

Geração de indicadores para periódicos científicos abertos

Bringing out indicators for open scientific journals

Patrícia Zeni MARCHIORI¹  0000-0002-7238-9268

Eduardo Michelotti BETTONI¹  0000-0001-9560-3514

Marcelo Batista CARVALHO¹  0000-0002-1401-3174

Andre Luiz APPEL²  0000-0002-9608-803X

Resumo

Equipes editoriais envidam constantes esforços no cumprimento de requisitos qualitativos e quantitativos exigidos por agências indexadoras de periódicos científicos. Aqueles periódicos ainda não consolidados nas suas áreas de interesse – e consequentemente ainda não inseridos nas bases de dados que produzem indicadores – normalmente deixam de coletar, calcular e analisar dados que venham a auxiliá-los na tomada de decisão. Assim, questiona-se a possibilidade e as condições de geração de indicadores de desempenho, utilizando-se os dados presentes na plataforma *Open Journal Systems* – enquanto ambiente de editoração e publicação de periódicos científicos –, em conjunto com outras fontes de dados sob domínio ou alcance dos seus gestores. Nesse sentido, o objetivo deste estudo foi explorar uma metodologia para geração de tais indicadores, integrando a criação e uso de uma base de referências com a consulta aos critérios de avaliação de periódicos verificados na literatura e, ainda, com o processo de extração de metadados. Tal metodologia, aplicada ao periódico “AtoZ: novas práticas em informação e conhecimento”, resultou na exploração de um *corpus* de referências em BibTeX e na conversão dos metadados existentes no arco temporal de 2011 a 2016, permitindo o planejamento, criação e análise de indicadores que, potencialmente, podem auxiliar na visão estratégica do periódico. O conjunto de dados coletados e o teste dos indicadores propostos reforçam a necessidade de preenchimento integral e constante dos metadados, bem como as vantagens do uso do formato BibTeX, além de permitir a sequência da investigação, voltada ao entendimento, avaliação e comparação do periódico com outros títulos disponíveis.

Palavras-chave: Produção científica. Revistas eletrônicas. Técnicas bibliométricas.

Abstract

Editorial teams usually engage in constant efforts to meet the qualitative and quantitative requirements demanded by indexing agencies. Those journals which, due to the fact that they have not been consolidated in the areas of interest for publication – and consequent non-insertion in indexing databases – usually don't apply procedures to collect, compute and analyze data that could help managers in the decision making process. Under this assumption, a question has been raised concerning the feasibility (and conditions) in which the use of data present in the Open Journal Systems platform - as an environment for publishing scientific journals – along with other sources under the domain/scope of the journal managers could generate performance indicators. The objective of this study was, therefore, to explore a methodology for the generation of such indicators, which integrates the creation and use of an article reference base; evaluation criteria for scientific journals derived from literature; and a process of metadata extraction. This methodology, applied to nine issues of the journal “AtoZ: novas práticas em informação e conhecimento”, gave rise to the exploration

¹Universidade Federal do Paraná, Centro de Ciências Sociais Aplicadas, Departamento de Ciência e Gestão da Informação. Av. Prefeito Lothário Meissner, 632, Campus III, Jd. Botânico, 80210-170, Curitiba, PR, Brasil. Correspondência para/Correspondence to: P.Z. MARCHIORI. E-mail: <pzeni@ufpr.br>.

²Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Escola de Comunicação, Programa de Pós-Graduação em Ciência da Informação. Rio de Janeiro, RJ, Brasil.

Apoio: Pesquisa realizada com recursos do Edital de Apoio à Editoração e Publicação de Periódicos Científicos, 2016 (UFPR/PRPPG/SIBI).

Recebido em 25 de junho de 2017, reapresentado em 26 de fevereiro de 2018 e aprovado para publicação em 26 de março de 2018.

Como citar este artigo/How to cite this article

Marchiori, P.Z. et al. Geração de indicadores para periódicos científicos abertos. *Transinformação*, v.30, n.3, p.324-335, 2018. <http://dx.doi.org/10.1590/2318-08892018000300005>



of a corpus of references in BibTeX and existing metadata. The results helped to design, create and analyze ten indicators which can, potentially, assist the editorial team in reviewing the strategic vision of the journal. The gathered data, followed by controlled tests of the proposed indicators, reinforces the need for a constant filling of metadata, as well as, stress the benefits of using the BibTeX format. The research offered some basis for a following investigation, which aims for a comparison between the AtoZ journal performance and other concurrent titles available.

Keywords: Scientific Production. Electronic journals. Bibliometric Techniques.

Introdução

Os periódicos considerados “de referência”, ou seja, aqueles reconhecidos como de excelência pela comunidade científica, combinam três características interdependentes: visibilidade, qualidade e credibilidade – acompanhadas de uma quarta condição desejável, a pontualidade. A condição de “ser referência” baseia-se em um círculo virtuoso com a comunidade de usuários (autores e leitores), o qual “[...] dá sustentação ao periódico principalmente com as taxas crescentes de artigos recebidos para publicação, o que provoca também taxas crescentes de rejeição de artigos que, por último, contribui para melhorar e estabilizar a qualidade dos artigos aprovados para publicação” (Packer; Meneghini, 2006, p.241). A publicação em inglês ou em mais de um idioma é igualmente considerada estratégica para a visibilidade do periódico, e pode ser uma variável importante na sua internacionalização.

Tais periódicos de referência são normalmente indexados por agências/empresas produtoras de bases de dados que pretendem “oferecer uma cobertura abrangente da mais importante e influente pesquisa realizada em todo o mundo [...] o que não significa necessariamente incluir tudo” (Garfield, 1990, p.186). A lógica que sustenta a seleção de revistas consideradas relevantes para a indexação em tais agências se baseia no fato de que elas apresentam um padrão editorial consistente e agregam artigos influentes, ou seja, aqueles que tendem a receber mais citações.

O número de citações de um artigo é entendido, normalmente, como uma contagem simples de menções que ele recebeu de outros. As bases *Web of Science* (WoS) e *Scopus* são reconhecidas por coletarem, armazenarem e indexarem sistematicamente os principais metadados dos artigos publicados em periódicos previamente selecionados, e que possibilitam a geração de indicadores, tais como o Fator de Impacto (FI), o *Journal Citation Report* (JCR) da WoS, o *SCImago Journal Rank* (SJR) da Scopus, e o índice “h” da WoS e *Scopus*, entre outros.

Periódicos não incorporados por essas bases contam, como alternativa, com a ferramenta de busca *Google Scholar* (GS) – explorada por Bar-Ilan (2008) e Harzing e Van der Wal (2008) –, que indexa de forma automatizada documentos de caráter científico publicados na *web*, e oferece alguns indicadores, tais como o número de citações de documentos e o índice “h” de autores e de periódicos. Estudos citados por Winter, Zadpoor e Dodou (2014) compararam a WoS e o *Google Scholar*, esclarecendo as condições de cobertura de fontes/materiais e processos de indexação, concluindo que o escopo do *Google Scholar* é mais amplo para periódicos de acesso aberto, quando comparado com a WoS.

Além dessas opções, surge um questionamento: quais seriam as possibilidades e condições para que gestores responsáveis por periódicos em consolidação (e ainda não inseridos em bases de dados indexadoras) estruturam a coleta e manipulação de dados que venham a compor indicadores “domésticos” de desempenho? Por indicadores domésticos de desempenho entendem-se, para fins deste estudo, aqueles baseados nos recursos, dados e tecnologias sob alcance do próprio editor, e que podem ser calculados e analisados sem necessidade de agências externas.

Neste estudo, portanto, objetivou-se apresentar uma metodologia que integra a constituição de uma base própria de referências, à consulta a critérios de avaliação de periódicos disponíveis na literatura, e o uso dos metadados derivada do *corpus* de artigos de um periódico científico interdisciplinar³ de acesso aberto para a geração de tais indicadores. Estes são descritos na sequência do estudo, acompanhados de um recorte selecionado de resultados.

³ “AtoZ: novas práticas em informação e conhecimento”, disponível em: <<http://www.revistas.ufpr.br/atoz>>. Criada e gerenciada pelo Grupo de Pesquisa Metodologias em Gestão da Informação (UFPR/CNPq) e atualmente vinculada ao Programa de Pós-Graduação (Mestrado) em Ciência, Gestão e Tecnologia da Informação, da Universidade Federal do Paraná.

Procedimentos Metodológicos

Em 2014, o LaTeX e o BibTeX foram adotados como ambientes para a edição e diagramação de manuscritos no periódico “AtoZ: novas práticas em informação e conhecimento”, o que permitiu a geração automatizada dos diferentes formatos de publicação (PDF, HTML e ePub) e o gerenciamento de todas as referências dos materiais publicados desde o lançamento do periódico em 2011, dando origem à base BibTeX/AtoZ (em um arquivo com extensão .bib). Tal base foi estruturada de acordo com o manual do pacote *apacite* (Meijer, 2013), e contava – até o primeiro semestre de 2016 – com 1.129 registros. O formato das referências e o conjunto de metadados em cada registro BibTeX foram definidos com base no *American Psychological Association (APA) Style*, aliados a dois complementos, a saber: (1) uma sintaxe própria para a *citation-key* (chave primária) dos registros BibTeX, permitindo que em novos artigos pudessem ser reutilizados os registros existentes na base; (2) uma *tag* customizada (“citado por”) para manter o vínculo bidirecional entre artigo citante e referência citada.

O periódico conta também com um conjunto de metadados armazenados sistemática e exaustivamente ao longo de sua existência, por meio do *Open Journal Systems (OJS)*; e acessíveis, de forma aberta, pelo protocolo *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*, sendo este um mecanismo facilitador da interoperabilidade entre repositórios de objetos digitais.

Para o estudo em questão, foram avaliadas as variáveis existentes nos dois conjuntos de dados disponíveis no ambiente analisado, a base BibTeX/AtoZ e os metadados do periódico AtoZ.

De início, delimitaram-se como *corpus* os documentos citantes (e referências citadas) presentes nas seções “Artigos” e “Short Papers” do periódico AtoZ (doravante denominados “artigos”) desde o volume 1, número 1 (2011) até o volume 5, número 1 (2016), tendo sido desconsiderados os editoriais, os expedientes e as entrevistas. Considerando-se os metadados dos artigos do *corpus*, o Quadro 1 apresenta as opções discutidas pela equipe de pesquisadores.

Os metadados listados podem, inicialmente, constituir um grupo próprio de indicadores, dada sua condição de disponibilidade orgânica no OJS. Em lógica semelhante, é possível derivar alguns indicadores de uma base BibTeX sem apoio dos metadados. Contudo, a associação entre essas duas fontes de dados se apresenta como desafio mais complexo: cada conjunto de metadados pode ser diretamente associado a um grupo de referências e analisado, por meio da relação citante/citado, como nos seguintes exemplos:

Quadro 1. Variáveis potenciais para indicadores derivadas dos metadados – Periódico AtoZ (2011-2016).

Metadados do OJS	Condições de análise para a exploração de indicadores
Autoria	a) análise via autor principal.
Afiliação do(s) autor(es)	b) análise autor a autor: a unidade de análise é cada um dos autores, e as contagens podem
País de origem do(s) autor(es)	superar o total de artigos.
Título/Title	a) análise pela extensão e frequência – <i>tag cloud</i> . b) análise de conteúdo/ <i>text mining</i> .
Palavras-chave/Keyword/Palabras-clave	Análise como variável discriminante, desde que preenchidas e controladas.
Seção	Análise como variável discriminante, pois é dado controlado no OJS.
Resumo/Abstract/Resumen	a) análise pela extensão e frequência – <i>tag cloud</i> . b) análise de conteúdo/ <i>text mining</i> .
Ano de publicação	
Área(s) do Conhecimento	
Idioma	Análise individual como variável discriminante, desde que os dados sejam preenchidos e controlados.
Tipo, método ou ponto de vista	
Agências de fomento	

Nota: OJS: *Open Journal Systems*.

Fonte: Elaborado pelos autores, maio 2017.

1) SE qualquer artigo citante tem a propriedade “ano de publicação” e SE esse mesmo artigo tem um conjunto de referências na base BibTeX, ENTÃO, é possível dizer quais das referências da base BibTeX foram citadas num dado ano; ou,

2) SE qualquer artigo citante tem a propriedade “ano de publicação”, ENTÃO é possível criar um indicador por ano de publicação, ou da média, ou, ainda, da distribuição de anos dos artigos citados.

Visando facilitar a reprodução deste estudo, definiu-se como premissa que todo e qualquer dado a ser coletado não exigiria permissão especial de acesso (ex.: *login*, consulta ao banco de dados) no OJS. Optou-se, para tanto, pelo protocolo OAI-PMH, o qual atende aos critérios de acesso, qualidade e interoperabilidade. No caso de metadados não disponíveis por esse método, foram utilizados *robots*, que são conjuntos de instruções de *software* que executam um *web scraping*⁴ nas páginas *web* do periódico. Assim, as etapas para a formação da base de metadados são as que seguem:

1) extração OAI-PMH: Dados descarregados a partir do endereço <<http://revistas.ufpr.br/atoz/oai>> e convertidos em XML único com apoio da ferramenta OAI-PMH Validator (Banos, 2017);

2) importação e manipulação do XML: Dados importados para manipulação no software R - um ambiente de *software* livre para a manipulação estatística e geração de gráficos (The R Foundation, 2017) - especificamente com o package::xml2;

3) verificação de consistência OAI-PMH: Dados compilados no R; avaliação de consistência de variáveis e qualidade dos metadados. Verificou-se a necessidade de utilizar o *web scraping* para a obtenção dos dados faltantes;

4) *web scraping*: Variáveis nome, afiliação institucional e país de origem dos autores foram obtidas por *web scraping* na página de cada artigo, com uso do package::RCurl.

Finalizada a preparação e validação dos metadados, passou-se à análise e adaptação da base BibTeX/AtoZ, visando a composição do *dataset* completo para a geração dos indicadores. Ainda em formato .bib, a base demandou uma conversão para XML por meio do BibTeXML (Zearin, 2016). Os seguintes procedimentos foram adotados:

1) pré-tratamento do arquivo .bib:

– mudança da posição das *custom tags* “citado”, que passaram a ser inseridas dentro das chaves de cada referência;

– *tags* “citado =” alteradas para uma *tag* por artigo citante (ex.: citado = {42215} / citado = {45315}) - os números se referem ao *id* do artigo na plataforma OJS (*ojsid*) – em substituição ao mecanismo executado anteriormente, no qual havia uma *tag* com a lista de citantes (ex.: citado = {42215,45315});

– verificação geral de inconsistências (quebras de linhas, espaços, *tags* não fechadas) com BibTeXML e conversor/indentador *xmllint*;

2) conversão para XML e testes de consistência: Execução do BibTeXML, em ciclos repetitivos, e aplicação do seguinte *checklist*:

– quantidade de referências no arquivo .bib é a mesma no “*dataframe*” do R?

– referências com mais de uma citação estão recebendo a respectiva multiplicação nas contagens?

– há referências sem pelo menos uma ocorrência da *tag* “citado”?

– quantidade de referências associadas a cada artigo citante é a mesma verificada na página de metadados deste no OJS?

Efetivou-se uma sequência de testes de consistência para as variáveis de cada indicador, como segue:

⁴ *Web scraping* é um conjunto de métodos que permite a extração automática de conteúdo na Internet. Diversas linguagens de programação – como o R – têm recursos (como estratégias de autenticação, identificação de padrões, tratamento) para simular a navegação humana e coletar dados (Munzert et al., 2014).

1) indicador baseado em títulos de periódicos: desambiguaram-se títulos registrados na base (ex.: Encontros Bibli; e Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação);

2) indicador baseado em nomes de autores: ao se identificarem valores "others" na tag "author", caso em que o autor do artigo escreveu apenas "et. al.", recuperaram-se os dados complementares originais para treze ocorrências verificadas;

3) indicador baseado em palavras-chave: desambiguaram-se e uniformizaram-se os termos (ex.: "dados abertos vinculados" e "dados abertos governamentais" = "dados abertos"). A redução da lista de possíveis termos auxiliou na representatividade mínima no *corpus* de palavras-chave para a geração dos indicadores;

4) indicador baseado em áreas de conhecimento: a uniformização resultou em uma categoria "Macroárea", a partir da combinação preenchida pelo autor (ex.: ciência da informação: estudos métricos, na macroárea = ciência da informação).

Para facilitar a leitura e possível uso dessa metodologia, os indicadores foram divididos em dois grandes grupos: (1) baseados apenas nos metadados; (2) cruzados entre metadados e base BibTeX/AtoZ. Essa distinção é importante ao se considerar o grau de complexidade e investimento de recursos para a elaboração dessa base de referências. Os metadados, ainda que possam ter problemas de preenchimento, são usualmente inseridos no OJS, o que permite uma indexação mais eficiente do periódico. O uso do BibTeX no processo editorial, por sua vez, não é frequente e exige um maior esforço e emprego de recursos.

Resultados

No total, foram estruturados, testados e avaliados dez indicadores (Quadro 2), sendo sete deles determinados apenas por metadados do OJS. Cinco indicadores foram considerados como tendo impacto para a presente discussão.

Quadro 2. Descrição e observações acerca dos indicadores selecionados – Periódico AtoZ (2011-2016).

1 de 2

Indicador/observações	Fonte dos dados/ tratamento	Resultados comentados
<i>Triangulação de dados</i> a) Conformidade dos metadados: total de artigos, total de <i>tags title</i> por idioma, total de <i>tags resumo</i> , total de documentos publicados. b) Distribuição de documentos por seção e número.	OAI-PMH	Teste: artigos com títulos em apenas dois idiomas, e não três (como proposto no periódico). Comentário: a verificação manual é possível, mas a compilação automática permite resposta mais rápida e permanente aos erros.
<i>i1. Autores por Artigos</i>	OAI-PMH	Ver tópico dedicado: 3.1 Autores por artigo (i1).
<i>i2. Colaboração internacional</i>	<i>Web scraping</i> nas páginas de metadados dos artigos.	Ver tópico dedicado: 3.2 Colaboração Internacional (i2).
<i>i3. Palavras-chave</i> Distribuição de frequência das Palavras-chave empregadas nos artigos.	OAI-PMH Tratamento: correção manual de redundâncias.	Teste: verificou-se o efeito cauda longa: 184 valores únicos/95% com apenas uma ocorrência. As palavras com maior frequência estiveram alinhadas à proposta editorial. Comentário: permite avaliar, periodicamente, o alinhamento entre a proposta editorial (foco e escopo) e o que está sendo publicado. A consistência (quantidade, redundâncias, tradução) das palavras-chave é fundamental para a avaliação e indexação do periódico.

Nota: OAI-PMH: *Open Archives Initiative Protocol for Metadata Harvesting*.

Fonte: Elaborado pelos autores, maio 2017.

Quadro 2. Descrição e observações acerca dos indicadores selecionados – Periódico AtoZ (2011-2016).

2 de 2

Indicador/observações	Fonte dos dados/ tratamento	Resultados comentados
<i>i4. Grau de formação</i> a) Distribuição de grau de formação máxima, dentre os autores de um artigo, ao longo dos volumes. b) Discriminado pela presença de pelo menos um(a) Doutor(a). c) Cálculo de percentual de autores, em cada volume, que possuíam doutorado completo.	<i>Web scraping</i> nas páginas de metadados dos artigos. Tratamento: no campo de biografia do autor, definiu-se um procedimento lógico para categorizar o conteúdo disponível em formato texto.	Teste: verificou-se um valor de 80% de autores doutores no primeiro número, mantendo-se a média de 60%, com um número considerado <i>outlier</i> (20% de autores doutores). Comentário: o valor ideal seria próximo de 100% de autores doutores, pois este é um atributo avaliado como um critério de observância das agências indexadoras.
<i>i5. Endogenia</i> a) Distribuição de instituição de origem por autor, ao longo dos volumes. b) Discriminado pelo percentual de autores da instituição de origem. c) Cálculo de percentual de autores, em cada número, filiados à instituição de origem do periódico.	OAI-PMH Tratamento: correção manual de redundâncias.	Teste: dois volumes obtiveram valor maior que 10%: o número de estreia (29%, associado a submissões por convite), e o volume 5 (24%). Comentário: editores buscam reduzir essa presença, demonstrando não somente poder de atração de material exógeno, mas reforçando seu papel como veículo de divulgação científica nacional e internacional sem conflito de interesses.
<i>i6. Métodos</i> Distribuição de métodos de pesquisa empregados (nenhum, um ou mais - por artigo).	OAI-PMH Tratamento: correção manual das redundâncias.	Teste: alinhamento ao escopo declarado nas diretrizes. Comentário: apesar de não ser um critério formal das agências indexadoras, pode auxiliar o periódico na verificação de volume/porcentagem de artigos estratificados por tipo de método de pesquisa e linha editorial.
<i>i7. Áreas e Macroáreas</i> Distribuição de áreas e macroáreas de conhecimento (um ou mais por artigo).	OAI-PMH Tratamento: correção manual das redundâncias e separação por macroáreas.	Teste: macroáreas alinhadas às linhas editoriais, com ascensão da área de tecnologia educacional. Comentário: é possível avaliar tanto por macroáreas (ex.: ciência da informação), como de forma específica (ex.: ciência da informação: processo de disseminação da informação), o que pode revelar conformidade com as linhas editoriais do periódico.
<i>i8. Autores</i>	OAI-PMH e base BibTeX/AtoZ.	Ver tópico dedicado: 3.3 Autores mais referenciados (i8).
<i>i9. Estatísticas bibliométricas: palavras-chave, macroáreas, métodos, corpus de referências.</i>	OAI-PMH e base BibTeX/AtoZ.	Ver tópico dedicado: 3.4 <i>Corpus</i> de referências por variável discriminante (i9).
<i>i10. Grafo de citações</i>	OAI-PMH e base BibTeX/AtoZ.	Ver tópico dedicado: 3.5 Grafo de citações (i10).

Nota: OAI-PMH: *Open Archives Initiative Protocol for Metadata Harvesting*.

Fonte: Elaborado pelos autores, maio 2017.

Autores por artigo (i1)

Nesse indicador, separou-se o conjunto de artigos de cada volume pela quantidade de autores. Por exemplo, no volume 2, três artigos com autoria única, seis artigos com autoria dupla e um artigo com cinco autores. Em posse dessa distribuição, é possível calcular a média autores/artigo em cada volume (2,10, no exemplo) e compará-lo com valores desejados na área.

Os resultados do periódico AtoZ demonstram um valor médio (cerca de dois autores/artigo) compatível com outros estudos na área de Ciências Sociais Aplicadas, como, por exemplo, os levantamentos feitos por Santos (2010) e por de Souza, Silva e Araújo (2013). O padrão não foi mantido no volume 4, onde há um artigo com cinco autores, um artigo com quatro autores, e um artigo com três autores, o que eleva a média para 3,30 autores/artigo.

Também foi identificado um *outlier* no volume 3, pois, apesar de a média ser compatível (2,54), um artigo foi assinado por oito autores.

Algumas políticas explícitas em diretrizes para a publicação tendem, segundo Garcia *et al.* (2010), a restringir o número de coautores, mesmo que se reconheça que distintos campos do conhecimento aceitam a chamada “hiperautoria”, na qual um único artigo pode ter dezenas de colaboradores (Cronin, 2001).

Colaboração internacional (i2)

Para o cálculo, extraiu-se a informação do país de afiliação institucional de cada autor do conjunto de dados, referindo-se tais contagens aos autores, e não aos artigos. Foi possível contabilizar a presença de autores de mais de um país, indicando uma distribuição institucional e geográfica de autores, o que pode potencializar a visibilidade do periódico (Packer; Meneghini, 2006) (Tabela 1). A análise da concentração ou dispersão das colaborações pode contextualizar a “capilaridade” do periódico, o que deve ser complementado pelo levantamento da quantidade de autores de cada país, e não apenas pelo percentual total de autores estrangeiros, como se comentará em seguida.

A prática desejada pelas agências indexadoras é que os periódicos tenham a maior quantidade possível de artigos publicados com a colaboração de autores estrangeiros, definindo, por exemplo, percentuais mínimos desejados, de acordo com a área do conhecimento, como o caso da Scientific Electronic Library Online (2014). Conforme os dados da Tabela 1, o periódico apresenta algum avanço nesse sentido, ao ter alcançado 30% e 33% nos volumes 2 e 3, respectivamente, sofrendo uma redução no volume 4, e alcançando um valor mais significativo no quinto (53%). Ressalva-se, como ponto de alerta, a concentração de material de pesquisadores de universidades equatorianas, cujos artigos – em comparação aos demais publicados no periódico –, contam com um maior número de coautorias. O aumento da dispersão geográfica dos autores é um indicador da capacidade do periódico em captar artigos de distintas comunidades de pesquisadores, o que é fundamental para o reconhecimento da qualidade do veículo de publicação (Packer; Meneghini, 2006).

Tabela 1. Distribuição de origem geográfica de autores, por país de afiliação institucional e por volume publicado – Periódico AtoZ (2011-2016).

Origem dos autores	Volumes					Total
	v1	v2	v3	v4	v5	
Nacional (%)	100	70	68	83	47	74
Brasil	24	14	27	20	8	93
Estrangeira (%)	0	30	33	17	53	26
Equador		6	10	4	6	26
Espanha					2	2
Portugal			3		1	4

Fonte: Elaborado pelos autores com base nos metadados do periódico AtoZ (2011-2016), maio de 2017.

Autores mais referenciados (i8)

Os indicadores i8, i9 e i10 estão alinhados aos pressupostos da análise bibliométrica (Thelwall, 2008), entendida no contexto da produção científica como um conjunto de métodos e técnicas apoiadas em recursos estatísticos, que permite explorar variáveis, tais como: distribuição de artigos por ano, *ranking* de autores mais frequentes, padrões de autoria, *ranking* geográfico, *ranking* de instituições mais produtivas, colaboração interinstitucional, distribuição de frequência de palavras-chave etc. (Keshava; Gireesh; Gowda, 2008).

No caso do i8, a contagem pela *tag author* na base BibTeX/AtoZ permitiu identificar os autores mais referenciados no *corpus* e, caso uma referência apresentasse mais de um autor, todos foram considerados. Aprimorou-se a análise com a admissão de três outras variáveis, a saber:

1) quantidade de artigos: um mesmo autor pode ter sido citado várias vezes em apenas um trabalho, afetando o contexto do peso que ele agrega na literatura citada no periódico;

2) quantidade de autocitações: a autocitação ocorre quando um autor referencia um trabalho em que ele mesmo também foi autor/coautor. Agências indexadoras tendem a descartá-las, alegando o risco de distorção nos índices;

3) homônimos: dada a dificuldade de desambiguação, têm reflexos na consistência da base de dados para análise.

Nos testes com o AtoZ, verificou-se o efeito cauda longa, com um total de 1.685 nomes unívocos. Pela limitação de tempo para a análise, apenas os 29 primeiros (aqueles com frequência ≥ 3) foram selecionados para avaliação de quantidade de artigos, autocitação e homonímia.

Caso se desejasse indicar – via quantidade geral de citações –, os “autores mais importantes para o conteúdo publicado”, haveria um viés, o qual foi testado em uma simulação: um determinado autor *alfa* foi citado apenas três vezes, porém com relação a três artigos diferentes, enquanto um autor *beta* obteve seis citações, porém todas de um mesmo artigo.

Quanto a autocitações, os resultados demonstraram que um dado autor *gamma* recebeu sete citações (todas, porém, de um mesmo artigo de sua própria autoria), enquanto o autor *omega* recebeu quatro citações, não sendo autor/coautor dos trabalhos citados. A discussão sobre autocitação é controversa, pois pode tanto revelar um dos aspectos da pesquisa em colaboração, como, ao contrário, uma posição “egoísta” (ou de isolamento) frente ao esforço científico (Lawani, 1982).

As questões de homonímia, frequentes em estudos bibliométricos (Narin *et al.*, 1976; Aksnes, 2008), também foram identificadas no conjunto de dados do periódico AtoZ. Dos 29 autores (frequência ≥ 3), quatro receberam citações com variações na grafia do nome (com ou sem acrônimos), acarretando erros na contagem. Nesse mesmo grupo, verificou-se que, de três citações recebidas pelo autor *epsilon*, duas não se referiam a ele quando se efetivou a análise detalhada (e manual) dos prenomes.

Corpus de referências por variável discriminante (i9)

Na análise proposta foram considerados:

1) A quantidade de referências: quantas referências foram utilizadas nos artigos citantes, e que tinham nos metadados determinada variável discriminante (palavra-chave; macroárea de conhecimento ou método empregado).

2) Os artigos citantes: quantos artigos citantes tinham a variável discriminante em seus metadados.

3) A proporção referências/artigos citantes: dividiu-se a quantidade de referências pelos artigos citantes, tendo-se como resultado uma média de referências que o *corpus* apresenta por artigo.

4) A meia-vida de citações: visou identificar a obsolescência das referências citadas, com base nas variáveis discriminantes consideradas.

5) A meia-vida de citações para artigos: idem à anterior, filtrada por referência tipo “artigo”.

6) As referências sem ano: foram informadas as quantidades de referências que, por não terem a *tag* ano, foram descartadas do cálculo de meia-vida.

7) A distribuição de tipos de fonte: para cada *corpus* de referências, qual o percentual por tipo de fonte empregado nas referências.

8) A distribuição de autores mais citados: para cada *corpus* de referências, qual o percentual por autores mais citados.

9) A distribuição de periódicos mais citados: para cada *corpus* de referências, qual o percentual por títulos de periódicos nas referências.

Dada a extensão original dos resultados, apresentam-se – como exemplo da potencialidade na geração de indicadores –, aqueles derivados da variável discriminante “macroárea: Ciência da Computação”, a qual totalizou treze artigos citantes e 231 referências (Tabela 2).

Dados idênticos foram gerados para as demais macroáreas (Ciência da Informação, Educação, Interdisciplinar), assim como para as palavras-chave mais frequentes (redes sociais, inteligência artificial, jogos educativos) e para os métodos de pesquisa mais utilizados (exploratório, qualitativo, documental, estudo de caso), totalizando dez seleções de *corpora* e, conseqüentemente, tabelas de resultados.

Em todos os conjuntos gerados pelo estudo foi possível perceber que as variáveis discriminantes foram determinantes para a construção de *corpora* distintos, comprovando seu papel para avaliação de diferentes comportamentos relativos ao uso de referências.

Tabela 2. Amostra de indicadores dos *corpora* de referências da macroárea de conhecimento Ciência da Computação – Periódico AtoZ (2011- 2016).

Macroárea: Ciência da Computação		
Tipo	n	
Referências	231	
Artigos Citantes	13	
Proporção Referências / Artigo Citante	17,8	
Meia-vida das citações – todos os tipos	5	
Meia-vida das citações – apenas artigo <article>	6	
Referências sem tag “ano” <year>	8	
Distribuição de fontes		
Tipo	Frequência (n)	%
Article	63	27.27
Book	73	31.60
Inbook	5	2.16
Valores suprimidos (...)
Distribuição de Autores		
Autor(a)	Frequência (n)	Artigos citantes (n)
C. Bizer	6	1
L. A. Digiampietri	5	1
R. Mugnaini	3	2
T. Berners-Lee	3	1
A. Illanas	2	1
Valores suprimidos (...)
Distribuição de Periódicos Citados		
Título	Frequência (n)	
Archival Science	2	
Revista Latinoamericana de Tecnología Educativa	2	
Accounting Research Journal	1	
AI Magazine	1	
Applied Ecology	1	
Valores suprimidos (...)	...	

Fonte: Elaborado pelos autores com base nos metadados e arquivo BibTeX do Periódico AtoZ (2011-2016), maio 2017.

Grafo de citações (i10)

De maneira similar ao i9 (*Corpus* de referências por variável discriminante [...]), foram selecionadas as referências citadas em artigos das seguintes macroáreas de conhecimento: Ciência da Informação, Ciência da Computação, Educação e Interdisciplinar.

De forma a amenizar o efeito de cauda longa, optou-se pela coleta, referência a referência, do título do periódico citado, o que conferiu maior densidade às relações. Esse indicador tomou a forma de um grafo de citações (Figura 1), no qual as arestas são as ligações entre macroárea de conhecimento (obtida nos metadados do artigo citante) e o periódico citado nas referências. Conforme a legenda, as variáveis contempladas na estrutura são: classe citado e citante (via formato do nó), quantidade de citações (de títulos de periódicos) por todas as macroáreas (tamanho do nó), e quantidade de diferentes macroáreas citantes (cor do nó).

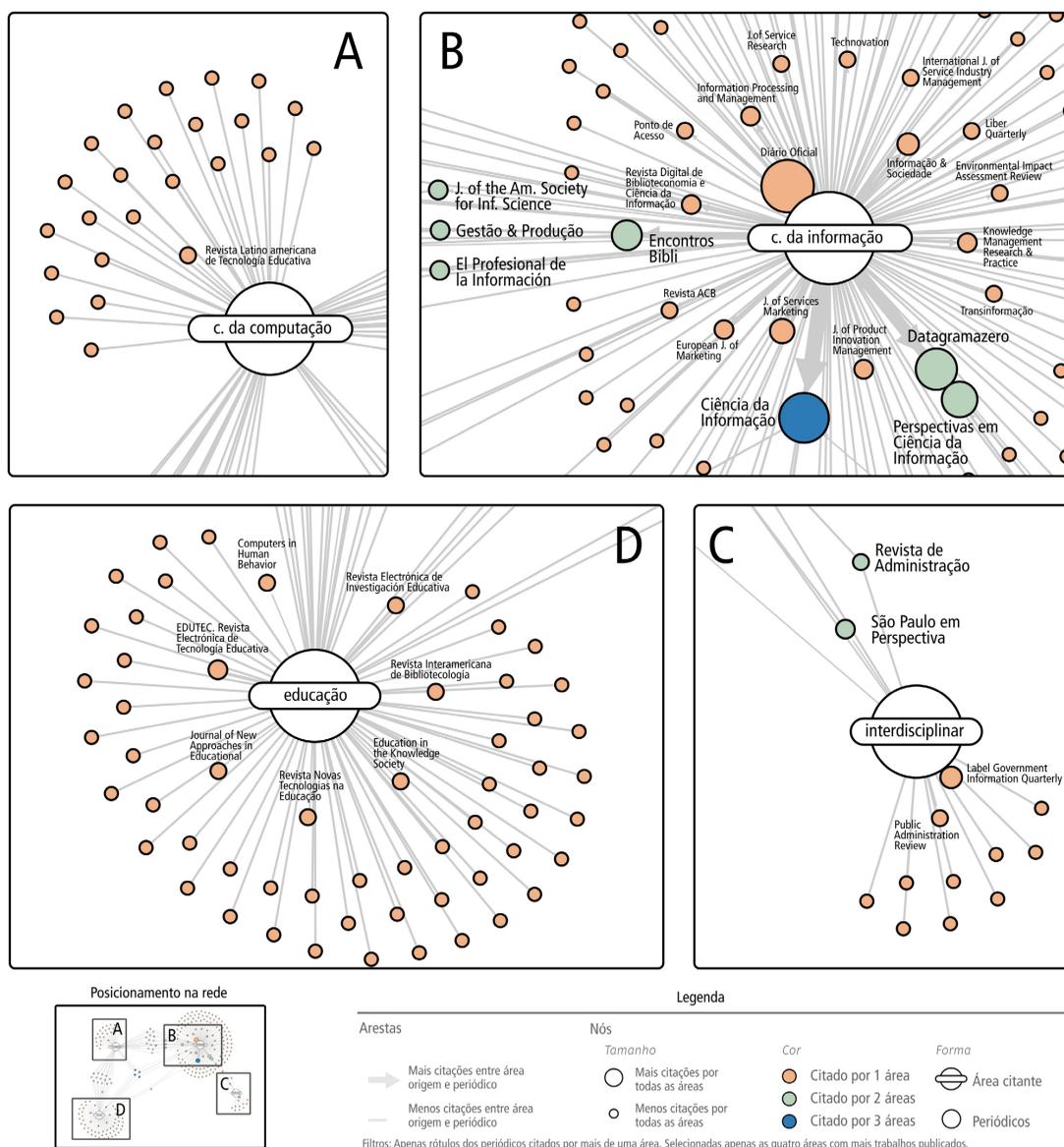


Figura 1. Representação de periódicos citados, de acordo com as macroáreas de conhecimento dos artigos citantes – Periódico AtoZ (2011- 2016).

Fonte: Elaborado pelos autores com base nos metadados e arquivo BibTeX do Periódico AtoZ (2011-2016), maio de 2017.

Ainda que o grafo não represente associações explícitas com os critérios que o periódico possa vir a atender segundo agências indexadoras, ele permite explorar a conformação das macroáreas de conhecimento e de seus periódicos nucleares, bem como os periódicos citados que interligam distintas macroáreas e ainda a distribuição em cauda longa em alguns casos.

Conclusão

Estudos de natureza bibliométrica – e outros de caráter estatístico –, apresentam resultados diretamente proporcionais à confiabilidade dos dados a serem manipulados. Igualmente, conjuntos mais robustos de dados (decorrentes de alimentação contínua e progressiva) aumentam a capacidade de gestores para projetar metas, estabelecer objetivos e tomar decisões frente às análises derivadas. Editores de periódicos que desejem produzir indicadores domésticos – e habilitar tecnicamente tais periódicos a indicadores externos –, devem estar atentos a tais quesitos de qualidade de dados e à garantia das fontes destes (no caso desta proposta, uma base BibTeX própria e um conjunto de metadados de todos os volumes e números do periódico). Tais condições foram verificadas nos resultados, pois, em quatro dos indicadores propostos (i3 – palavras-chave; i5 – endogenia; i6 – método; e i7 – áreas e macroáreas), foi necessário o tratamento manual para retirar redundâncias, tendo sido feitos testes de consistência voltados à desambiguação. O fato de a investigação ter provocado mudanças na alimentação dos metadados e nos procedimentos de registro na base BibTeX revela-se como um efeito colateral positivo, assim como um alerta para outros gestores que se interessem em repetir o estudo.

Para o periódico AtoZ, os resultados indicam pontos de atenção a serem considerados pelos gestores, tais como: a redução da endogenia; a captação de material que amplie a representatividade na distribuição institucional e geográfica dos autores, assim como de autores/coautores com titulação de doutorado; o ajuste (já aplicado) nas diretrizes, voltado à limitação no número de coautores; a avaliação da necessidade de ampliação do corpo de consultores (comitê consultivo) que inclua especialistas na área de Educação (em especial tecnologias aplicadas ao ensino-aprendizagem) – como resultado visualizado na Figura 1; a captação de submissões na macroárea de Administração, presente na linha editorial do periódico, mais fracamente representada quando se analisam os resultados; e a verificação cuidadosa dos métodos de pesquisa utilizados pelos autores das submissões em relação à proposta editorial do periódico. Resultados obtidos na análise das palavras-chave revelam que o periódico está publicando alguns temas emergentes, tais como redes sociais, jogos educativos e inteligência artificial.

No caso específico do periódico AtoZ, projeta-se – a médio prazo –, a revisão/aprimoramento da diagramação em LaTeX e sua migração para um *Schema Extensible Markup Language/Journal Article Tag Suite* (XML/JATS), facilitando sua integração às bases indexadoras e outros sistemas em rede. A base de dados de referências, atualmente em BibTeX, tem suficiente granularidade para ser submetida a um processo de conversão quando da transição para formatos utilizados por tais bases, sem exigências de retrabalho. A máxima aderência pretendida, na proposta, com as exigências de agências indexadoras (ainda que limitada ao conjunto de dados atualmente disponíveis no periódico AtoZ) acompanha um objetivo colateral de qualquer periódico acadêmico-científico em sua trajetória, ou seja, reduzir/racionalizar recursos físicos, financeiros e humanos quando de sua incorporação a respeitadas índices de citação.

O processo apresentado tem limitações inerentes a um estudo aplicado em um único ambiente, ainda que seja orientado às exigências de agências indexadoras e voltado à geração de tais levantamentos de forma semiautomatizada. As ferramentas utilizadas para a manipulação das referências, dos metadados e dos procedimentos relativos aos cruzamentos efetuados nos testes podem ser modificadas de acordo com as necessidades daqueles que considerarem factível a reprodução deste estudo em seus ambientes. Um segundo movimento de extrapolação pode ser encaminhado ao se reproduzir o conjunto de procedimentos, ferramentas e indicadores definidos nesta investigação em um recorte mais amplo de periódicos. O fato de as fontes dos dados (tanto os metadados, quanto

as referências) serem de domínio público, em parte significativa dos periódicos que utilizam o OJS, é um ponto positivo e estimulante para a continuidade dos estudos. Contudo, a qualidade dessas fontes de dados pode limitar o espectro dos indicadores sugeridos. Um estudo adicional, apoiado nesses pressupostos, está em progresso pelo presente grupo de pesquisadores, com o objetivo de ampliar essa discussão inicial e verificar o posicionamento do periódico AtoZ frente a um universo mais amplo de publicações.

Colaboradores

P. Marchiori, E. Bettoni e M. Carvalho participaram da concepção, coleta e tabulação dos dados. P. Marchiori, E. Bettoni e A. Appel colaboraram na análise dos dados, na redação e na revisão do artigo.

Referências

- Aksnes, D.W. When different persons have an identical author name. How frequent are homonyms? *Journal of the Association for Information Science and Technology*, v.59, n.5, p.838-841, 2008. <http://dx.doi.org/10.1002/asi.20788>
- Banos, V. *Open archives initiative protocol for metadata harvesting (OAI-PMH): Validator e data extractor Tool*. Tessalônica, Grécia: OAI-PMH Validator, 2017. Available from: <<http://validator.oaipmh.com>>. Cited: May 16, 2017.
- Bar-Ilan, J. Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, v.74, n.2, p.257-271, 2008. <http://dx.doi.org/10.1007/s11192-008-0216-y>
- Cronin, B. Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, v.52, n.7, p.1532-2890, 2001. <http://dx.doi.org/10.1002/asi.1097>
- Garcia, C.C. *et al.* Authorship for scientific papers: The new challenges. *Brazilian Journal of Cardiovascular Surgery*, v.25, n.4, p.559-567, 2010. <http://dx.doi.org/10.1590/S0102-76382010000400021>
- Garfield, E. How ISI selects journals for coverage. *Current Contents*, v.13, n.22, p.185-193, 1990. Available from: <<http://www.garfield.library.upenn.edu/essays/v13p185y1990.pdf>>. Cited: May 16, 2017.
- Harzing, A.W.; Van der Wal, R. Google Scholar as a new source for citation analysis? *Ethics in Science and Environmental Politics*, v.8, n.1, p.61-73, 2008. <http://dx.doi.org/10.3354/esep00076>
- Keshava, G.; Gireesh, A.; Gowda, M.P. ACM transaction on information systems (1989-2006): A bibliometric study. *Information Studies*, v.14, n.4, p.223-234, 2008. Available from: <https://www.researchgate.net/publication/303683455_acm_transactions_on_information_systems_1989-2006_a_bibliometric_study>. Cited: May 16, 2017.
- Lawani, S.M. On the heterogeneity and classification of author self-citations. *Journal of the American Society for Information Science and Technology*, v.33, n.5, p.280-284, 1982. <http://dx.doi.org/10.1002/asi.4630330506>
- Meijer, E. The apacite package: Citation and reference list with LaTeX and BibTeX according to the rules of the American Psychological Association. In: *Comprehensive TEX Archive Network*. Heidelberg: CTAN, 2013. Available from: <<https://ctan.org/pkg/apacite>>. Cited: May 16, 2017.
- Munzert, S. *et al.* *Automated data collection with R: A practical guide to web scraping and text mining*. New Delhi: Wiley, 2014.
- Narin, F. *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Washington: Computer Horizons, 1976. Available from: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.473.8004&rep=rep1&type=pdf>>. Cited: May 16, 2017.
- Packer, A.L.; Meneghini, R. Visibilidade da produção científica. In: Población, D.A.; Witter, G.P.; Silva, J.F.M. (Org.). *Comunicação e produção científica: contexto, indicadores e avaliação*. São Paulo: Angellara, 2006. p.235-259.
- Santos, S.M. *Perfil dos periódicos científicos de Ciências Sociais e de Humanidades: mapeamento das características extrínsecas*. 176f. 2010. Dissertação (Mestrado em Ciência da Informação) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2010. Disponível em: <<http://www.scielo.org/local/content/pdf/058.pdf>>. Acesso em: 16 maio 2017.
- Scientific Electronic Library Online. *Critérios SciELO Brasil: critérios, políticas e procedimentos para a admissão e a permanência de periódicos científicos na Coleção SciELO Brasil*. 2014. Disponível em: <http://www.scielo.br/avaliacao/20141003NovosCritérios_SciELO_Brasil.pdf>. Acesso em: 31 maio 2017.
- Souza, F.J.V.; Silva, M.C.; Araújo, A.O. Uma análise da produção científica da área de contabilidade governamental nos periódicos que utilizam o SEER e o SciELO. *Revista de Contabilidade da UFBA*, v.7, n.2, p.22-37, 2013. Disponível em: <<https://portalseer.ufba.br/index.php/rcontabilidade/article/view/6268/6450>>. Acesso em: 16 maio 2017.
- The R Foundation. *The R Project for Statistical Computing*. 2017. Available from: <<https://www.r-project.org/>>. Cited: May 16, 2017.
- Thelwall, M. Bibliometrics to webometrics. *Journal of Information Science*, v.34, n.4, p.605-621, 2008. <http://dx.doi.org/10.1177/0165551507087238>
- Winter, J.C.F.; Zadpoor, A. A.; Dodou, D. The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics*, v.98, n.2, p.1547-1565, 2014. <http://dx.doi.org/10.1007/s11192-013-1089-2>
- Zearin. *BibTeXML*. San Francisco: GitHub Repositor, 2016. Available from: <<https://github.com/Zearin/BibTeXML>>. Cited: May 16, 2017.