching string is found, the source document is added to the retrieved result set. Thus, in case a word is typed incorrectly, the process is performed anyway, bringing negative results.

Rosenfeld, Morville and Arango (2015) note that retrieving large result sets is annoying for the user, thus they recommend providing instructions on how to restrict search results. On the other hand, they consider a search without results frustrating for the users and they recommend the adoption of policies for the solution of the problem, such as to provide another option, even if they have recovered zero results or to present tips or advice that enable to improve the search. The journals analyzed adopt the suggested measures and also allow advanced searches from Boolean operators.

The inspection tested if Nature, Science and PLOS Biology display search results in an organized way, if they can be separated hierarchically, if they are allowed to be classified by criteria such as alphabetical, chronological, relevance and popularity and if they allow refinement by means of adjustment filters. There are many ways to display search results. Rosenfeld, Morville, and Arango (2015) argue that there are two issues to be considered in this case: what components of the retrieved document should be presented and how to list the results of a search?

The answer to the first question is not conclusive, since it is related to the content type of the document (a text, image, videos etc.) and the public. In the context of scientific journals, the archives are predominantly textual, even though the majority contains images (figures, tables and pictures). As for the public, most of them are academic users - undergraduate and postgraduate students, professors, researchers, members of the scientific community in general.

Rosenfeld, Morville and Arango (2015) recommend that small amounts of information should be offered to users who know what they are looking for and greater amounts to users who are not sure about what they want. In the case of scientists, the authors note that there is a greater interest in search results with a high rate of retrieval items, than with high rates of precision. This indicator may be linked to the researcher's activity, which seeks, at first, to encompass all registered knowledge in relation to a given subject.

In the second question, regarding the organization and presentation of the search results, Rosenfeld, Morville and Arango (2015) point out that good organization and hierarchy help the user to quickly visualize the page in search for important parts of each result. Evaluation has shown that, in general, the three journals analyzed present satisfactory organization and well-defined hierarchy in the presentation of the results of the search system. In all three journals inspected it is possible to identify the total number of results retrieved, but in Nature it is not possible to set up the number of results that will be displayed per page.

4 CONCLUSION

Search systems are easily mistaken by the search box image, positioned in the upper right corner of the digital page. However, searching on a website involves more than that. In this article, we dealt with the mechanisms and results of a query, but we could consider that the search involves all the user's contact with the site.

When accessing a site, the user is looking for information, and his look scrolls the page, stopping at the points that interest him. He browses the hypertext and accesses the links that call his attention. During these actions, the four Information Architecture systems are present, all serving his purpose. Navigation, labeling, organization and search systems are part of a whole. Thus, this complex architecture works together to meet user goals.

Therefore, the components recognized as a search system are nothing more than complements to the improvement of the user experience. In this sense, they work together to offer the user information and opportunities for interaction. Thus, this system assists in the retrieval of information and lends credibility to the site, essential factors for a scientific journal.

It was observed in this investigation that highly visible journals, whether open access or paid, follow a similar pattern of presentation, probably because both have the financial and human resources that enable them to implement adequate solutions. When this is not the case, as with smaller open access journals, it is necessary to optimize the available resources. In this sense, the results of this investigation can contribute for improvements in scientific journals, that improve the experience of its users.

REFERENCES

AGNER, L. **Ergodesign e arquitetura de informação**: trabalhando com o usuário. 2. ed. Rio de Janeiro, RJ: Quartet, 2009.

BOAI15. Budapest Open Access Initiative (BOAI). 2017. Disponível em: < http://www.budapestopenaccessinitiative.org/boai15-1>. Acesso em: 5 jan. 2018.

DOWNEY, L.; BANERJEE, S. Building an Information Architecture checklist. **Journal of Information Architecture**, v. 2, n. 2, 2011. Disponível em: <http://journalofia.org/volume2/issue2/03-downey/>. Acesso em: 5 mar. 2018.

GARRETT, J. J. **The elements of user experience**: user centered design for the Web and beyond. 2nd. ed. Berkeley: New Riders, 2011.

GOOGLE. **Busca.** 2015. Disponível em: <http://www.google.com.brr>. Acesso em: 5 jan. 2015.

KALBACH, J. **Design de navegação Web**. Porto Alegre: Bookman, 2009.

MORVILLE, P.; ROSENFELD, L. **Information Architecture for the World Wide Web**. 3rd. d. Sebastopol: O'Reilly, 2006.

NATURE. **[Web page].** 2015. Disponível em: < http://www.nature.com>. Acesso em: 5 dez. 2015.

NIELSEN, J. **10 Usability Heuristics for User Interface Design**. Nielsen Norman Group. 1995. Disponível em: <http://www.nngroup.com/articles/ten-usability-heuristics/>. Acesso em: 26 nov. 2015.

PLOS BIOLOGY. **[Web page].** 2015. Disponível em: <http://journals.plos.org/plosbiology>. Acesso em: 5 jan. 2015.

PRESSMAN, R. S. **Engenharia de Software**: uma abordagem profissional. Porto Alegre: AMGH Editora., 2011.

ROSENFELD, L.; MORVILLE, P.; ARANGO, J. **Information Architecture**: for the Web and beyond. North, Sebastopol: O'Reilly, 2015.

SCIENCE. **[Web page].** 2015. Disponível em: <http://www.sciencemag.org>. Acesso em: 5 dez. 2015.

ISSN 1678-765X
17 >
9 771678 765041