

Nova fórmula revisada para o cálculo do fator de impacto WEB (FIW)

Nadia Vanti*

José Alfredo F. Costa**

Ilaydiany Cristina Oliveira da Silva***

Resumo Apresenta um histórico da medição do impacto da produção científica e mostra a necessidade de estender tal avaliação à web. Descreve como vem sendo calculado o fator de impacto em sítios web e propõe uma mudança na fórmula original, com a inclusão de um logaritmo natural ao seu denominador, a fim de se obter resultados mais acurados. A nova fórmula é aplicada aos sítios web das universidades federais da região sudeste do Brasil, obtendo-se resultados mais acurados.

Palavras-chave Impacto da produção científica, Indicadores webométricos, Fator de Impacto Web, Fórmula do FIW, Logaritmo natural.

New revised formula for calculating Web Impact Factor

Abstract Displays a history of measuring the impact of scientific articles and shows the need to extend this assessment to the web. Describes how the impact factor on websites has been calculated and proposes a change in the original formula, with the inclusion of a natural logarithm in its denominator in order to obtain more accurate results. The new formula is applied to websites of federal universities in Southeastern Brazil.

Keywords Impact of scientific production, Webometric indicators, Web Impact Factor Formula FIW, Natural logarithm.

* Doutora em Comunicação e Informação pela Universidade Federal do Rio Grande do Sul. Professora Adjunta do Departamento de Ciência da Informação, Centro de Ciências Sociais Aplicadas, Universidade Federal do Rio Grande do Norte. Endereço profissional: Campus Universitário – UFRN – CCSA – DECIN - sala 07, térreo - Av. Salgado Filho, s/n, Lagoa Nova - 59072-970 - Natal, RN. Telefone: (84) 3215-3515. E-mail: nvanti@ufrnet.br

** Doutor. em Engenharia Elétrica e de Computação pela Universidade Estadual de Campinas – SP. Professor Associado do Departamento de Engenharia Elétrica, Centro de Tecnologia, Universidade Federal do Rio Grande do Norte. Endereço profissional: Campus Universitário - UFRN - CT – DEE - Av. Salgado Filho, s/n, Lagoa Nova - 59072-970 - Natal, RN. Telefone: (84) 3215-3882. E-mail: alfredo@ufrnet.br

*** Mestra em Engenharia de Produção pela Universidade Federal do Rio Grande do Norte. Professora Substituta do Departamento de Ciência da Informação, Centro de Ciências Sociais Aplicadas, Universidade Federal do Rio Grande do Norte. Endereço profissional: Campus Universitário – UFRN – CCSA – DECIN - sala 07, térreo - Av. Salgado Filho, s/n. - Lagoa Nova - CEP 59072-970 - Natal, RN. Telefone: (84) 3215-3515. E-mail: ilaydiany18@hotmail.com

Introdução¹

Ao longo do tempo, a produção científica foi alvo de avaliações por parte de pesquisadores e agências de fomento com o interesse de analisar a produtividade de autores, áreas do conhecimento e entidades acadêmicas, tendo em vista a elaboração de políticas voltadas para a gestão da ciência e da tecnologia. Para tais análises, foram criados indicadores que possibilitaram a padronização dos procedimentos, servindo de parâmetro para diferentes avaliações.

A ciência utiliza-se de indicadores para medir as ações sistemáticas que têm relação com a produção, difusão, transferência e aplicação dos conhecimentos científicos e tecnológicos, que podem ser chamados de indicadores da ciência.

A fim de descobrir o impacto das publicações periódicas na comunidade científica, foi proposto o Fator de Impacto de Revistas. (GARFIELD, 1955). O Fator de Impacto é o indicador utilizado para calcular o número médio de citações recebidas por uma revista científica e é obtido com base na relação entre o número de vezes que a revista foi citada e o número de artigos que ela publicou num determinado período de tempo (usualmente dois anos). (JOURNAL, 1998 apud STRELL, 2005).

Com o surgimento de novas mídias, como a Internet e a web, houve a necessidade da criação de um índice análogo para a avaliação da produção científica disponível nestes ambientes. Foi, então, que surgiu o Fator de Impacto Web (FIW) (RODRÍGUEZ-GAIRÍN, 1997; INGWERSEN, 1998), que vem sendo utilizado para medir o grau de atratividade e influência destes novos recursos de informação. No entanto, a fórmula criada para o cálculo deste indicador não tem trazido resultados satisfatórios.

A proposta deste trabalho, portanto, é entender por que motivo ao se utilizar a fórmula criada para calcular o FIW dos sítios web não se obtiveram resultados satisfatórios. E, a partir deste entendimento, avançar, buscando uma maneira mais adequada de calcular tal indicador.

O fator de impacto de revistas

O primeiro autor a propor a medição do impacto da produção científica foi Eugene Garfield em 1955, no seu conhecido artigo *Citation Indexes for Science*, segundo o qual o número de citações recebidas por uma publicação mostra a sua repercussão na literatura científica. Garfield compreendeu que os artigos de revisão na literatura periódica são fortemente dependentes das citações bibliográficas, induzindo o leitor a buscar a fonte original motivado por uma ideia notável ou um conceito. Ao capturar as citações, Garfield acreditava que o pesquisador poderia ter uma visão imediata da abordagem utilizada por outro cientista ao apoiar uma ideia ou uma metodologia, tendo como base as fontes que o autor consultou e citou. (THOMSON REUTERS, 2010).

¹ Os autores agradecem as sugestões feitas pelo professor e pesquisador Piotr Trzesniak com relação à introdução de um logaritmo natural ao invés de um logaritmo decimal à fórmula original do FIW, o que permitiu um maior desenvolvimento da pesquisa e a ratificação da necessidade de modificar a forma de cálculo deste indicador.

Motivado por estas ideias, Garfield lançou as bases para a criação do ISI (*Institute for Scientific Information*) em 1958, na Filadélfia. A partir de 1960 desenvolveu novas ferramentas para facilitar o acesso às informações científicas de diferentes campos do conhecimento: *Current Contents*, *Social Sciences Citation Index* (SSCI) e *Arts and Humanities Citation Index* (A&HCI). As bases de dados do *Institute for Scientific Information* (ISI/Thomson) têm sido importantes fontes de informação para a avaliação do impacto de publicações científicas. A lista do fator de impacto das revistas (FI) é publicada anualmente no *Jornal Citation Reports* (JCR) pela *Thomson Reuters* e pode ser consultada por meio do sítio do próprio *ISI Web of Knowledge*, disponibilizado no Brasil pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Contudo, com o surgimento de um novo paradigma informacional, denominado Sociedade do Conhecimento, surgiram diferentes recursos gerados pelas tecnologias da informação e comunicação, como a internet e a web. Estes novos ambientes, por suas facilidades de acesso e divulgação da informação, acabaram por se tornar os canais de comunicação científica mais utilizados pelos produtores e consumidores deste tipo de informação, fazendo com que ela migrasse e continue migrando rapidamente do suporte tradicional para a web.

O fator de impacto WEB

Houve, a partir desta nova realidade, portanto, a necessidade de se estender a avaliação das publicações também para este novo ambiente. Era preciso então criar um indicador análogo ao sugerido por Garfield (1955) para calcular, agora, o fator de impacto dos recursos eletrônicos. Foi então que Rodríguez-Gairín (1997) e Ingwersen (1998) sugeriram a utilização de um indicador para calcular a média de citações a um sítio web em seus estudos e o chamaram de Fator de Impacto Web (FIW). (AGUILLO, 2005). Tal medida, segundo Ingwersen (1998), é a soma do número de páginas web internas e externas de um sítio dividido pelo número de páginas deste próprio sítio em um determinado momento. Ingwersen sugeriu este cálculo como um meio para medir e comparar a atratividade e influência de distintos espaços web.

De acordo com Smith (1999), o FIW de um espaço na web é o número de páginas que apontam para um espaço web, dividido pelo número de páginas deste espaço. Para este autor, podem ser calculados três distintos tipos de Fator de Impacto Web: o FIW externo, aquele que reflete o número de páginas que apresentam um link de fora do espaço web que está sendo medido, o FIW interno, refletindo os links que partem de dentro do espaço web analisado, e um FIW geral, que combina links externos e externos de um mesmo espaço web. Cabe ressaltar que para este mesmo autor o FIW externo seja, provavelmente, o mais significativo deles.

Thelwall (2002), dentro desta mesma perspectiva, optou por deixar de lado o FIW interno nas suas medições por considerá-lo como análogo a uma auto-citação na produção científica e sugeriu uma fórmula que contabilizava apenas os links externos e os dividia pelo número total de páginas de um sítio web:

FIW = $\frac{\text{N}^\circ \text{ de páginas externas que apontam para determinado sítio}}{\text{N}^\circ \text{ de páginas que integram o sítio apontado}}$

Assim, o indicador sugerido por Rodríguez-Gairín (1997) e Ingwersen (1998) e depurado por Thelwall (2002) na sua fórmula, passou a ser utilizado como um dos valores para determinar a posição de determinados sítios no ranking da web. No entanto, foi Smith (1999) quem estabeleceu os critérios para a interpretação dos números obtidos por meio deste indicador. Segundo ele, sítios considerados com alto FIW são aqueles que possuem índices significativamente maiores do que 1, sítios com médio FIW são os que possuem índices em torno de 1 e sítios com baixo FIW são aqueles que apresentam índices significativamente menores do que 1.

Várias pesquisas foram feitas empregando a mesma metodologia para calcular o Fator de Impacto de sítios web e os números que se obtiveram foram sempre desanimadores. O que se tem percebido é que, diferentemente do que ocorre com os valores apresentados para compor o cálculo do Fator de Impacto de Revistas (FI), onde normalmente o número de citações é muito superior ao número de artigos publicados em uma revista, daí resultando um valor de FI bastante significativo, os *inlinks* externos ou as “citações”² apresentam números muito menores do que o total de páginas de um sítio web, o que leva, nestes casos, a se obter um valor de FIW irrelevante ou até nulo.

Cabe destacar, também, que Thelwall (2001), Kousha (2005) e Vanti (2007) já apresentaram questionamentos acerca da funcionalidade deste indicador no âmbito web. Smith (1999), inclusive, afirma que para grandes organizações, como universidades ou instituições de pesquisa, o FIW parece ser uma medida útil da influência global do espaço web, mas para espaços menores, como revistas eletrônicas, o uso do FIW é menos confiável.

Metodologia

Tomando-se como base a fórmula original idealizada por Rodríguez-Gairín (1997) e Ingwersen (1998) e depurada por Thelwall (2002) buscou-se adaptá-la para que cumprisse o seu objetivo: apontar quais são os sítios que recebem o maior número de *links* externos em relação ao seu tamanho para, assim, oferecer uma medida da utilidade e do impacto do sítio analisado.

Na busca de elementos para entender a distorção dos resultados obtidos com o emprego da fórmula proposta pelos autores citados, observou-se que o problema do cálculo do FIW estava exatamente no número excessivamente grande de páginas dos sítios web com relação ao número de “citações” recebidas (visibilidade), o que é comum a todos os recursos disponíveis neste ambiente (veja-se mais adiante exemplo desta situação nas Tabelas 3 e Tabela 4), chegando-se à conclusão de que para se conseguir resultados mais acurados era necessário modificar a fórmula de aplicação do FIW. Foi quando se testou incluir um logaritmo ao denominador da sua fórmula, junto ao tamanho do sítio, potencializando-o, de modo a que o indicador relacional (FIW) melhor cumprisse o seu objetivo. Desta forma, foi possível observar que utilizando este recurso obtêm-se resultados mais próximos à realidade, pois a transformação LOG reduz a faixa dinâmica dos valores, diminuindo a diferença entre eles. Acredita-se, ademais, que o logaritmo natural (ln) é o mais adequado neste caso pelo fato de ele ser comumente usado em cálculos

² Tradução para o português do neologismo “sitations” criado por Rousseau (1997) para designar as citações (links) entre os sítios *web*.

matemáticos que apresentam números que crescem de forma exponencial, de modo a suprir a real necessidade aqui apresentada.

A fórmula ficou assim:

$$\text{FIW} = \frac{\text{Visibilidade}}{\ln(\text{Tamanho do Sítio})}$$

Para testar a nova maneira de calcular o FIW, foram aplicadas as duas fórmulas aos sítios web das universidades federais da região sudeste do Brasil. Foram inicialmente coletados os dados referentes ao tamanho e à visibilidade de cada sítio por meio do motor de busca Yahoo no mês de outubro de 2009. Com estes dados em mãos, aplicou-se então a fórmula sugerida por Thelwall (2002) - FIW - e depois a fórmula alternativa criada pelos autores deste trabalho - representada pelo acrônimo FIWIn - para, assim, poder comparar os resultados obtidos.

Resultados

Os resultados obtidos para o FIW das universidades federais da região sudeste do Brasil a partir das duas fórmulas supracitadas, bem como os indicadores tamanho e visibilidade, também envolvidos no cálculo, podem ser visualizados nas tabelas a seguir.

Tabela 1 – Ranking das universidades federais da região sudeste do Brasil com a aplicação do FIW₁

RANK	UNIV	FIW ₁
1º	UFJF	0,04
2º	UFTM	0,04
3º	UFABC	0,01
4º	UFSJ	0,01
5º	UFRRJ	0,00
6º	UFOP	0,00
7º	UFES	0,00
8º	UNIRIO	0,00
9º	UFLA	0,00
10º	UNIFEI	0,00
11º	UFV	0,00
12º	UFSCAR	0,00
13º	UFRJ	0,00
14º	UNIFESP	0,00
15º	UFMG	0,00
16º	UFU	0,00
17º	UFF	0,00
18º	UNIFAL	0,00
19º	UFVJM	0,00

Fonte: Elaborada pelos autores.

Tabela 2 – Ranking das universidades federais da região sudeste do Brasil com a aplicação do FIW₂

RANK	UNIV	FIW ₂
1º	UFRJ	613
2º	UFMG	393
3º	UNIFESP	377
4º	UFF	333
5º	UFV	255
6º	UFJF	225
7º	UFES	175
8º	UFSCAR	170
9º	UFTM	157
10º	UFU	129
11º	UFOP	128
12º	UFRRJ	112
13º	UFLA	101
14º	UNIRIO	84
15º	UFABC	31
16º	UFSJ	30
17º	UNIFEI	29
18º	UNIFAL	2
19º	UFVJM	0

Fonte: Elaborada pelos autores.

Tabela 3 – Ranking das universidades federais da região sudeste do Brasil por tamanho do site

RANK	UNIV	TAM
1º	UFF	19.500.000
2º	UFRJ	5.470.000
3º	UFMG	4.120.000
4º	UNIFESP	3.450.000
5º	UFV	2.060.000
6º	UFU	1.470.000
7º	UFSCAR	1.360.000
8º	UFES	812.000
9º	UFOP	560.000
10º	UFLA	540.000
11º	UNIRIO	439.000
12º	UFRRJ	395.000
13º	UNIFEI	149.000
14º	UNIFAL	89.700
15º	UFJF	55.400
16º	UFSJ	47.700
17º	UFTM	38.500
18º	UFABC	30.100
19º	UFVJM	20.300

Fonte: Elaborada pelos autores.

Tabela 4 – Ranking das universidades federais da região sudeste do Brasil pelo indicador visibilidade

RANK	UNIV	VISIBIL
1º	UFRJ	9.507
2º	UFMG	5.993
3º	UNIFESP	5.678
4º	UFF	5.592
5º	UFV	3.704
6º	UFJF	2.457
7º	UFSCAR	2.405
8º	UFES	2.384
9º	UFU	1.832
10º	UFOP	1.693
11º	UFTM	1.657
12º	UFRRJ	1.444
13º	UFLA	1.336
14º	UNIRIO	1.095
15º	UNIFEI	344
16º	UFSJ	326
17º	UFABC	318
18º	UNIFAL	20
19º	UFVJM	4

Fonte: Elaborada pelos autores.

Na leitura dos resultados mostrados na Tabela 1, pode-se observar primeiramente que das 19 universidades analisadas, 15 apresentam valores de FIW nulos (0,00), o que chama a atenção. Depois verifica-se que o valor do FIW (Tabela 1) e o do FIWIn (Tabela 2) são diferentes para a mesma universidade analisada e que a colocação de cada uma delas no *ranking* web também não é a mesma, dependendo da maneira como o FIW é calculado.

O que se modifica com a inclusão do logaritmo natural ao tamanho do sítio é que este propicia valores mais acordes de FIW. O logaritmo transforma faixas de valores muito elevadas (como tamanho de sítios da ordem de milhões de páginas) em valores mais baixos, porém, sendo uma transformação monotônica e crescente, preserva a ordem dos dados, apenas diminuindo a faixa dos valores. O resultado final é a mudança de lugar das universidades no *ranking* web, quando calculado pelo FIWIn.

A UFRJ - Universidade Federal do Rio de Janeiro, universidade importante no cenário nacional, por exemplo, que alcança o maior índice de visibilidade ou “situações”, 9.507 (Tabela 4), apresenta um FIW de valor 0,00 (Tabela 1), enquanto que no FIWIn, o valor obtido por esta mesma universidade é de 1.410 (Tabela 2). Além disto, o que reforça a idéia de que o FIW da forma como era calculada não era a mais adequada é que a posição desta mesma universidade com relação a este indicador passa do 13º lugar (Tabela 1) para o 1º lugar quando calculado de acordo com a nova fórmula (Tabela 2).

Como já foi exposto, existe uma forte relação entre o indicador de visibilidade e o FIW. Neste sentido, outro fator que leva a acreditar que a inclusão de um logaritmo ao denominador da fórmula original resolveria os problemas de resultados espúrios reside no fato de que, ao se comparar as posições alcançadas pelas mesmas universidades no *ranking* com a aplicação do indicador visibilidade (Tabela 4) e do FIWIn (Tabela 2), elas são praticamente as mesmas: até o 6º lugar, as universidades coincidem na sequência tanto em um indicador quanto no outro: a UFRJ, a UFMG, a UNIFESP, a UFF, a UFV, a UFJF ocupam o mesmo lugar tanto em visibilidade quanto em FIWIn. Por outro lado, o que ocorre na comparação entre as posições obtidas pelas mesmas universidades, se forem examinados os resultados para o indicador visibilidade e para o FIW, é que eles são muito diferentes. Neste caso não há coincidência alguma entre o lugar obtido por estas instituições no *ranking*.

Percebe-se ainda que, de acordo com a fórmula ora sugerida, a interpretação dos resultados também deve ser diferente da sugerida por Smith (1999), pois não há um valor fixo estabelecido como critério para saber se o valor obtido por cada sítio é alto ou baixo, como antes. A idéia, neste caso, é calcular a média obtida pelo conjunto de sítios analisados e, a partir deste valor, determinar quais os resultados que estão acima da média (considerados FIWs altos) e quais estão abaixo da média (considerados FIWs baixos).

A partir do entendimento de que a incorporação de um logaritmo natural ao denominador da fórmula original traz resultados mais acurados ao cálculo do FI de páginas web, sugere-se, neste estudo, a sua alteração, a fim de que este alcance a mesma credibilidade já há muito tempo conseguida com o cálculo do fator de impacto de revistas.

Considerações finais

O mais importante em uma análise como a exposta neste trabalho não é exatamente se determinada instituição aparece em primeiro ou em último lugar no *ranking* por indicador, mas sim o que tal posição pode representar em termos de aprovação ou não da sua política de disponibilização de informações na web, de motivação para modificar ou ampliar a forma de divulgação de informações úteis ou de conscientização em disseminar informações de qualidade, que possam ser atualizadas e consultadas com frequência.

Foi com a intenção de contribuir neste sentido que se buscou adaptar uma fórmula que já havia sido pensada e criada por outros autores com a mesma idéia: avançar em um campo da área de Ciência da Informação que, naquele momento (1997), pouco se conhecia: a webometria, campo que ainda hoje apresenta grandes desafios.

Foi possível comprovar, neste estudo, que a incorporação de um logaritmo natural ao denominador da fórmula original traz resultados mais acurados do que a simples aplicação desta. Ainda pôde-se constatar que as posições das universidades no que diz respeito ao fator de impacto no *ranking* webométrico continuaram se mantendo também muito próximas daquelas ocupadas no *ranking* quando o indicador se referia à visibilidade. Neste sentido, pode-se destacar que a coincidência em quase todas as posições obtidas pelas universidades por meio do indicador visibilidade e da aplicação do FIWIn vem reafirmar a pertinência desta nova forma de cálculo.

Artigo recebido em 15/02/2013 e aprovado em 26/03/2013.

Referências

AGUILLO, Isidro F. (Coord.) *Factor de impacto y visibilidad de 4.000 sedes web universitarias españolas*. Madrid: CINDOC, 2005. Disponível em: <http://www.cindoc.csic.es/estudios_ea2004_0020_informe.doc>. Acesso em: 15 mar. 2006.

GARFIELD E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, v. 122, n. 3159, p. 108-111, 1955. Disponível em: <http://www.garfield.library.upenn.edu/papers/science_v122v3159p108y1955.html>. Acesso em: 03 out. 2010.

INGWERSEN, P. The calculation of web impact factors. *Journal of Documentation*, v. 54, n. 2, p. 236-243, 1998.

KOUSHA, Kayvan. Webometrics and scholarly communication: an overview. *Quarterly Journal of the National Library of Iran*, v. 14, n. 4, p. 7-16, 2005.

RODRÍGUEZ-GAIRÍN, J. M. Valorando el impacto de la información en internet: AltaVista, el "Citation Index" de la red. *Revista Española de Documentación Científica*, v. 20, n. 2, p. 175-181, 1997. Disponível em: <<http://bd.ub.es/pub/rzgairin/altavis.htm>>. Acesso em: 03 out. 2010.

ROUSSEAU, Ronald. Sitations: an exploratory study. *Cybermetrics*, v. 1, n. 1, 1997. Disponível em: <http://www.cindoc.csis.es/cybermetrics_volume_1_issue_1_paper1_sitations_an_exploratory_study_by_rousseau.htm>. Acesso em: 25 set. 2011.

SMITH, A.G. A tale of two web spaces: comparing sites using web impact factors. *Journal of Documentation*, v. 55, n. 5, p. 577-592, 1999.

THELWALL, Mike. Results from a web impact factor crawler. *Journal of Documentation*, v. 57, n. 2, Mar. 2001.

THOMSON REUTERS. History of citation indexing. *Science*, Oct. 2010. Disponível em: <http://thomsonreuters.com/products_services/science/free/essays/history_of_citation_indexing/>. Acesso em: 28 jun. 2010.

VANTI, Nadia. Application of web indicators to the latin-american academic sites in social sciences. *Brazilian Journal of Information Science*, v.1, n.2, p. 22-46, 2007. Disponível em: <<http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/37>>. Acesso em: 04 out. 2010.