

UFPR/SCSA/Grupo de Pesquisa “Metodologias em Gestão da Informação”  
Revista eletrônica AtoZ: novas práticas em informação e conhecimento

MSc Eduardo Michelotti Bettoni<sup>1</sup>  
Marcelo Batista de Carvalho  
Profa. Dra. Patricia Zeni Marchiori

**GERAÇÃO DE INDICADORES PARA PERIÓDICOS CIENTÍFICOS:  
UM ESTUDO NA ATOZ**

Apêndice do Relatório final referente ao  
Edital de Apoio à Editoração e Publicação  
de Periódicos Científicos - 2016  
(UFPR/PRPPG/SIBI)

Curitiba  
2016

1 Email para contato com os autores: [contatoatoz@gmail.com](mailto:contatoatoz@gmail.com)

## RESUMO

Projeto aprovado no Edital de apoio à editoração de periódicos científicos (PRPPG/UFPR - 2016), para aplicação do recurso nas atividades do periódico AtoZ: Novas práticas em informação e conhecimento (PPGCGTI/UFPR). Objetivou constituir uma base de referências dos trabalhos publicados e planejar indicadores de desempenho da revista. Verificou-se que não somente o corpus de referências mas também a conversão dos metadados da revista em indicadores podem servir como orientação para melhores tomadas de decisão no processo editorial em vista dos critérios postulados pelas agências indexadoras. Por sua vez, a análise da base constituída também permite um melhor entendimento sobre as relações entre o periódico e seus pares, além de dar suporte à um processo de editoração mais assertivo com o uso do formato BibTeX.

**Palavras-chave:** Periódico. Bibliometria. Análise de Produção Científica.

## SUMÁRIO

1	INTRODUÇÃO.....	4
2	BASE DE REFERÊNCIAS - A BASE BIB (BibTeX/AtoZ).....	8
3	GERENCIAMENTO DE REFERÊNCIAS NO FORMATO BibTeX: TESTES DE FERRAMENTAS.....	14
4	ESTUDO EMPÍRICO COM INDICADORES: TESTES COM A REVISTA ATOZ.....	19
4.1	Exploração das variáveis.....	19
4.2	Formação da base de dados para os indicadores: OJS e BibTeX/AtoZ.....	23
4.3	Definição de indicadores e resultados dos testes.....	27
4.3.1	Indicadores baseados apenas nos metadados.....	29
4.3.1.1	Triangulação de dados.....	30
4.3.1.2	i1. Autores por trabalho.....	31
4.3.1.3	i2. Internacionalização.....	32
4.3.1.4	i3. Palavras-chave.....	32
4.3.1.5	i4. Grau de formação.....	34
4.3.1.6	i5. Endogenia com base nas instituições de origem.....	35
4.3.1.7	i6. Métodos de pesquisa empregados.....	36
4.3.1.8	i7. Áreas de conhecimento.....	37
4.3.2	Indicadores com metadados e base BibTeX.....	39
4.3.2.1	i8. Autores mais referenciados.....	39
4.3.2.2	i9. Corpus de referências por variável discriminante.....	41
4.3.2.3	i10. Grafo de citações - áreas de conhecimento e periódicos.....	48
5	CONSIDERAÇÕES.....	52
	REFERÊNCIAS.....	54

## 1 INTRODUÇÃO

Os **Indicadores** constituem um instrumento de planejamento em relação a definição de objetivos e estratégias de uma organização. São parâmetros qualificados e/ou quantificados que servem para detalhar em que medida os objetivos de um projeto foram ou serão alcançados dentro de um prazo limitado de tempo e numa localidade específica (FERREIRA; CASSIOLATO; GONZALES, 2009).

Identificar o impacto de uma pesquisa é fundamental para autores, avaliadores, editores e patrocinadores da ciência e a **bibliometria** é uma das formas de responder este tipo de problema. (KOSTOFF et. al., 2001).

A **Análise bibliométrica** é um método para sumarizar estudos publicados na literatura científica por meio de alguns indicadores (THELWALL, 2008). Sob esta abordagem as publicações são analisadas com o apoio de recursos estatísticos, permitindo explorar variáveis como: distribuição de artigos por ano, *ranking* de autores mais frequentes/produtivos, padrões de autoria, *ranking* geográfico, *ranking* de instituições mais produtivas, colaboração interinstitucional, referências por documento, distribuição de frequência de palavras-chave etc (KESHAVAL; GIREESH; GOWDA, 2008).

A bibliometria costuma ser alvo de críticas por reduzir a complexidade das relações no contexto acadêmico em indicadores de caráter numérico. Porém, em geral, é o um passo inicial necessário para a realização de análises mais robustas. Kostoff et. al (2001) e Hung (2011) aplicaram técnicas de bibliometria como base para posterior realização de *text mining* em *corpora* científicos, ressaltando a importância da primeira para os estudos.

Uma das técnicas bibliométricas mais exploradas é a análise de citações. Em literatura científica, **citações** são instâncias de referências a outros trabalhos científicos já publicados (HU; CHEN; LIU, 2013). Price (1986) defende uma convenção para os termos “citação” e “referência”. Se um artigo R contém uma nota bibliográfica usando e descrevendo um artigo C, então R contém uma referência a C e, C é citado por R. Com isso, o número de referências que um artigo possui pode ser medido pela quantidade de itens em sua seção de notas

bibliográficas, também conhecida como **lista de referências**. O número de citações de um artigo é, normalmente, identificado em um índice de citações (base de dados contendo registros de artigos citantes e citados) e se relaciona à contagem de menções que recebeu de outros artigos. As bases mais conhecidas atualmente são Web of Science, Scopus e Google Scholar.

A Web of Science surgiu da junção entre as bases Science Citation Index, Social Sciences Citation Index e Arts and Humanities Citation Index. Tais bases foram criadas originalmente junto ao Institute for Scientific Information (ISI), instituído por Eugene Garfield, também responsável pela criação do índice Fator de Impacto (FI) (Garfield, 1955; 1964; 1972). É uma base comercial acessível mediante assinatura, pertencente até o ano de 2016 ao grupo Thomson Reuters, ano em que foi vendida (THOMSON REUTERS, 2016) aos grupos Onex Corporation, sediado em Toronto, e Baring Private Equity Asia, sediado em Hong Kong. A WoS coleta e armazena (indexa) sistematicamente os principais metadados dos artigos publicados em periódicos previamente selecionados, quais sejam: título do artigo, autores e suas afiliações institucionais, resumo, palavras-chave, citações etc., que possibilitam a geração do FI e do Journal Citation Report (JCR), a compilação anual dos índices de FI dos periódicos indexados na base.

O fator de impacto de um periódico científico, principal indicador de análise de citações, é calculado, de acordo com Borgman e Furner (2002, p. 17), da seguinte forma: divide-se a) o número de citações recebidas no ano corrente pelos artigos publicados no periódico nos últimos dois anos, pelo b) número de artigos publicados pelo periódico nos dois últimos anos. Os autores destacam que o FI de um periódico não é usado somente como fator de competitividade com outros periódicos, mas também como um substituto para a contagem de citações de um determinado artigo, especialmente na avaliação da produção dos autores-pesquisadores para a tomada de decisões acerca de promoções ou nomeações para cargos (Borgman e Furner, 2002, p. 18).

A Scopus, assim como a Web of Science, é uma base bibliográfica comercial acessível mediante assinatura. Pertence ao mesmo grupo da editora Elsevier e é responsável pela emissão do SCImago Journal Rank (SJR), do índice h, além de outros indicadores. O SJR é um índice que, além do número de citações, também

leva em consideração a influência e o prestígio dos periódicos citantes, ou seja, um artigo citado será melhor posicionado no ranking quanto mais ele for citado por periódicos de maior influência que, por sua vez, são muito citadas por outros artigos (GONZÁLEZ-PEREIRA et al., 2010).

O Google Scholar (ou Google Acadêmico) é reconhecido como uma ferramenta de busca, de acesso gratuito, de variadas formas de publicações científicas, tais como artigos científicos, livros, artigos publicados em eventos científicos, teses e dissertações etc. Tais publicações são indexadas automaticamente pela ferramenta com base em características previamente definidas e que permitem a identificação dessas publicações como científicas. A partir do índice elaborado, o Google Scholar também possibilita a geração de indicadores bibliométricos, tais como o número de citações recebidas pelos documentos indexados e o índice  $h$  de pesquisadores e periódicos indexados. Alguns autores (HARZING; WAL, 2008; BAR-ILAN, 2008) apontam vantagens e desvantagens no uso do Google Scholar como ferramenta para análises bibliométricas, contudo, não deixam de reconhecer a sua maior abrangência e diversidade de conteúdos indexados em relação às bases Web of Science e Scopus.

Outro indicador para análise de citações, o índice  $h$ , proposto por Jorge Hirsch (2005), leva em conta o número de publicações e o número de citações de um determinado autor/pesquisador. De acordo com Hirsch (2005, p. 16569), um cientista tem índice  $h$  se  $h$  de seus artigos  $N$  (número de artigos) tiverem pelo menos  $h$  citações cada um e os demais artigos ( $N - h$ ) tiverem  $\leq h$  citações cada. Por exemplo, um autor terá índice  $h$  igual a 6 se tiver publicado ao menos seis 6 artigos com seis citações ou mais.

O índice  $h$  foi idealizado inicialmente por Hirsh para quantificar a produção científica de um indivíduo, contudo Braun, Glänzel e Schubert (2006) sugeriram a utilização desse índice também ao nível dos periódicos, como complementação ao fator de impacto. O índice  $h$ , tanto de autores como de periódicos, pode ser obtido por meio das bases Scopus, Web of Science e Google Acadêmico. Bar-Ilan (2008), contudo, chama a atenção para o fato de que cada base pode apresentar um valor diferente de índice  $h$  para um mesmo autor ou uma mesma publicação, em função das bases apresentarem grande variação na cobertura.

Para periódicos não indexados, existe a possibilidade de realização de análises bibliométricas com base em dados e indicadores derivados de seus próprios materiais. Embora não permitam a comparação direta com outros periódicos, podem fornecer *insights* para os editores nos seus processos de gestão editorial e com vistas à elaboração de estratégias para a ampliação da qualidade do material que é publicado, bem como para a ampliação da visibilidade e alcance do junto ao seu público-alvo. Para criação e consolidação de um indicador é essencial que os parâmetros possam ser medidos com base em fontes de dados confiáveis. Buscando-se a realização de análises bibliométricas, faz-se necessária a constituição de uma base própria de referências e metadados do corpus de artigos do periódico alvo.

Com o intuito de explorar e testar indicadores dessa natureza, este documento descreve um estudo realizado com dados coletados a partir da revista eletrônica AtoZ: novas práticas em informação e conhecimento. A trajetória metodológica relatada envolve não somente a construção de bases de dados de referências e metadados do periódico, mas também um estudo para a composição de indicadores, tal como um levantamento e análise de ferramentas para o gerenciamento de referências e outros mecanismos de apoio. Como resultados, apresenta-se a descrição dos indicadores elencados, assim como análises e visualizações a partir de tais indicadores.

## 2 BASE DE REFERÊNCIAS - A BASE BIB (BibTeX/AtoZ)

A revista eletrônica AtoZ: novas práticas em informação e conhecimento<sup>2</sup> foi definida como periódico teste para os estudos. Em 2014, como resultado de uma bolsa de Iniciação Científica, criou-se e implantou-se uma *template* LaTeX como o ambiente padrão para a edição e diagramação dos manuscritos. A geração automatizada, a partir do LaTeX, dos diferentes formatos finais dos artigos da AtoZ (PDF, HTML e ePub) e a adoção do BibTeX têm, desde então, facilitado o reuso de referências. O BibTeX é um projeto *open-source* e funciona como uma extensão não obrigatória para o pacote de macro LaTeX. **Os arquivos com extensão .bib** são aqueles associados à uma base de dados bibliográfica BibTeX (*BibTeX Bibliography Database File*).

Seu uso é recomendável, pois assim como ocorre com outros gerenciadores de referências (tais como o Endnote), o formato das referências pode ser modificado dinamicamente (*on the fly*) a partir da armazenagem de todas as referências em uma base de dados *flat-file*<sup>3</sup>. Assim como no LaTeX, a formatação é feita pela ferramenta e a relação do conjunto de dados BibTeX com o conteúdo, via citações, se dá no momento da compilação.

A base BibTeX/AtoZ é resultado da compilação das referências dos materiais publicados na revista AtoZ, registradas em formato BibTeX conforme orientações do manual do pacote *apacite* (MEIJER, 2013) na sequência de passos relatada a seguir.

Para os volumes 1 a 3 da revista, as referências foram convertidas com apoio da ferramenta AnyStyle<sup>4</sup>. Essa ferramenta é um *parser* de referências bibliográficas, ou seja, ela analisa a estrutura de cada referência, separando os segmentos segundo seu significado (nomes dos autores, título do material, local de publicação etc). Até o volume 3/número 1 as referências do periódico estavam normalizadas segundo a norma NBR6023 da Associação Brasileira de Normas Técnicas (ABNT).

2 Disponível em: <<http://www.revistas.ufpr.br/atoz>>.

3 *Flat file* é um arquivo de texto, geralmente contendo um registro por linha, utilizado quando não [se tem] uma quantidade grande de informações a ser trabalhada, ou não há necessidade de uma manipulação mais exigente dos dados de uma determinada aplicação. (REVISTABW, 2013).

4 Disponível em: <<https://anystyle.io/>>.



Do volume 3/ número 2 em diante seguiram-se orientações da Associação Americana de Psicologia (APA).

Devido à adoção do LaTeX no processo editorial do periódico, as referências do volume 4 (e posteriores) já se encontravam em arquivos BibTeX. Elas foram revisadas e incluídas na base BibTeX/AtoZ. Esta base consiste em um único arquivo BibTeX contendo, até o momento, 1.129 referências. Adicionou-se cada nova referência a ele via um editor de documentos TeX/LaTeX para ambientes KDE, intitulado Kile, conforme a estrutura do Código 1:

```
@tipo_de_entrada{chave,  
  nome_do_campo_1 = {valor_do_campo_1},  
  nome_do_campo_2 = {valor_do_campo_2},  
  ...  
  nome_do_campo_N = {valor_do_campo_N},  
} citado = {id_do_material_citante}
```

Código 1 – Estrutura das entradas de referência na base BibTeX/AtoZ

A primeira linha do Código 1 contém o tipo de entrada (dentre aqueles previstos pelo pacote *apacite* - Quadro 1) e uma *chave* única identificadora da entrada.

<b>Tipo de entrada</b>	<b>Usado para</b>	<b>Observações</b>
article	Artigos em periódicos científicos ou similares	
magazine	Artigos em revistas	
newspaper	Artigos em jornais noticiosos	
book	Livros completos	
incollection	Capítulos ou artigos em livros, anais de congresso ou similares	usado inbook e inproceedings
techreport	Relatórios	
intechreport	Artigos em relatórios	
lecture	Apresentações	usado inproceedings
phdthesis	Teses de doutorado	
mastersthesis	Dissertações de mestrado, monografias e trabalhos de conclusão de curso	
unpublished	Pré-prints e documentos não publicados	
misc	Sites, filmes, discos, softwares e outros materiais que não se encaixam nas outras categorias	
literal	Documentos em que seja necessário ajuste manual em sua aparência na lista de referências	não usado
manual	Manuais	
booklet	Panfletos	não usado
inbook	Capítulos em livros ou similares	
inproceedings	Trabalhos em anais de congresso ou similares	
proceedings	Anais de congresso como um todo ou similares	

Quadro 1 - Tipos de referência aceitos no pacote *apacite*.

Cada linha subsequente apresenta em *valor\_do\_campo* uma característica da entrada (nome dos autores, ano de publicação etc.) identificada por *nome\_do\_campo* (também previsto pelo *apacite* - Quadro 2). Na última linha, em *id\_do\_material\_citante*, são apresentados um ou mais códigos identificadores dos

materiais que citam a referência. Assim, o campo *citado* identifica quantos e quais itens/materiais do periódico citaram determinada referência.

Campo (continua)	Tipo de referência							
	article	book	inbook e inpro- ceedings	tech- report	intech- report	masters- thesis e phdthesis	unpu- blished	manual, misc e proceed- ings
address		✓	✓	✓	✓	✓	✓	✓
author	✓	✓	✓	✓	✓	✓	✓	✓
booktitle			✓		✓			
chapter			✓					
day	✓		✓	✓	✓		✓	✓
edition		✓	✓	✓	✓	✓		✓
editor	✓	✓	✓	✓	✓		✓	✓
firstkey	✓	✓	✓	✓	✓	✓	✓	✓
howpublished	✓	✓	✓	✓	✓	✓	✓	✓
institution				✓	✓			
journal	✓					✓		
key	✓	✓	✓	✓	✓	✓	✓	✓
lastchecked	✓	✓	✓	✓	✓	✓	✓	✓
month	✓		✓	✓	✓		✓	✓
note	✓	✓	✓	✓	✓	✓	✓	✓
number	✓	✓		✓	✓	✓	✓	✓
organization							✓	
originaladdress	✓	✓	✓		✓			
originalbooktitle	✓	✓	✓		✓			
originaledition	✓	✓	✓		✓			
originaleditor	✓	✓	✓		✓			

Campo (conclusão)	Tipo de referência							
	article	book	inbook e inpro- ceedings	tech- report	intech- report	masters- thesis e phdthesis	unpu- blished	manual, misc e proceed- ings
originaljournal	✓	✓	✓		✓			
originalnumber	✓	✓	✓		✓			
originalvolume	✓	✓	✓		✓			
originalyear	✓	✓	✓	✓	✓	✓	✓	✓
pages	✓		✓		✓	✓		
publisher		✓	✓					✓
school						✓		
title	✓	✓	✓	✓	✓	✓	✓	✓
translator	✓	✓	✓	✓	✓	✓	✓	✓
type	✓	✓	✓	✓	✓	✓	✓	✓
url	✓	✓	✓	✓	✓	✓	✓	✓
volume	✓	✓	✓	✓	✓	✓		✓
year	✓	✓	✓	✓	✓	✓	✓	✓

Quadro 2 - Campos de dados usados na conversão de referências.

Fonte: adaptado de (MEIJER, 2013).

As chaves identificadoras foram estabelecidas com base nos sobrenomes dos autores e ano de publicação com o intuito de evitar chaves repetidas e, ao contrário do que ocorre com uma chave numérica, por exemplo, facilitar sua aplicação. O arquivo foi ordenado alfabeticamente pelas chaves identificadoras. Entradas duplicadas foram mescladas mantendo-se os códigos de todos os citantes no campo *citado* na última linha de cada registro BibTeX.

Dadas as especificidades apresentadas para conversão e manutenção da base de referências BibTeX - e o aumento a cada edição do número de referências registradas num único arquivo - a inclusão de novas entradas pode se tornar não-trivial com o uso de um editor (como o Kile, por exemplo). Conforme abordado na

sequência, a adoção de uma ferramenta gerenciadora de referências é crítica para esta consistência.

Observaram-se alguns problemas e inconsistências nas referências. Uma delas é a possível existência de homônimos, potencializada pelo uso de abreviações nos primeiros nomes dos autores. Para o estudo de indicadores (detalhado posteriormente neste documento) foi necessário modificar as referências que utilizavam a expressão *et al* para indicar a existência de mais autores, substituindo-a pelos nomes dos autores faltantes. Ainda que adequado para o gerenciamento de referências, o registro tipo único com múltiplos citantes se mostrou inadequado para a criação de indicadores, pois tal adaptação foi uma decisão da equipe gestora do periódico e não uma funcionalidade já existente no BibTeX. Verificou-se, neste momento, que tal decisão incompatibilizou a base BibTeX/AtoZ para seu uso nas ferramentas existentes, o que exigiu a criação de uma estrutura própria para tratamento e análise dos dados.

### 3 GERENCIAMENTO DE REFERÊNCIAS NO FORMATO BibTeX - TESTES DE FERRAMENTAS

Realizaram-se testes com ferramentas gratuitas<sup>5</sup> a fim de identificar a mais adequada para o gerenciamento das referências do periódico. As ferramentas testadas foram:

- a) Aigaion<sup>6</sup>: sistema de gerenciamento bibliográfico baseado em Web;
- b) Bebop<sup>7</sup>: interface Web para referências salvas em arquivo BibTeX;
- c) BibBase<sup>8</sup>: gerenciador de páginas Web de publicação científica;
- d) Bibdesk<sup>9</sup>: gerenciador de referências BibTeX para MacOS X;
- e) Bibus<sup>10</sup>: gestor de referências bibliográficas compatível com editores de texto;
- f) BibSonomy<sup>11</sup>: sistema de compartilhamento de publicações científicas e *bookmarks*;
- g) CiteULike<sup>12</sup>: serviço online de descoberta e gerenciamento de referências;
- h) Colwiz<sup>13</sup>: gerenciador de pesquisa científica para Web, computadores e celulares;
- i) Docear<sup>14</sup>: gerenciador pessoal de referências e literatura científica;
- j) JabRef<sup>15</sup>: gerenciador *open source* de referências bibliográficas em BibTeX;
- k) KBibTeX<sup>16</sup>: editor de referências BibTeX para a suíte de aplicativos KDE;
- l) Mendeley<sup>17</sup>: gerenciador de referências e rede social científica;
- m) Pybliographer<sup>18</sup>: ferramenta de gerenciamento de bases de dados bibliográficas;
- n) Qiqqa<sup>19</sup>: gerenciador de referências e pesquisas acadêmicas;

5 Com base em: <[https://en.wikipedia.org/wiki/Comparison\\_of\\_reference\\_management\\_software](https://en.wikipedia.org/wiki/Comparison_of_reference_management_software)>.

6 Disponível em: <<https://sourceforge.net/projects/aigaion>>.

7 Disponível em: <<http://people.alari.ch/derino/Software/Bebop>>.

8 Disponível em: <<http://bibbase.org>>.

9 Disponível em: <<http://bibdesk.sourceforge.net>>.

10 Disponível em: <<http://bibus-biblio.sourceforge.net>>.

11 Disponível em: <<https://www.bibsonomy.org>>.

12 Disponível em: <<http://www.citeulike.org>>.

13 Disponível em: <<https://www.colwiz.com>>.

14 Disponível em: <<http://www.docear.org>>.

15 Disponível em: <<https://www.jabref.org>>.

16 Disponível em: <<http://home.gna.org/kbibtex>>.

17 Disponível em: <<https://www.mendeley.com>>.

18 Disponível em: <<http://pybliographer.org>>.

19 Disponível em: <<http://www.qiqqa.com>>.

- o) ReadCube<sup>20</sup>: programa para leitura, gerenciamento e descoberta de artigos acadêmicos;
- p) Refbase<sup>21</sup>: interface Web para gerenciamento de literatura científica e citações;
- q) RefME<sup>22</sup>: ferramenta Web de gerenciamento de pesquisa, referências e citações;
- r) RefDB<sup>23</sup>: ferramenta para bases de dados de referências bibliográficas;
- s) Referencer (ou Referenciador)<sup>24</sup>: programa do projeto GNOME para gerenciar referências e gerar arquivos BibTeX;
- t) RefTeX<sup>25</sup>: pacote para o editor de textos Emacs que permite o uso de rótulos, referências, citações e índices em documentos LaTeX;
- u) Wikindx<sup>26</sup>: sistema Web de autoria de artigos e gerenciamento de referências e citações;
- v) Zotero<sup>27</sup>: ferramenta para coletar, organizar, citar e compartilhar referências de pesquisa.

Na fase de testes inseriu-se a base de referências da revista AtoZ em cada uma das ferramentas, avaliando-as em seis quesitos, a saber:

- a) *foco*: se a ferramenta é voltada ao gerenciamento de referências, permitindo adicionar, editar e excluir entradas (assinalado no quadro como “R”); à leitura de documentos para pesquisas, com suporte ao vínculo do documento original à referência (assinalado “L”); ou outros focos (“O”). A aderência a este quesito foi considerada crítica na avaliação;
- b) *manipulação de referências*: se existe a possibilidade de inserção, edição e exclusão de entradas pela ferramenta. Esta é uma característica desejável, pois o editor pode, com base nesta, interagir com a biblioteca de referências por meio de uma interface, e não via código;

20 Disponível em: <<https://www.readcube.com>>.

21 Disponível em: <<http://www.refbase.net>>.

22 Disponível em: <<https://www.refme.com>>.

23 Disponível em: <<http://refdb.sourceforge.net>>. **Observação:** não foi possível instalar esta ferramenta para testes, tendo sido desconsiderada.

24 Disponível em: <<https://launchpad.net/referencer>>.

25 Disponível em: <<http://www.gnu.org/software/auctex/reftex.html>>.

26 Disponível em: <<http://wikindx.sourceforge.net>>.

27 Disponível em: <<https://www.zotero.org>>.

- c) *adição de campos*: a possibilidade de se trabalhar com campos extras não previstos pela ferramenta. Esta característica é essencial para que não se percam itens no banco de dados, principalmente no campo *citado*;
- d) *uso do arquivo .bib*: o modo como o arquivo BibTeX é utilizado. O uso pode ser direto, ou seja, o programa lê e escreve diretamente no arquivo BibTeX (marcado no quadro como “D”); ou via importação (“I”) e exportação (“E”) do arquivo. O uso direto é desejável, assim como o fato da importação e exportação ocorrerem sem perda de dados;
- e) *relatórios*: se a ferramenta apresenta a possibilidade de gerar relatórios e compilação da base de referências, como frequências de autores e periódicos, por exemplo; e
- f) *busca externa*: a opção de obter dados externos usando o *Digital object identifier* (DOI), o nome do autor, o título do periódico etc. Esta característica é desejável para reduzir o trabalho de digitação dos dados e a chance de erros.

Os resultados das avaliações são indicados no Quadro 3.



Ferramenta	Foco	Manipulação de referências	Adição de campos	Uso do arquivo .bib	Relatórios	Busca externa
Aigaion	R	✓		I/E		
Bebop	R	✓		D	✓	
BibBase	O			I/E		✓
Bibdesk	R	✓		D		✓
BibSonomy	O	✓		D		
Bibus	R	✓		I/E		✓
CiteULike	R	✓		I/E		
Colwiz	R	✓		I/E		✓
Docear	L	✓	✓	I/E		✓
JabRef	R	✓	✓	D		✓
KBibTeX	R	✓	✓	D	✓	✓
Mendeley	L	✓		I/E		
Pybliographer	R	✓	✓	D		
Qiqqa	L	✓		I/E		✓
ReadCube	L	✓		I/E		✓
RefBase	R	✓		I/E		✓
RefME	R	✓		E		✓
Referencer	L	✓	✓	I/E		✓
RefTeX	O					
Wikindx	R	✓	✓	I/E	✓	✓
Zotero	L	✓		I/E		

Quadro 3 - Análise de ferramentas para manipulação de referências BibTeX

Foco: R - manipulação de referências; L - leitura de documentos para pesquisas; O - outros focos.

Uso do arquivo .bib: D - uso direto; I - importação; E - exportação;

As ferramentas Docear, Mendeley, Qiqqa, ReadCube, Referencer e Zotero têm foco no gerenciamento de leitura e pesquisa científica, o que não é adequado ao propósito deste trabalho. O Docear, baseado na ferramenta JabRef, permite a

criação de mapas mentais para representar a biblioteca. O Qiqqa, exclusivo para Windows, faz sugestões de leituras. O ReadCube e o Referencer (ou Referenciador) importam dados de fontes externas, como o Google Acadêmico. O Zotero permite a criação de linhas do tempo. Dessas, apenas o Docear e o Referencer permitem o uso de campos adicionais. Essas ferramentas não suportam a criação de relatórios a partir da base de referências.

O BibBase é uma ferramenta para criação de páginas Web com listas de referências científicas. O BibSonomy é uma ferramenta voltada principalmente para o compartilhamento de publicações científicas e criação de *bookmarks*. Já o RefTeX é um *plugin* para o editor de textos Emacs, voltado para a criação de referências e citações em documentos LaTeX. Os focos do BibBase, BibSonomy e RefTeX não são adequados para este trabalho.

Todas as ferramentas com foco no gerenciamento de referências permitem a manipulação (adição, edição e remoção) de registros. Dentre elas, o JabRef, o KBibTeX, o Pybliographer e o Wikindx aceitam o uso de campos adicionais automaticamente ou através de configuração prévia. As ferramentas Bebo, KBibTeX e Wikindx emitem relatórios com estatísticas descritivas, tais como frequências de autores ou tipos de referência. O Bibdesk, Bibus, Colwiz, JabRef, KBibTeX, RefBase, RefME e Wikindx importam dados de bases externas, como o Google Acadêmico, Web of Science e CrossRef. As que trabalham diretamente no arquivo BibTeX são: Bebo, Bibdesk, JabRef, KBibTeX e Pybliographer.

As ferramentas KbibTeX e JabRef se mostraram mais adequadas para o gerenciamento de referências por apresentarem mais características em conformidade com aquelas desejadas pelo grupo de trabalho.

Nenhuma das ferramentas analisadas, porém, foi compatível com o estudo de indicadores. Embora algumas delas tenham a capacidade de geração de relatórios, estes se limitam a listas de frequências e, mesmo assim, com algumas falhas em função da estrutura adotada no periódico alvo.

## 4 ESTUDO EMPÍRICO COM INDICADORES: TESTES COM A REVISTA ATOZ

No escopo deste estudo os indicadores foram constituídos com base em dados internos do periódico alvo, ou seja, algo de domínio do corpo do editorial. A decisão foi pautada no grau de complexidade menor e no objetivo indireto de facilitar o acesso dos editores aos indicadores e, principalmente, permitir que estes sejam partícipes na sua formação a partir da busca pela qualidade dos metadados e referências em seus periódicos.

A exploração dos indicadores nesta etapa de testes se deu por duas fontes principais de dados: (i) base BibTeX/AtoZ, formulada e gerenciada conforme definido nos tópicos anteriores; (ii) conjunto de metadados dos artigos citantes (ou simplesmente, metadados do periódico alvo).

### 4.1 Exploração das variáveis

Os dados no formato BibTeX podem ser resumidos, em um primeiro momento, por frequência e/ou distribuição com distintos graus de complexidade. Pode-se, inicialmente, “contar” uma variável (ex.: “Instituições citadas”, via a *tag bibcorporate*), avançando-se na definição de cruzamentos mais robustos como, por exemplo, a “distribuição de instituições, por cidade e ano em que foram citadas, e apenas em periódicos”. Em ambos casos, a origem do dado é a própria base BibTeX, sem qualquer associação à uma fonte externa. Uma lista das possíveis contagens é apresentada no Quadro 4.

<b>Estatística (frequência)</b>	<b>Tags do BibTeX</b>	<b>Tipo de entrada BibTeX</b>
Instituições citadas	bibcorporate	todas
Distribuição temporal das publicações citadas	year	todas
Recursos (artigo, <i>site</i> , livro) citados	title id year	todas
Periódicos citados	journal	article
Fascículo de um periódico	journal volume	article
Distribuição temporal de acesso à fonte	lastchecked	todas
Editoras de livros citadas	publisher	book
Eventos citados	booktitle howpublished	inproceedings
Instituições de pós-graduação com teses/dissertações/monografias citadas	school	phdthesis, masterthesis
Cidades das instituições de pós-graduação com teses/dissertações/monografias citadas	address	phdthesis, masterthesis
Média de autores por publicações citadas	author	todas
Autores citados	author	todas
Livro citado (inteiro ou capítulo)	title	book inbook

Quadro 4 - Estatísticas univariadas baseadas apenas no BibTeX

Essas variáveis podem ser associadas e resumidas de acordo com o objetivo do estudo, ou seja, a pergunta que se deseja responder com os dados. Em determinada situação, apenas a frequência simples de uma delas pode responder à uma demanda. Adicionalmente, cruzamentos entre elas podem aumentar o grau de complexidade e também de significado. O quadro 5 contém alguns exemplos (não exaustivos) dessas associações:

Variáveis	Tags
Distribuição de <i>periódicos citados</i> por ano de publicação	journal, year
Distribuição de trabalhos publicados em eventos, citados, por ano de publicação (tipo de entrada)	booktitle, year
Distribuição de fontes (tese de doutorado, dissertação de mestrado) por instituição	phdthesis, masterthesis, address

Quadro 5 - Estatísticas multivariadas baseadas apenas no BibTeX

Mesmo limitando a fonte de dados à base BibTeX/AtoZ, é possível explorar os indicadores acima, na forma de um extenso sumário - ou uma lista de estatísticas sobre todas as referências. Porém, existem outros encaminhamentos que podem enriquecer essa análise. O mais relevante, no contexto deste estudo, é estabelecer relações entre recursos citantes - quais artigos do periódico alvo citam determinada(s) referências - e os recursos citados - elementos da base BibTeX/AtoZ.

Nesta base são reunidas todas referências dos artigos publicados no periódico alvo. Alguns fundamentos que regem a sua estrutura - e, conseqüentemente, as possibilidades de análise -, além daqueles que já foram citados no tópico da constituição de uma base de referências, são:

- a) qualquer referência que esteja na base BibTeX/AtoZ teve pelo menos uma ocorrência dentro do *corpus* de artigos do periódico alvo;
- b) qualquer referência presente em qualquer artigo do periódico alvo será encontrada na base BibTeX/AtoZ;
- c) um mesmo artigo pode ter sido referenciado por um ou mais artigos de autores do periódico alvo e, caso isto ocorra, não haverá redundância de registros. Ou seja, a base BibTeX/AtoZ registra apenas uma ocorrência. Uma única exceção se aplica a referências do tipo "sites", dado seu aspecto dinâmico, pois a data de acesso costuma ser diferente a cada citação. Neste caso, a base apresenta dois registros diferentes.

Esse princípios conduzem as relações entre a base BibTeX/AtoZ (e todos os seus campos supracitados) e os artigos citantes (e todos seus metadados). É fundamental verificar se há conformidade e, portanto, consistência na base. Os metadados essenciais dos artigos citantes - potencialmente existentes na plataforma

OJS - e que podem dar base à cruzamentos na formação de indicadores são listados no quadro 6.

Metadado do OJS - propriedade do artigo citante	Observações
Autoria	Pode ser explorado pelo autor principal ou autor a autor. Neste último caso é importante ressaltar que cada autor do trabalho contabiliza uma variável por citação.
Filiação do autor(es)	
País de origem do autor(es)	
Título / Title	Como indicador, pois pode ser explorado por extensão, por meio de <i>tag cloud</i> ou alguma análise de conteúdo/ <i>text mining</i> .
Palavras-Chave / Keywords / Palabras-clave	Caso sejam preenchidas e controladas, a análise é facilitada.
Seção	Como é um dado controlado e com poucos valores possíveis, pode servir de variável discriminante.
Resumo / Abstract / Resumen	Como indicador por ser explorado por extensão, por meio de <i>tag cloud</i> ou alguma análise de conteúdo / <i>text mining</i> .
Ano de publicação	
Área(s) do Conhecimento	Caso sejam preenchidas e controladas, a análise é facilitada.
Idioma	
Tipo, método ou ponto de vista	Caso sejam preenchidos e controlados, a análise é facilitada.
Agências de fomento	Necessita de normalização e devem ser preenchidas e controladas para facilitar a compilação

Quadro 6 - Variáveis dos metadados do periódico-alvo que podem ser exploradas nos indicadores

Os metadados listados podem, inicialmente, constituir um próprio grupo de indicadores que seja independente das referências. Ainda que eles já estejam disponíveis de maneira orgânica para o editores de revistas, via plataforma OJS; não estão consolidados e/ou sumarizados. Um levantamento de frequência de áreas de conhecimento, por exemplo, demandaria acessar artigo a artigo, algo que pode ser facilitado por outros métodos e oferecer *insights* para o editor em relação ao alinhamento da proposta editorial com o *corpus* de trabalhos publicados.

Portanto, tanto a base BibTeX, como o conjunto de metadados podem oferecer indicadores isolados. Contudo, associar estas duas fontes de dados apresenta um desafio mais complexo. Cada conjunto de metadados pode ser diretamente associado à um grupo de referências e sumarizado, por meio da relação citante > citado, como no exemplo:

*qualquer artigo citante tem a propriedade “ano de publicação”.*

*Esse mesmo artigo tem um conjunto de referências na base .bib.*

*dada esta ligação, é possível dizer quais das referências da base bib foram citadas num dado ano.*

*um segundo passo seria também utilizar uma propriedade da base .bib (o ano do material citado, ex.) e criar um indicador por ano de publicação, ou da média ou, ainda, da distribuição de anos dos artigos citados.*

Em resumo, cada estatística dos Quadros 4 e 5 pode ser cruzada / sumarizada com cada variável do Quadro 6, possibilitando a construção de indicadores.

#### **4.2 Formação da base de dados para os indicadores: OJS e BibTeX/AtoZ**

Considerando o estágio já avançado da BibTeX/AtoZ enquanto fonte de dados, inicialmente foi realizado um trabalho para coleta e tratamento daqueles oriundos dos metadados do periódico alvo.

Como o enfoque desse estudo é permitir sua reprodução pelos editores em diversos periódicos, a abordagem da obtenção dos metadados foi aquela em que não fossem demandadas permissões, como acesso especial ao banco de dados do OJS. Para atingir este objetivo, três métodos de obtenção merecem destaque: (a) via exportação pelos plugins nativos, de acesso pelo editor; (b) via protocolo OAI-PMH, com o qual é possível ter acesso aos dados sem qualquer *login*; (c) via *robots* que executam um *web scraping* na página da revista.

Pelo acesso, qualidade e interoperabilidade optou-se, primeiramente, pelo método (b) - com qual é possível chegar aos mesmos resultados de (a), conjugado aos *robots* (c). Esta última opção, apenas quando o dado não está disponível via

protocolo OAI-PMH. Na formação de uma base única de metadados para a geração dos indicadores, as seguintes etapas foram realizadas.

- a) **Dados do periódico alvo extraídos via protocolo OAI-PMH.** Por meio do endereço <http://revistas.ufpr.br/atoz/oai> os dados foram descarregados e convertidos em XML único com apoio da ferramenta OAI-PMH Validator<sup>28</sup>, especificamente em sua função `ListRecords OAI_DC`. Outras ferramentas foram investigadas - como o OAI CSV Parser<sup>29</sup> - sem efetivo uso nesta fase.
- b) **Verificação de consistência do XML.** Após um teste de compatibilidade do XML, verificou-se que alguns atributos da *tag root* <OAI-PMH> causavam ruído no método de *parser xpath*, exigindo uma correção manual.
- c) **Importação e manipulação do XML.** O software escolhido para manipulação dos dados em XML foi o R - por meio da IDE R Studio, especificamente com o `package::xml2`.
- d) **Relatórios iniciais gerados apenas com a fonte OAI-PMH.** Após a importação no R, os dados do XML foram explorados quanto à contagens, consistência de variáveis, e qualidade dos metadados. Verificou-se a necessidade de utilizar o *web scraping* para a obtenção de alguns dados.
- e) **Execução de Web scraping para complemento de dados.** As variáveis nome, origem institucional e país de origem dos autores foram obtidas por um *web scraping* realizado na página de cada um dos artigos, com uso do `package::RCurl`, dentro do ambiente R e, posteriormente, processado com `package::xml2`.

Com a associação entre os dois métodos foi possível acercar a todos os metadados necessários para realização do estudo, partindo das variáveis desejáveis elencadas no tópico Exploração de Variáveis.

Deve-se observar que a base BibTeX/AtoZ foi constituída no formato adequado à compilação (o BibTeX), mas este não é o mais adequado para a sumarização e análise de dados. Para a continuidade do trabalho optou-se, assim como para os metadados, pela conversão para XML.

28 Disponível em: <<http://validator.oaipmh.com>>.

29 Disponível em: <<https://github.com/openstate/oai-csv-parser>>.



Testaram-se algumas ferramentas (Bibutils<sup>30</sup> e JabRef) mas estas apresentaram falhas, principalmente na codificação (pela presença de caracteres latinos) e de não interpretação de alguns tipos de documentos (ex.: *newspaper*). A solução adotada foi o BibTeXML<sup>31</sup> (ZEARIN, 2016), um projeto ainda embrionário no github. Apesar do BibTeXML estar no estágio inicial, é o resultado da herança de um projeto homônimo que havia sido abandonado há alguns anos. O conversor não apresentou problemas sendo capaz de: (i) manter acentos e caracteres latinos, (ii) aceitar todos os tipos de documentos, (iii) aceitar *custom tags* como “citado”. Assim como nos metadados, alguns procedimentos foram necessários.

a) **Tratamento do arquivo base BibTeX/AtoZ.** Antes mesmo da conversão algumas alterações foram necessárias:

- i. exclusão de todos os comentários e marcações que eram apenas referenciais (ex.: separadores por letra inicial do autor);
- ii. inserção da *tag* “citado” dentro das chaves da referência para, durante a conversão, ser interpretada;
- iii. *tags* “citado =” alteradas para uma *tag* por artigo citante (citado = {42215<sup>32</sup>} / citado = {45315}), e não uma *tag* com a lista de citantes (citado = {42215,45315});
- iv. verificação geral de inconsistências geradas pela edição manual (quebras de linhas, espaços, *tags* não fechadas). Para tal utilizou-se o conversor BibTeXML e o conversor/identador *xmllint*.

b) **Conversão para XML e testes de consistência.** Execução do *script* da ferramenta BibTeXML, em várias repetições, de maneira a efetuar o checklist indicado no Quadro 7:

30 Disponível em: <<https://sourceforge.net/p/bibutils/home/Bibutils/>>.

31 Disponível em: <<https://github.com/Zearin/BibTeXML>>.

32 Os números se referem ao id do artigo na plataforma OJS (ojsid).

Checklist	Resultado
quantidade de referências na base é a mesma de referências no <i>data.frame</i> dentro do R?	não houve necessidade de correção
as referências com mais de uma citação estão recebendo a respectiva multiplicação na contabilização?	correções feitas no cálculo de cada indicador
existem referências sem pelo menos uma ocorrência da tag "citado"?	corrigido
cada artigo / short paper tem pelo menos uma referência na base? Não, havia um código trocado (41314 na verdade era 41341)	corrigido
a quantidade de referências associadas à um ojsid é o mesmo que verificado na página de metadados do mesmo no OJS?	corrigido <b>obs:</b> Para esse teste fez-se checagem manual dos 70 artigos
todas as entradas possuem uma chave única?	encontraram-se duplicatas que foram corrigidas

Quadro 7 - Checklist e resultado dos testes de consistência da conversão para XML.

Após essa verificação prévia de inconsistências e da respectiva correção, os indicadores definidos foram processados e novas rodadas de ajuste foram necessárias. Por uma questão de organização do documento, estes testes serão apresentados no presente tópico, previamente ao capítulo de detalhamento dos indicadores. Abaixo, listam-se a sequência de testes efetuados.

- a) **Testes nos dados do indicador de periódicos.** Após a geração de uma lista de valores únicos na sumarização dos periódicos citados foi necessária uma desambiguação de nomes na base BibTeX/AtoZ (ex.: *Encontros Bibli* e *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*)
- b) **Testes nos dados do indicador de autores.** Após a geração de uma lista de valores únicos de autores citados:
  - i. foi identificado "others" como autor, quando o autor do artigo escreveu apenas *et. al.*, inclusive nas referências. Foi necessário recuperar os originais das treze ocorrências.
  - ii. foram encontrados trabalhos da tag <author>, os quais foram excluídos em qualquer contabilização de indicador de autoria

- c) **Testes nos dados do indicador de palavras-chave.** Após a geração de uma lista de valores únicos de palavras-chave dos metadados foi necessária uma desambiguação (ex.: *dados abertos vinculados* e *dados abertos governamentais* = *dados abertos*). Essa redução da lista de possíveis palavras-chave foi necessária para que o *corpus* de palavras-chave atingisse uma representatividade mínima na geração dos indicadores.
- d) **Testes nos dados do indicador de áreas de conhecimento.** O mesmo procedimento que foi realizado com as palavras-chave, exceto pelo fato da criação de uma categoria *Macroárea*, a partir da combinação preenchida pelo autor. (ex.: *ciência da informação : estudos métricos*, na macroárea = *ciência da informação*).

Estes são, portanto, todos os procedimentos relacionados à alteração de valores nas fontes de dados (recodificação) - e/ou criação de novas variáveis - que apoiaram a formulação dos indicadores.

### 4.3 Definição de indicadores e resultados dos testes

Após uma pré-análise dos dados e ações voltadas para a redução de inconsistências/ruídos, a equipe de trabalho realizou rodadas de priorização e selecionou os indicadores arrolados no Quadro 8.

Conforme apresentado anteriormente, o espectro de variáveis das duas fontes de dados (base BibTeX/AtoZ e metadados OJS) tem potencial para geração de diversos indicadores, requerendo uma seleção daqueles mais alinhados ao propósito do estudo enquanto suporte ao trabalho dos editores. Cada um deles será detalhado em conjunto com os respectivos resultados no presente tópico.

Indicador (continua)	Definição	Fonte dos dados e observações
Triangulação de Dados	Verificação de conformidade dos metadados da revista (total de artigos, total de tags title por idioma, total de tags resumo, total de galleys), distribuição de documentos por seção e número da revista.	Protocolo OAI-PMH
i1_autoresartigos	Distribuição da quantidade de autores por trabalho, ao longo dos volumes da revista	Protocolo OAI-PMH
i2_internacionalização	Distribuição de países de origem dos autores, ao longo dos volumes da revista. Discriminado pelo percentual entre nacional e internacional.	<i>Web scraping</i> nas páginas de metadados dos artigos. Tratamento: separação de valores no campo "bio" do autor.
i3_palavraschave	Distribuição de frequência das palavras chave em português empregadas em artigos da revista.	Protocolo OAI-PMH Tratamento: correção manual de redundâncias
i4_graudeformacao	Distribuição de grau de formação máxima, dentre os autores de um artigo, ao longo dos volumes da revista. Discriminado pela presença de pelo menos um(a) Doutor(a).	<i>Web scraping</i> nas páginas de metadados dos artigos Tratamento: definição de lógica para categorizar as formações disponíveis em formato texto no campo de bio do autor.
i5_endogenia	Distribuição de instituição de origem por autor, ao longo dos volumes da revista. Discriminado pelo percentual de autores da instituição de origem	Protocolo OAI-PMH Tratamento: correção manual de redundâncias.
i6_metodos	Distribuição de métodos de pesquisa empregados (nenhum, um ou mais por artigo)	Protocolo OAI-PMH Tratamento: correção manual das redundâncias.
i7_areas e i7b_areas	Distribuição de áreas de conhecimento (um ou mais por artigo)	Protocolo OAI-PMH Tratamento: correção manual das redundâncias e separação por macroáreas.
i8_autores	Frequência de autores mais referenciados com sinalização de distribuição (citados em quantos artigos diferentes); autocitação (em quantos também era co-autor do trabalho) <ul style="list-style-type: none"> <li>• atenção para possíveis homônimos: o autor é o mesmo das "x" obras).</li> </ul>	Protocolo OAI-PMH e base BibTeX/AtoZ Tratamento: avaliação de autocitação e homonímia feita apenas para os mais citados (n>=3), de forma manual.

Indicador (conclusão)	Definição	Fonte dos dados e observações
i9a_palavraschave	<i>corpus</i> de referências (e nele, distribuição de fontes, periódico mais citado, meia-vida, autor mais citado)	Protocolo OAI-PMH e base BibTeX/AtoZ
i9b_areas	<i>corpus</i> de referências (e nele, distribuição de fontes, periódico mais citado, meia-vida, autor mais citado)	Protocolo OAI-PMH e base BibTeX/AtoZ
i9c_metodos	<i>corpus</i> de referências (e nele, distribuição de fontes, periódico mais citado, meia-vida, autor mais citado)	Protocolo OAI-PMH e base BibTeX/AtoZ
i10_grafodecitacoes	grafo de citações tendo como nó de origem as áreas de conhecimento dos citantes e, como nó destino, os periódicos das referências citadas	Protocolo OAI-PMH e base BibTeX/AtoZ

Quadro 8 - Indicadores propostos para os testes na AtoZ

Delimitou-se o *corpus* dos documentos citantes apenas para as seções *Artigos* e *Short Papers* do periódico-alvo (doravante denominados apenas de “artigos”), sendo excluídos o *Editorial*, *Expediente* e *Entrevistas*. A janela temporal foi composta por artigos desde o volume 1/número 1 (2011) até o volume 5/número 1 (2016). A mesma separação foi empregada no *corpus* de referências, sendo consideradas apenas aquelas citadas por artigos inseridos nesse recorte.

#### 4.3.1 Indicadores baseados apenas nos metadados

Para facilitar a leitura e possível uso da metodologia de indicadores proposta, eles foram divididos em dois grandes grupos: (i) baseados apenas no metadados; e (ii) cruzados entre metadados e base BibTeX/AtoZ. Essa distinção é importante ao considerar o grau de complexidade e investimento de recursos necessários para a elaboração da fonte de dados BibTeX. Os metadados, ainda que possam ter problemas de preenchimento, já fazem parte de todos os periódicos que são indexados. O uso do BibTeX no processo editorial, por sua vez, não é frequente e exige esforço considerável.

Nesse primeiro tópico são apresentados indicadores originados exclusivamente dos metadados.

#### 4.3.1.1 Triangulação de dados

Uma avaliação inicial, ainda que não possa ser definida como indicador, diz respeito à algumas conformidades básicas no conjunto de metadados do periódico-alvo. Trata-se da contagem de campos essenciais para definição dos artigos e, com isso, verificação de eventuais *gaps* no preenchimento durante o processo editorial (Tabela 1). Excepcionalmente neste teste são considerados documentos de todas as seções do periódico alvo.

Tabela 1 - Avaliação básica de conformidade dos metadados

<b>Metadado avaliado</b>	<b>Qtd.</b>	<b>Valor Esperado</b>
Quantidade de registros recuperados	70	70
Títulos em Português	70	70
Títulos em Espanhol	67	70
Títulos em Inglês	67	70
Resumos em Português	69	70
Resumos em Inglês	65	70
Resumos em Espanhol	67	70
Publisher	70	70
Trabalho com método informado	66	66
Arquivo formato tipo epub+zip	64	66
Arquivo formato tipo pdf	70	70
Arquivo formato tipo html	70	70

Na Tabela 1 foram elencadas algumas métricas para conferência nos metadados, com a sinalização em verde para “conforme” e amarelo para “não conforme”. Como exemplo, nos casos em que faltaram títulos e resumos em determinado artigo, o editor pode efetuar as correções e com isso ampliar a consistências dos metadados, aumentando também as chances de melhor indexação.

Ainda que sejam métricas simples e não se configurem como indicadores, essa triangulação de dados é útil para o corpo editorial avaliar a conformidade do

preenchimento dos metadados do periódico alvo, algo que pode ser julgado por um comitê de avaliação para indexadores. Uma verificação manual também é possível, mas a compilação é mais rápida e prática, permitindo olhar diretamente para as falhas e efetuar correções. No caso do periódico-alvo, no qual foi proposta a publicação de metadados em três idiomas, é fundamental atualizar aqueles artigos em que isso não foi inserido.

#### 4.3.1.2 i1. Autores por trabalho

Em alguns indexadores a quantidade de autores em cada artigo é considerado critério de qualidade. Ainda que isso apresente variação entre diferentes áreas de conhecimento, em geral há um movimento para limitação da quantidade de autores por trabalho (em geral, até cinco autores). O cenário do periódico-alvo pode ser visualizado na Tabela 2.

Tabela 2 - Distribuição de quantidade de autores por volume

		Volume					Total de autores
		v1	v2	v3	v4	v5	
Quantidade de autores por trabalho	1 autor	3	3	1	2	1	10
	2 autores	2	6	5	5		18
	3 autores	3		3	1	1	8
	4 autores	2		3	1	2	8
	5 autores		1		1	1	3
	8 autores			1			1
<b>Total de artigos</b>		10	10	13	10	5	48
~ <i>autores/artigo</i>		2.40	2.10	2.54	3.30	1.20	

No exemplo, o periódico-alvo manteve um valor médio (cerca de 2 autores/artigo) compatível com outros estudos na área de Ciências Sociais Aplicadas como, por exemplo, os levantamentos feitos por Santos (2010), e de Souza, Silva e Araújo (2013). Os casos em que o padrão não foi mantido foi no volume 4, onde 1 artigo teve 5 autores, 1 artigo teve 4 autores e 1 artigo teve três autores, elevando a média para 3.30 autores/artigo. Também foi identificado um

*outlier* no volume 3; pois, apesar da média ser compatível (2.54), havia 1 artigo com 8 autores.

#### 4.3.1.3 i2. Internacionalização

Outro indicador reconhecido é o grau de internacionalização da revista, em geral expresso pela quantidade de autores estrangeiros que publicam artigos no periódico. Para o cálculo foram analisados o currículo de cada autor e extraída a informação do país de origem. Observa-se que todas as contagens são referentes à autores e não aos artigos. Dessa maneira é possível contabilizar trabalhos com cooperação internacional (artigo com autores de mais de um país), conforme Tabela 3.

Tabela 3 - Distribuição de origem geográfica de autores por volume

		Volume					
		v1	v2	v3	v4	v5	Grand Total
<b>País de origem do autor</b>	Brasil	24	14	27	20	8	93
	Equador		6	10	4	6	26
	Espanha					2	2
	Portugal			3		1	4
	<i>Nacional</i>	100%	70%	68%	83%	47%	74%
	<i>Internacional</i>	0%	30%	33%	17%	53%	26%

Neste indicador, a prática desejada pelos indexadores é que os periódicos tenham a maior quantidade de autores estrangeiros publicados em seu *corpus*. Conforme pode ser visto na Tabela 3, a revista já apresenta algum avanço nesse sentido, ao ter alcançado 30% e 33% nos volumes 2 e 3, sofrendo uma redução no v4 e alcançando um valor mais significativo no volume 5 (53%). O ideal é que esse número seja ampliado continuamente.

#### 4.3.1.4 i3. Palavras-chave

Um indicador de emprego e frequência de palavras-chave nos artigos publicados pode dar luz ao corpo editorial da relação que existe entre a proposta



editorial (foco e escopo) e o que de fato está sendo publicado. Ainda que essa avaliação possa estar presente, de maneira tácita, pelo dia-a-dia do trabalho editorial, um indicador é capaz de quantificar e facilitar o entendimento. Os resultados do periódico alvo são vistos na Tabela 4.

Tabela 4 - Frequência decrescente de palavras-chave

Palavra-chave	Ocorrências	Ordem
redes sociais	7	1
jogos educativos	3	2
inteligência artificial	3	3
usabilidade	2	4
twitter	2	5
tecnologias de informação e comunicação (tic)	2	6
sustentabilidade	2	7
sistemas de informação	2	8
produção tecnológica	2	9
processo ensino-aprendizagem	2	10
organização do trabalho	2	11
ontologia	2	12
moda	2	13
mineração de dados	2	14
inovação tecnológica	2	15
dados abertos	2	16
compartilhamento de conhecimento	2	17
ciência da informação	2	18
aprendizagem	2	19
ambiente pessoal de aprendizagem	2	20
<i>Valores suprimidos (...)</i>	...	21-183
wii	1	184

A frequência de palavras-chave apresenta o efeito de cauda longa, totalizando 184 valores únicos, sendo 164 (95%) destes com apenas uma aparição. Conforme apontado anteriormente, foi necessário definir uma regra de consistência (e atualizar a contagem de frequência em relação ao valor original da base), dada a reconhecida proximidade de significado.

A avaliação da Tabela 4 não foi considerada conclusiva, pois há a alta variabilidade e um expressivo alongamento de cauda. Uma possibilidade, no futuro, seria o agrupamento das palavras-chave em temas macro. Ainda assim, as três

palavras-chave mais frequentes (redes sociais, jogos educativos e inteligência artificial) foram testadas no âmbito de variáveis cruzadas (metadados com base BibTeX AtoZ), apresentados nos tópicos “i9. Corpus de referências por variável discriminante”.

#### 4.3.1.5 i4. Grau de formação

A presença de autores com maior grau de formação acadêmica também é alvo de avaliação por parte de agência e indexadores. No cenário mais favorável, ou seja, contar com uma maior proporção de pesquisadores com Doutorado Completo publicando em seus números, confere maior notoriedade ao periódico.

Para criação desse indicador, o currículo de cada autor, em cada artigo foi avaliado. Buscou-se qual deles (autores) possuía maior grau de formação e este representaria determinado artigo. Ex.: em um dado artigo *alpha*, com três autores (um com graduação completa, um com mestrado incompleto e um com doutorado incompleto), foi considerado como grau máximo o “doutorado incompleto”. Neste exemplo, aquele com graduação completa - por óbvio - não teria qualquer título superior à esse. Os resultados estão na Tabela 5.

Tabela 5 - Distribuição de maior grau de formação de autores por artigo de acordo com o volume

Maior grau de formação no artigo	Volume					Total
	v1	v2	v3	v4	v5	
Doutorado Completo	8	2	9	6	3	28
Doutorado Incompleto	1	3	4	1		9
Graduação ou Bacharelado Completos		1		1		2
Mestrado Completo	1	3		2		6
Mestrado Incompleto		1			1	2
Outros					1	1
<b>Total</b>	10	10	13	10	5	48
<i>Trabalhos com pelo menos 1 autor Doutor</i>	80%	20%	69%	60%	60%	58%

Os resultados mostram que o periódico-alvo ainda não conseguiu atingir, de maneira satisfatória, a captação de trabalhos com autores Doutores. O índice do

primeiro volume (80%) é relevante mas não chegou a ser mantido ou recuperado nos volumes seguintes. O mais preocupante foi o caso do volume 2, onde apenas 2 (20%) - entre 10 artigos - tiveram pelo menos um doutor como coautor do trabalho.

A responsabilidade pelo grau de formação não chega a ser algo diretamente do controle/domínio do corpo editorial, uma vez que selecionar trabalhos por esse critério é algo que fere o processo *double blind review* minimamente recomendado. Ainda assim, por ser algo que, indiretamente, reflete uma dimensão “científica” dos estudos publicados, é um critério de observância das agências.

#### 4.3.1.6 i5. Endogenia com base nas instituições de origem

No contexto da publicação científica a endogenia está associada a presença de autores filiados à instituição responsável pela revista, em seu *corpus* de artigos publicados. No cenário ideal, o periódico precisa ter a menor quantidade (ou zero) de trabalhos em que os autores sejam da própria instituição (ou, minimamente, da unidade que edita o periódico), demonstrando um maior alcance e, conseqüentemente, sua importância como veículo de divulgação acadêmico-científica no cenário nacional e internacional.

Na construção do indicador, foi avaliada a filiação autor a autor do *corpus* de artigos da revista, sumarizada por instituição e distribuída ao longo dos volumes, conforme pode ser visto na Tabela 6.

Tabela 6 - Distribuição de afiliação institucional de autores por volume

Instituição de origem	v1	v2	v3	v4	v5	Total	Ordem
Universidade Federal do Paraná - UFPR	7			2	4	13	1
Universidad de las Fuerzas Armadas - ESPE		5	6			11	2
Universidade Federal de Pernambuco - UFPE	4	2	4			10	3
Universidade de São Paulo - USP			7			7	4
Universidade Federal de Santa Catarina - UFSC			2	4		6	5
Universidade Federal de Minas Gerais - UFMG		2			4	6	6
Universidade Paulista - UNIP, São Paulo				5		5	7
Universidade Federal do Pará - UFPA			3	2		5	8
Universidad Casa Grande - UCG		1		3	1	5	9
Universidade Federal de São Carlos - UFSCar	2			2		4	10
<i>Valores suprimidos (...)</i>	<i>... ...</i>	<i>... ...</i>	<i>... ...</i>	<i>... ...</i>	<i>... ...</i>	<i>...</i>	<i>11-49</i>
Associação Brasileira de Engenharia Sanitária e Ambiental - ABES	1					1	49
<b>Total</b>	<b>24</b>	<b>20</b>	<b>40</b>	<b>24</b>	<b>17</b>	<b>125</b>	

<i>% autores do volume que são da Instituição de origem do periódico alvo - UFPR</i>	29%	0%	0%	8%	24%	10%
--	-----	----	----	----	-----	-----

Com base na Tabela 6, verifica-se que foi no volume 1 a maior ocorrência de endogenia, principalmente associado ao lançamento da revista, o qual demandou a solicitação de submissões por convite. Nos outros volumes os índices estão adequados àquilo que é desejável pelas agências indexadoras, com exceção do volume 5, o qual tem - no momento em que se elabora este documento - o número 2 ainda não publicado.

#### 4.3.1.7 i6. Métodos de pesquisa empregados

Os métodos de pesquisa empregados não são um ponto comum no escopo de critérios para avaliação de periódicos mas, dependendo do foco e escopo da revista (ou até de determinada agência indexadora), podem revelar aspectos importantes sobre o que está sendo publicado (Tabela 7). Por exemplo, uma linha editorial que se propõe a publicar pesquisas experimentais pode se perceber distante do foco ao identificar uma grande proporção de revisões de literatura.

Tabela 7 - Frequência decrescente de métodos de proteção empregados

<b>Método empregado</b>	<b>Frequência</b>	<b>Ordem</b>
exploratória	11	1
qualitativa	9	2
estudo de caso	7	3
documental	7	4
quantitativa	6	5
exploratório-descritiva	6	6
descritiva	6	7
quali-quantitativa	5	8
experimental	5	9
empírica	4	10
comparada	3	11
teoria fundamentada em dados	2	12
revisão de literatura	2	13
prototipagem	2	14
etnográfica	2	15
campo	2	16
bibliográfica	2	17
<i>Valores suprimidos (...)</i>	...	18-29
teórica	1	30

Na situação do periódico alvo avaliado, os resultados encontrados na frequência de métodos empregados não estão em desacordo com a proposta de escopo e, também, não se opõe àquilo que é posicionado pelas agências indexadoras.

#### 4.3.1.8 i7. Áreas de conhecimento

Assim como o emprego de palavras-chave e métodos de pesquisa, a associação com a área de conhecimento na qual o artigo se enquadra também pode elucidar o quanto determinado periódico alvo está alinhado à suas definições de foco e escopo. Ainda, esse é um indicador a ser explorado por um agência avaliadora, principalmente se ela for originada de alguma área específica do conhecimento. Conforme posicionado anteriormente, neste indicador exploram-se macroáreas (Tabela 8) e áreas (Tabela 9) .

Tabela 8 - Frequência decrescente de macroáreas de conhecimento

Macroárea	Frequência
ciência da informação	34
ciência da computação	15
educação	9
interdisciplinar	7
comunicação	3
engenharia de produção	2
direito	2
sociologia	1
engenharias	1
desenho industrial	1
ciências da saúde	1
ciência política	1
arquitetura e urbanismo	1

Na avaliação das áreas e macroáreas de conhecimento verificou-se que existe alinhamento entre o foco e escopo da revista e a classificação empregada nos artigos publicados. Um destaque é a educação: tecnologia educacional que, apesar de não tratada diretamente no propósito do periódico alvo, já apresenta relevância na visão geral dos artigos.

Tabela 9 - Frequência decrescente de áreas de conhecimento

Área de conhecimento	Frequência	Ordem
educação: tecnologia educacional	6	1
ciência da informação: processos de disseminação da informação	4	2
ciência da informação: processos da comunicação	4	3
ciência da informação: biblioteconomia	4	4
ciência da informação	4	5
ciência da computação: sistemas de informação	4	6
interdisciplinar	3	7
engenharia/tecnologia/gestão	3	8
ciência da informação: representação da informação	3	9
ciência da informação: métodos quantitativos, bibliometria	3	10
educação: métodos e técnicas de ensino	2	11
ciência da computação: software	2	12
ciência da computação: engenharia de software	2	13
ciência da computação	2	14
Valores suprimidos (...)	...	15-45
técnicas de recuperação da informação	1	46

O indicador de macroáreas também é subsídio para formação de um outro mais complexo (detalhes em i9. “Corpus de referências por variável discriminante”), servindo como ponte entre os metadados e a Base BibTeX/AtoZ. As macroáreas mais frequentes foram empregadas em tais indicadores: ciência da informação (27 artigos), ciência da computação (13 artigos), educação (9 artigos) e interdisciplinar (4 artigos).

#### 4.3.2 Indicadores com metadados e base BibTeX

Os indicadores seguintes são baseados no cruzamento entre os metadados do OJS do periódico-alvo e a base BibTeX/AtoZ. De modo geral, tratam das relações existentes entre literatura citante e literatura citada. Ao contrário dos anteriores, a associação com critérios de agências indexadoras não é tão explícita, cabendo a interpretação de relevância de acordo com os objetivos e necessidade dos gestores do periódico.

Por se tratar do processamento dos dados das listas de referências, esses indicadores estão mais alinhados ao pressupostos da análise bibliométrica - contextualizada na introdução deste documento, entendendo que a literatura utilizada pelos autores de artigos publicados no periódico alvo podem trazer reflexões acerca de sua qualidade.

##### 4.3.2.1 i8. Autores mais referenciados

Para sumarizar os autores mais referenciados nos artigos publicados na revista a base BibTeX foi utilizada num primeiro momento, com a contagem da tag *author*, conforme Tabela 10. Caso uma referência tivesse mais de um autor, todos foram considerados. Para aprimorar a análise dessa lista foram estudadas mais três variáveis, a saber:

- a) **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. Tomando-se um exemplo da Tabela 10, ainda que A. C. Gil tenha sido citado apenas três vezes, foram três os artigos em

que ele apareceu. O exemplo contrário seria Etienne Wenger que, apesar de receber três citações a mais que A. C. Gil, todas estas estão concentradas em apenas um trabalho.

- b) **Quantidade de autocitações.** Outro indicativo já reconhecido no campo de indicadores bibliométricos é a autocitação, quando um autor referênciava um trabalho em que ele mesmo também foi autor/co-autor. Esse é um fator controverso por distorcer os índices e, por essa razão, tais referências costumam ser descartadas em sua composição.
- c) **Homônimos.** Essa variável associa-se às métricas de forma indireta pois implica na consistência da base de dados. Há um reconhecido problema com homonímia na compilação de dados sobre autores, algo que este estudo igualmente revela.

Tabela 10 - Extrato de autores mais referenciados - frequência igual ou superior a 3 (continua)

Autor	Qtd.	Qtd. Artigos	Autocitação	Obs. Homônimos
Instituição - Brasil	20	5	NA	
L. S. Bufrem	7	1	7	
C. Bizer	6	1	0	
Etienne Wenger	5	1	0	
Instituição - Republica del Ecuador	5	1	NA	
L. A. Digiampietri	5	1	5	outra citação como L. Digiampietri
M. P. Bax	5	1	5	
R. Mugnaini	5	2	0	
E. Stal	4	2	0	
Instituição - Brasil, Ministério da Educação [MEC]	4	2	NA	outra citação com sigla entre parênteses
A. C. Gil	3	3	0	outra citação como A. Gil
A. Fujino	3	2	0	
A. V. Roth	3	1	0	
C. Storey	3	1	0	
E. T. Albergaria	3	1	3	
I. Chiavenato	3	3	0	
Instituição - Instituto Brasileiro de Geografia e Estatística [IBGE]	3	1	NA	outra citação com a sigla entre parênteses
Instituição - Society for Human Resource Management	3	1	NA	



Tabela 10 - Extrato de autores mais referenciados - frequência igual ou superior a 3 (conclusão)

<b>Autor</b>	<b>Qtd.</b>	<b>Qtd. Artigos</b>	<b>Autocitação</b>	<b>Obs. Homônimos</b>
J. C. C. Terra	3	2	0	
J. Lee	3	2	0	dois autores diferentes: Jungwoo Lee, Johnny Chung Lee
J. Leskovec	3	1	0	
L. Prusak	3	2	0	
M. Castells	3	3	0	
M. L. de A. Campos	3	1	0	
M. N. González de Gómez	3	2	0	
M. O. E. de Oliveira	3	1	0	
M. Santos	3	2	0	três autores diferentes: Milton Santos, Marcelo Rodrigues dos Santos, Marcio de Miranda Santos
R. O. Prates	3	1	3	
T. Berners-Lee	3	1	0	

Análises que envolvem o nome dos autores, nas condições de homonímia, costumam apresentar mais falhas dada a dificuldade de separá-los com um identificador único. Para preservar o dado original e promover esta discussão optou-se por não corrigir os possíveis casos de homonímia, mantendo-os evidentes na tabela de resultados.

Na frequência de aparição dos autores, nas referências citadas nos artigos, verifica-se o efeito de cauda longa, com um total de 1.685 nomes diferentes. Pela limitação de tempo para a análise, apenas os 29 primeiros (aqueles com frequência maior ou igual a três) foram considerados. As variáveis autocitação e quantidade de artigos e potenciais homônimos foram computadas com busca manual, em função do grau de complexidade que envolveria a implantação semi-automática desse cálculo.

#### 4.3.2.2 i9. Corpus de referências por variável discriminante

O indicador i9 deriva-se de três grupos: palavras-chave, macroáreas de conhecimento e métodos empregados. Os três levam o nome das variáveis

discriminantes que são empregadas, em momento diferentes, para o filtro da base BibTeX/AtoZ, criando diferentes *corpus* que são submetidos a uma mesma estrutura de indicadores.

Como exemplo, verificou-se que palavra-chave mais frequente é *redes sociais*. Ela foi empregada em 7 trabalhos. Nos dados que serão apresentados no i9, grupo palavras-chave, foram consideradas todas as referências que foram empregadas pelo menos uma vez dentre os 7 trabalhos. A mesma lógica se aplica às outras duas palavras-chave mais frequentes (jogos educativos, inteligência artificial) e às outras variáveis discriminantes.

A estrutura comum de análise é formada com alguns estatísticas bibliométricas que, apesar da limitação do escopo do *corpus* - considerado pequeno, principalmente por se tratar de apenas um periódico alvo e com cinco volumes - oferece *insights* para o corpo editorial e para possíveis estudos futuros. Nela foram considerados:

- a) **Quantidade de Referências.** No *corpus* referente ao conjunto de referências empregadas em artigos que utilizaram tal variável discriminantes, qual a quantidade de entradas.
- b) **Artigos Citantes.** Quantos foram os artigos que utilizaram a variável discriminante.
- c) **Proporção Referências/Artigos citantes.** Divisão entre os dois anteriores, sugerindo uma compreensão de quantas referências em média aquele *corpus* apresenta por artigo.
- d) **Citing Half-Life.** No contexto deste estudo visa identificar a obsolência das referências citadas no periódico alvo, a partir de variáveis discriminantes empregadas nos artigos.  
**Citing Half-Life apenas tipo artigo.** Mesma métrica anterior mas considerando apenas as referências do tipo artigo.
- e) **Referências sem ano.** Como o cálculo do Citing Half-Life depende da relação entre o ano que foi citado e o ano que foi publicado, foram informadas as quantidades de referências que, por não terem a *tag* ano, foram descartadas.
- f) **Distribuição de tipos de fonte.** Para cada *corpus*, qual o percentual de tipo de fonte empregado nas referências.

- g) **Distribuição de autores mais citados.** Para cada *corpus*, qual o percentual de autores mais citados.
- h) **Distribuição de periódicos mais citados.** Para cada *corpus*, qual o percentual de tipo de fonte empregado nas referências.

Dada a extensão dos resultados, foram apresentados apenas alguns excertos que facilitam o entendimento das possibilidades em se gerar esses indicadores. Na Tabela 11 estão os indicadores das palavras-chave *redes sociais* e *jogos educativos*.

Tabela 11 - Amostra de indicadores dos *corpora* de referências das palavras-chave redes sociais e inteligência artificial

redes sociais			inteligência artificial		
Referências	206		Referências	58	
Artigos Citantes	7		Artigos Citantes	3	
Proporção Referências / Artigo Citante	29.4		Proporção Referências / Artigo Citante	19.3	
Citing Half-Life - todos tipos	5		Citing Half-Life - todos tipos	5	
Citing Half-Life - apenas article	5		Citing Half-Life - apenas article	5	
Referências sem tag <year>	4		Referências sem tag <year>	1	
<b>Distribuição de fontes</b>			<b>Distribuição de fontes</b>		
Tipo	Qtd	%	Tipo	Qtd	%
article	71	34.47%	article	14	24.14%
book	38	18.45%	book	24	41.38%
inbook	14	6.80%	inbook	0	0.00%
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...
<b>Distribuição de Autores</b>			<b>Distribuição de Autores</b>		
person	freq	qtd_artigos	person	freq	qtd_artigos
L. A. Digiampietri	5	1	A. Illanas	2	1
Instituição - Society for Human Resource Management	3	1	E. Rich	2	2
R. Mugnaini	3	2	J. Lee	2	2
B. L. Berkelaar	2	1	K. Knight	2	2
C. C. Trucolo	2	1	M. Buckland	2	2
Denis Gillet	2	1	R. Bello	2	2
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...
<b>Distribuição de periódicos citados</b>			<b>Distribuição de periódicos citados</b>		
Título	Qtd		Título	Qtd	
Educação, Cultura e Comunicação	2		AtoZ: novas práticas em informação e conhecimento	1	
Education in the Knowledge Society	2		Computers & Education	1	
EDUTEC. Revista Electrónica de Tecnología Educativa	2		Educational Technology & Society	1	
Valores suprimidos (...)	...		Valores suprimidos (...)	...	

Os *corpora* de referências que têm como base a variável discriminante macroárea de conhecimento - excerto de *ciência da informação* e *ciência da computação* - são apresentados na Tabela 13.

Tabela 13 - Amostra de indicadores dos *corpora* de referências das macroáreas de conhecimento *ciência da informação* e *ciência da computação* (continua)

ciência da informação			ciência da computação		
Referências	580		Referências	231	
Artigos Citantes	24		Artigos Citantes	13	
Proporção Referências / Artigo Citante	24.2		Proporção Referências / Artigo Citante	17.8	
Citing Half-Life - todos tipos	7		Citing Half-Life - todos tipos	5	
Citing Half-Life - apenas article	8		Citing Half-Life - apenas article	6	
Referências sem tag <year>	11		Referências sem tag <year>	8	
<b>Distribuição de fontes</b>			<b>Distribuição de fontes</b>		
Tipo	Qtd	%	Tipo	Qtd	%
article	225	38.79%	article	63	27.27%
book	156	26.90%	book	73	31.60%
inbook	26	4.48%	inbook	5	2.16%
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...
<b>Distribuição de Autores</b>			<b>Distribuição de Autores</b>		
person	freq	qtd_artigos	person	freq	qtd_artigos
Instituição - Brasil	17	5	C. Bizer	6	1
L. S. Bufrem	7	1	L. A. Digiampietri	5	1
Etienne Wenger	5	1	R. Mugnaini	3	2
L. A. Digiampietri	5	1	T. Berners-Lee	3	1
M. P. Bax	5	1	A. Illanas	2	1
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...

Tabela 13 - Amostra de indicadores dos *corpora* de referências das macroáreas de conhecimento *ciência da informação* e *ciência da computação* (conclusão)

ciência da informação			ciência da computação		
Distribuição de periódicos citados			Distribuição de periódicos citados		
Título	Qtd		Título	Qtd	
Diário Oficial [da] República Federativa do Brasil	14		Archival Science	2	
Ciência da Informação	11		Revista Latinoamericana de Tecnología Educativa	2	
DataGramZero	9		Accounting Research Journal	1	
Perspectivas em Ciência da Informação	7		AI Magazine	1	
Encontros Bibli	6		Applied Ecology	1	
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...

O último grupo de indicadores na estrutura comum é o de métodos empregado na pesquisa, especificamente um excerto de resultados de *exploratória* e *qualitativa*, Tabela 14.

Tabela 14 - Amostra de indicadores dos *corpora* de referências dos métodos *pesquisa exploratória* e *pesquisa qualitativa* (continua)

exploratória			qualitativa		
Referências	231		Referências	241	
Artigos Citantes	11		Artigos Citantes	9	
Proporção Referências / Artigo Citante	21.0		Proporção Referências / Artigo Citante	26.8	
Citing Half-Life - todos tipos	6		Citing Half-Life - todos tipos	8	
Citing Half-Life - apenas article	6.5		Citing Half-Life - apenas article	8	
Referências sem tag <year>	8		Referências sem tag <year>	3	
Distribuição de fontes			Distribuição de fontes		
Tipo	Qtd	%	Tipo	Qtd	%
article	95	41.13%	article	94	39.00%
book	59	25.54%	book	70	29.05%
inbook	8	3.46%	inbook	13	5.39%
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...

Tabela 14 - Amostra de indicadores dos *corpora* de referências dos métodos *pesquisa exploratória* e *pesquisa qualitativa* (conclusão)

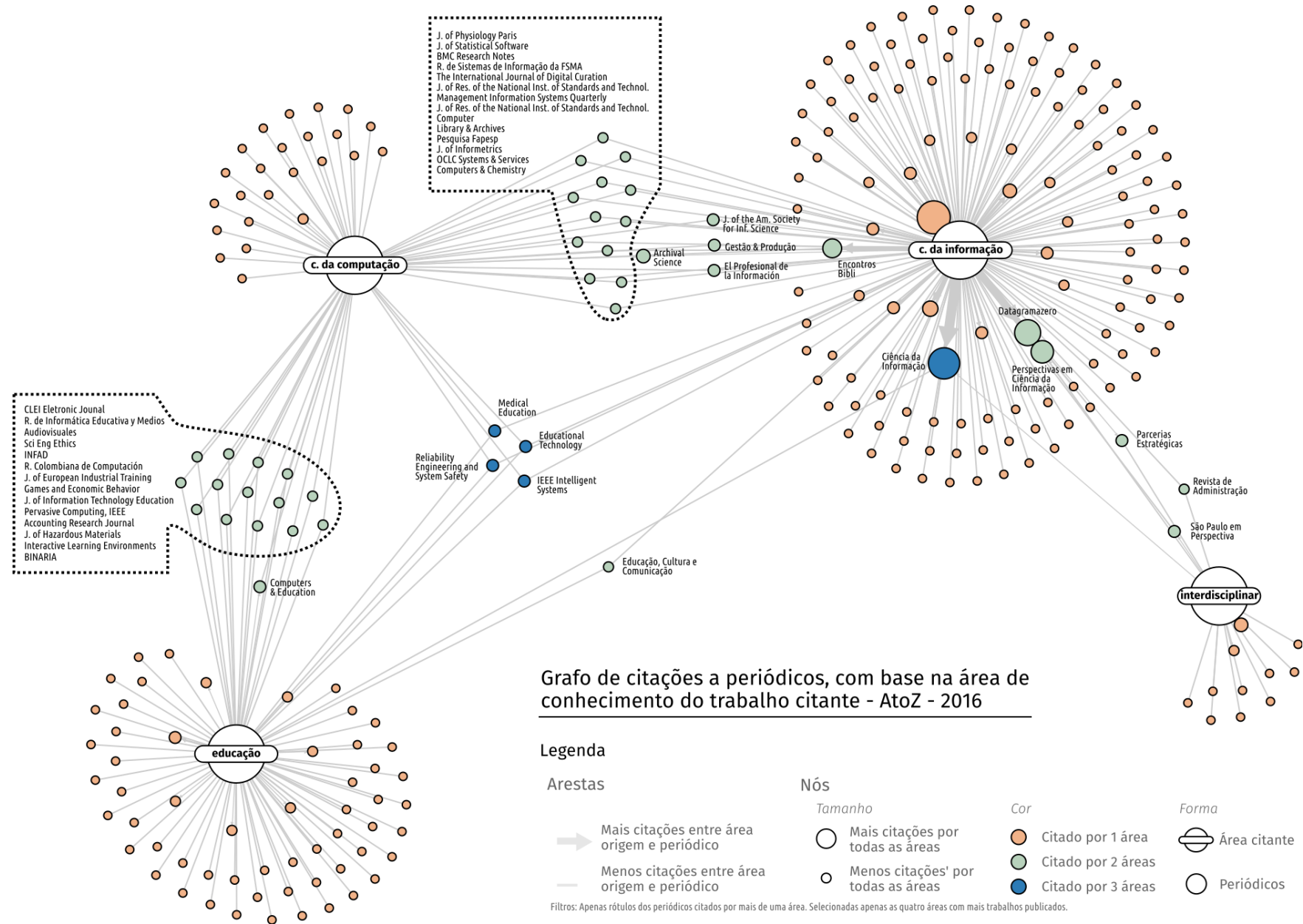
exploratória			qualitativa		
Distribuição de Autores			Distribuição de Autores		
person	freq	qtd_artigos	person	freq	qtd_artigos
Etienne Wenger	5	1	Instituição - Republica del Ecuador	5	1
J. Leskovec	3	1	Instituição - Brasil	4	5
A. Illanas	2	1	A. V. Roth	3	1
B. A. Huberman	2	1	C. Storey	3	1
Denis Gillet	2	1	E. Stal	3	2
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...
Distribuição de periódicos citados			Distribuição de periódicos citados		
Título	Qtd		Título	Qtd	
Ciência da Informação	5		Journal of Services Marketing	5	
Perspectivas em Ciência da Informação	4		Ciência da Informação	4	
DataGramZero	3		Government Information Quarterly	4	
Encontros Bibli:	3		European Journal of Marketing	3	
Knowledge Management Research & Practice	3		Journal of Product Innovation Management	3	
Valores suprimidos (...)	...	...	Valores suprimidos (...)	...	...

#### 4.3.2.3 i10. Grafo de citações - áreas de conhecimento e periódicos

O último indicador desenvolvido tratou das relações existentes entre macroáreas de conhecimento e os periódicos referenciados. De maneira similar ao i9, foram selecionadas as referências citadas em artigos cujas macroáreas de conhecimento nos metadados foram: ciência da informação, ciência da computação, educação e interdisciplinar.

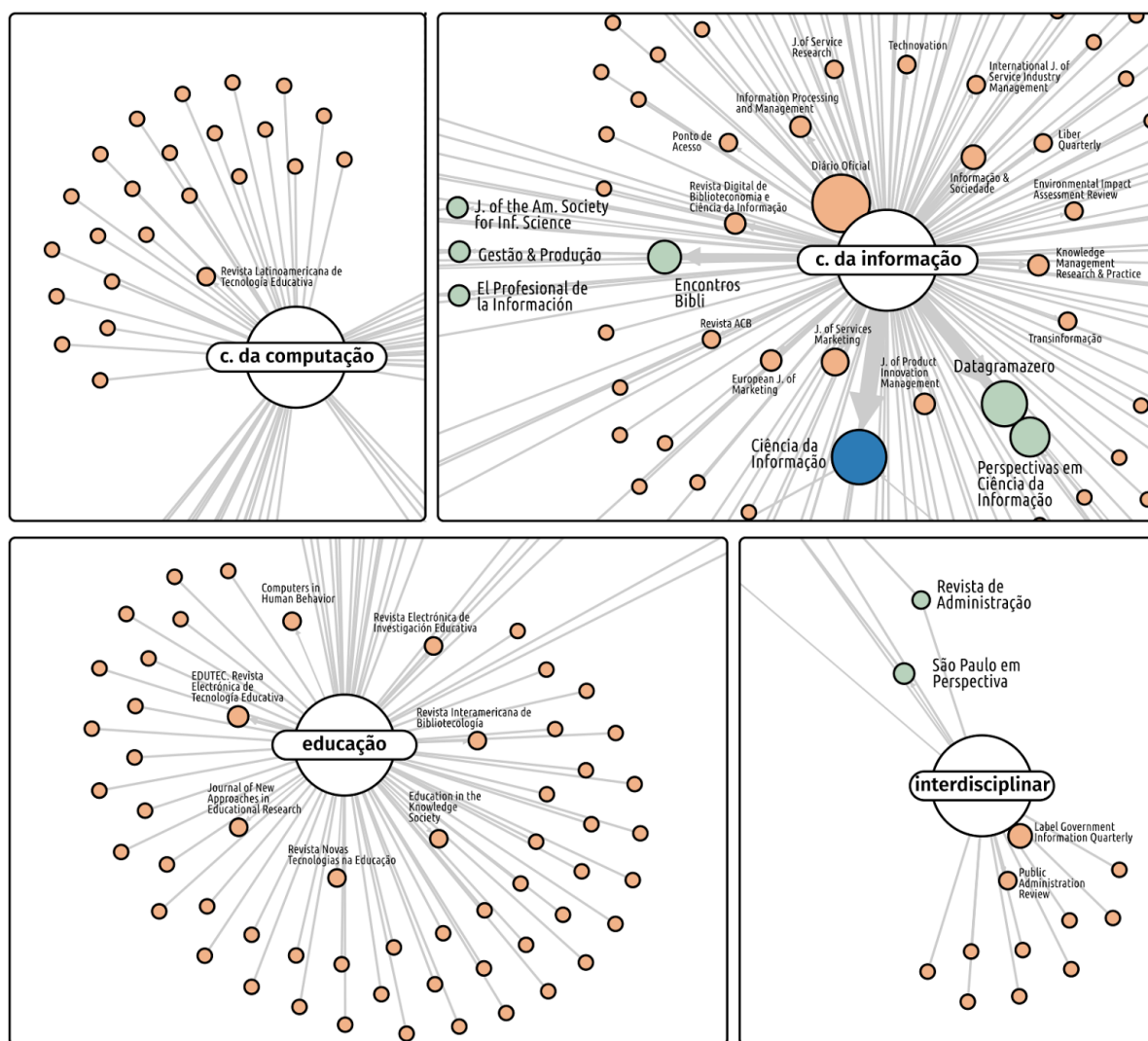
Em sua construção, não foram consideradas as referências em si - por uma questão de efeito cauda longa - mas sim uma seleção dos periódicos citados. Isso conferiu maior densidade às relações, reduzindo os efeitos de ter um *corpus* reduzido. Para representar esse indicador foi utilizado um grafo de citações. Nele, as arestas são as ligações entre macroárea de conhecimento e periódico citado.





Grafo 1 - Visão Geral das relações entre macroáreas de conhecimento e periódicos citados

Conforme indicado na legenda do Grafo 1, foram reproduzidas as seguintes variáveis no desenho: classe citado e citante (formato do desenho), quantidade de citações por todas as áreas (tamanho do nó no caso de periódicos), quantidade de diferentes áreas citando (cor do nó). Para melhor a visualização da composição de cada macroárea, foi construído um mosaico do Grafo com aproximação (Grafo 2).



Grafo 2 - Mosaico com aproximação do núcleo de cada uma das macroáreas

Ainda que os grafos não representem associações explícitas à critérios que o periódico-alvo possa vir a atender na exigência de uma agência indexadora, eles permitem explorar a posição que o alvo ocupa dentro de seu ambiente. A

representação corrobora a conformação das macroáreas de conhecimento e de seus periódicos nucleares, tal como a distribuição em cauda longa em alguns casos.

## 5 CONSIDERAÇÕES

Este apêndice é parte integrante do Relatório final referente ao Edital de Apoio à Editoração e Publicação de Periódicos Científicos - 2016 (UFPR/PRPPG/SIBI) no contexto da proposta apresentada pelo grupo gestor do periódico acadêmico-científico AtoZ: novas práticas em informação e conhecimento e sob a responsabilidade do Grupo de Pesquisa UFPR/CNPq Metodologias em Gestão da Informação. O grupo gestor da AtoZ agradece, novamente, ao Conselho de Editores e à Coordenadoria de Pesquisa da PRPPG pelo apoio e recursos oferecidos, sem os quais a implantação da base BibTeX/AtoZ não teria sido possível, assim como as demais etapas de geração dos indicadores, testes e resultados.

Neste particular, reforça-se que estudos de natureza bibliométrica - e outros de caráter estatístico - apresentam resultados diretamente proporcionais à confiabilidade dos dados a serem manipulados. Igualmente, universos mais robustos de dados (decorrentes de alimentação contínua e progressiva) aumentam a capacidade de gestores em projetar metas, estabelecer objetivos e tomar decisões frente às análises derivadas. Editores de periódicos que desejem produzir indicadores internos - e habilitar tecnicamente as revistas sob suas responsabilidades à indicadores externos - devem estar atentos a estes quesitos de qualidade de dados e a garantia das fontes destes (no caso desta proposta, uma base BibTeX própria e um conjunto de metadados de todos os volumes e números da revista).

Obviamente, 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. A máxima aderência pretendida, na proposta, com as exigências de agências indexadoras (ainda que limitada ao conjunto de dados atualmente disponível na AtoZ) acompanha um objetivo colateral de qualquer periódico acadêmico-científico em sua trajetória, ou seja, reduzir/racionalizar recursos físicos, financeiros e humanos ao quando de sua incorporação à respeitadas índices de citação.

Para a equipe gestora da AtoZ, finalizar as atividades e cumprir com os objetivos do Edital 2016 - Apoio à Editoração e Publicação de Periódicos Científicos trouxe significativo valor agregado aos seus processos de gestão editorial, apontando caminhos para ajustes e estabelecimento de estratégias a serem definidas para curto, médio e longo prazo.

## REFERÊNCIAS

BAR-ILAN, J. Which h-index? A comparison of WoS, Scopus and Google Scholar. **Scientometrics**, v. 74, n. 2, p. 257-271, Feb. 2008. doi: 10.1007/s11192-008-0216-y.

BORGMAN, C. L.; FURNER, J. Scholarly communication and bibliometrics. **Annual Review of Information Science and Technology**, v. 36, n. 1, p. 2-72, 2002. doi: 10.1002/aris.1440360102

BRAUN, T.; GLÄNZEL, W.; SCHUBERT, A. A Hirsch-type index for journals. **Scientometrics**, v. 69, n. 1, p. 169-173, Oct. 2006. doi: 10.1007/s11192-006-0147-4

FERREIRA, H.; CASSIOLATO, M.; GONZALEZ, R. **Uma experiência de desenvolvimento metodológico para avaliação de programas: o modelo lógico do programa segundo tempo**. Brasília, Instituto de Pesquisa Econômica Aplicada, 2009. (Texto para Discussão, 1369). Disponível em: <[http://repositorio.ipea.gov.br/bitstream/11058/1545/1/TD\\_1369.pdf](http://repositorio.ipea.gov.br/bitstream/11058/1545/1/TD_1369.pdf)>. Acesso em: 16 ago. 2013.

GARFIELD, E. Citation indexes for science: a new dimension in documentation through association of ideas. **Science**, v. 122, n. 3159, p. 108-111, 15 July 1955. doi: 10.1126/science.122.3159.108

GARFIELD, E. Science Citation Index: a new dimension in indexing. **Science**, v. 144, n. 3619, p. 649-654, 8 May 1964. doi: 10.1126/science.144.3619.649

GARFIELD, E. Citation analysis as a tool in journal evaluation: journals can be ranked by frequency and impact of citations for science policy studies. **Science**, v. 178, n. 4060, p. 471-479, 3 Nov. 1972. doi: 10.1126/science.178.4060.471

GONZÁLEZ-PEREIRA, B.; GUERRERO-BOTE, V. P.; MOYA-ANEGÓN, F. A new approach to the metric of journals' scientific prestige: The SJR indicator. **Journal of Informetrics**, v. 4, n. 3, p. 379-391, Jul. 2010. doi: 10.1016/j.joi.2010.03.002

HARZING, A. W.; WAL, R. van der. Google Scholar as a new source for citation analysis?. **Ethics in Science and Environmental Politics**, v. 8, n. 1, p. 61-73, 2008.

HIRSCH, J. E. An index to quantify an individual's scientific research output. **Proceedings of the National Academy of Sciences**, v. 102, n. 46, p. 16569-16572, 15 nov. 2005. doi: 10.1073/pnas.0507655102

- HU, Z.; CHEN, C.; LIU, Z. Where are citations located in the body of scientific articles? A study of the distributions of citation locations. **Journal of Informetrics**, v. 7, n. 4, p. 887–896, Oct. 2013. doi: 10.1016/j.joi.2013.08.005
- HUNG, J. Trends of e-learning research from 2000 to 2008: use of text mining and bibliometrics. **British Journal of Educational Technology**, v. 43, n. 1, p 5 - 16, Jan. 2012. Disponível em: <<http://dx.doi.org/10.1111/j.1467-8535.2010.01144.x>>. Acesso em: 09 dez. 2016.
- KESHAVA; GIREESH, A. G.; GOWDA, M. P. ACM transaction on information systems (1989–2006): a bibliometric study. **Information Studies**, v. 14, n. 4, p. 223–234, 2008.
- KOSTOFF, R. N.; DEL RIO, J. A.; HUMENIK, J. A.; GARCÍA, E. O.; RAMIREZ, A. M. Citation mining: Integrating text mining and bibliometrics for research user profiling. **J. Am. Soc. Inf. Sci.**, v. 52, p. 1148–1156, 2001.
- MEIJER, E. **The apacite package**: citation and reference list with LaTeX and BibTeX according to the rules of the American Psychological Association. v. 6.03, 2013. Disponível em: <<ftp://192.115.211.70/ctan/biblio/bibtex/contrib/apacite/apacite.pdf>>. Acesso em: 20 set. 2016.
- PRICE, D. J. de S. **Little science, big science and beyond**. New York: Columbia University Press, 1986.
- REVISTABW. Flat File (Arquivo Simples). **Revista Brasileira de Web: Tecnologia**. Disponível em <<http://www.revistabw.com.br/revistabw/flat-file/>>. Criado em: 02/01/2013. Última atualização: 24/07/2015. Acesso em: 01 dez. 2016
- SANTOS, S. M. dos. **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: 09 dez. 2016.
- SOUZA, F. J. V. de; SILVA, M. C. da; 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**, Salvador, v. 7, n. 2, p. 22-37, maio/ago. 2013. Disponível em: <<https://portalseer.ufba.br/index.php/rcontabilidade/article/view/6268/6450>>. Acesso em: 09 dez. 2016.

THELWALL, M. Bibliometrics to webometrics. **Journal of Information Science**, v. 34, n. 4, p. 605–621, Aug. 2008. doi: 10.1177/0165551507087238

THOMSON REUTERS. **Thomson Reuters Announces Definitive Agreement to Sell its Intellectual Property & Science Business to Onex and Baring Asia for \$3.55 billion**. 11 jul. 2016. Disponível em: <<http://thomsonreuters.com/en/press-releases/2016/july/thomson-reuters-announces-definitive-agreement-to-sell-its-intellectual-property-science-business.html>>. Acesso em: out. 2016.

ZEARIN. **BibTeXML**. GitHub repository: 2016. Disponível em: <<https://github.com/Zearin/BibTeXML>>. Acesso em: 30 ago. 2016.