

FATOR DE IMPACTO WEB (FIW): um novo cálculo para a obtenção de resultados mais acurados

Eixo temático: Informetria e Webometria

Modalidade: Apresentação oral

1 INTRODUÇÃO

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 tecnologia. Para tais análises foram criados determinados indicadores como forma de padronizar os procedimentos de análise e servir 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. 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).

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), 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 nas pesquisas realizadas com tal finalidade.

A proposta deste trabalho, portanto, é entender porque que 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 a fim de obter resultados mais acurados ao tipo de recurso utilizado: sítios web.

2 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 publicou, consultou e citou. (THOMSON REUTERS, 2010).

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 *Institut 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 *Thomsom Reuters* e pode ser consultada por meio do sítio do próprio *ISI Web of Knowledge*, disponibilizado também 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.

3 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 Rodriguez-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), representa a razão entre o número de links recebidos por um sítio web e o seu tamanho em um determinado tempo.

Neste sentido, assim como para as análises quantitativas de materiais impressos utilizam-se as bases de dados do ISI como ferramenta, os motores de busca são utilizados para efetuar buscas, coletar dados e quantificar as informações contidas nos sítios web. Pode-se afirmar, portanto, que há certa equivalência entre as bases de dados do ISI e os motores de busca na web tais como o Google, Yahoo, Alta Vista, entre outros.

De acordo com Smith (1999) o FIW serve para medir e comparar a atratividade e influência de distintos espaços web e deve ser calculado por meio de uma fórmula pela qual se divide o número de links que recebe um determinado sítio (*inlinks* externos) pelo número de páginas deste mesmo sítio (tamanho do sítio), em certo espaço de tempo. A fórmula do FIW costuma ser representada da seguinte maneira:

$$\text{FIW} = \frac{\text{Nº de páginas que apontam para determinado sítio}}{\text{Nº de páginas do sítio apontado}}$$

A partir de então, os estudos webométricos que utilizavam o FIW como um dos indicadores para determinar a posição de determinados sítios no ranking da web recorriam à fórmula sugerida por Ingwersen (1998) e Smith (1999) para calculá-lo. Este último autor também estabeleceu os critérios para os valores obtidos por meio deste indicador: sítios considerados com alto FIW são aqueles que possuem índices significativamente maiores do que 1, sítios com médio FIW são os 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 tem se 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 na revista, daí resultando um valor de FI bastante significativo, os *inlinks* externos ou as citações 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 autores como Thelwall (2001), Kousha (2005) e Vanti (2007), em seus estudos e pesquisas sobre o assunto, já apresentaram questionamentos acerca da funcionalidade deste indicador no âmbito web.

4 METODOLOGIA

Tomando-se como base a fórmula original ideada por Rodriguez-Gairín (1997) e Ingwersen (1998), tentou-se adaptá-la para que cumprisse o seu objetivo que é o de apontar quais são os sítios que recebem o maior número de links externos em relação ao seu tamanho em um determinado momento 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 a fórmula original, 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 (o que é comum a todos os recursos disponíveis neste ambiente), 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 fórmula, potencializando-o de modo a cumprir o seu objetivo. A fórmula ficou assim:

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

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 Ingwersen (1998) e Smith (1999) num primeiro momento (FIW_1) e depois a fórmula alternativa criada pelos autores deste trabalho (FIW_2), para assim poder comparar os resultados obtidos.

5 RESULTADOS

Os resultados obtidos para o FIW das universidades federais da região sudeste do Brasil com as duas fórmulas supracitadas, bem como os indicadores tamanho e visibilidade, também envolvidos no cálculo, podem ser visualizados na Tabela 2.

Tabela 2 – Ranking das universidades federais da região sudeste do Brasil por indicador

RANK	UNIV	TAM	UNIV	VISIBIL	UNIV	FIW_1	UNIV	FIW_2
1º	UFF	19.500.000	UFRJ	9.507	UFJF	0,04	UFRJ	1.410
2º	UFRJ	5.470.000	UFMG	5.993	UFTM	0,04	UFMG	905
3º	UFMG	4.120.000	UNIFESP	5.678	UFABC	0,01	UNIFESP	868
4º	UNIFESP	3.450.000	UFF	5.592	UFSJ	0,01	UFF	767
5º	UFV	2.060.000	UFV	3.704	UFRRJ	0,00	UFV	586
6º	UFU	1.470.000	UFJF	2.457	UFOP	0,00	UFJF	517
7º	UFSCAR	1.360.000	UFSCAR	2.405	UFES	0,00	UFES	403
8º	UFES	812.000	UFES	2.384	UNIRIO	0,00	UFSCAR	392
9º	UFOP	560.000	UFU	1.832	UFLA	0,00	UFTM	361
10º	UFLA	540.000	UFOP	1.693	UNIFEI	0,00	UFU	297
11º	UNIRIO	439.000	UFTM	1.657	UFV	0,00	UFOP	294
12º	UFRRJ	395.000	UFRRJ	1.444	UFSCAR	0,00	UFRRJ	258
13º	UNIFEI	149.000	UFLA	1.336	UFRJ	0,00	UFLA	233
14º	UNIFAL	89.700	UNIRIO	1.095	UNIFESP	0,00	UNIRIO	194
15º	UFJF	55.400	UNIFEI	344	UFMG	0,00	UFABC	71
16º	UFSJ	47.700	UFSJ	326	UFU	0,00	UFSJ	69,6
17º	UFTM	38.500	UFABC	318	UFF	0,00	UNIFEI	66,4
18º	UFABC	30.100	UNIFAL	20	UNIFAL	0,00	UNIFAL	4,03
19º	UFVJM	20.300	UFVJM	4	UFVJM	0,00	UFVJM	0,92

Fonte: Elaborado pelos autores.

Por meio dos resultados mostrados na Tabela 2, pode-se observar primeiramente que na coluna 4 (FIW_1), das 19 universidades analisadas 15 apresentam valores nulos (0,00), o que é preocupante. Depois se verifica que o FIW_1 e o FIW_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. A UFRJ - Universidade Federal do

Rio de Janeiro, universidade importante no cenário nacional, por exemplo, apresenta no FIW_1 o valor de 0,00 (nulo), enquanto que no FIW_2 , o valor obtido por esta mesma universidade é de 1.410. Além disso, o que reforça a idéia de que o FIW da forma como era calculada não era a mais adequada, é que a colocação desta universidade passa do 13º lugar de acordo com o valor do FIW_1 para o 1º lugar seguindo o valor do FIW_2 .

Outro fator que contribui para se acreditar que a inclusão de um Logaritmo no denominador da fórmula proposta por Ingwersen (1998) e Smith (1999) pode resolver os problemas de resultados espúrios do FIW como era calculado anteriormente, é que se forem comparados os números alcançados pelas distintas universidades no que diz respeito aos indicadores visibilidade e FIW_2 é perceptível a proximidade existente entre eles. Até o 6º lugar no ranking, 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 FIW_2 .

Por outro lado, o que ocorre na comparação entre as posições obtidas pelas universidades se forem examinados a visibilidade, o tamanho e o FIW_1 é muito diferente. 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 está 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 FIW's altos) e quais estão abaixo da média (considerados FIW's baixos).

4 CONCLUSÕES

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, etc.

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

REFERÊNCIAS

- AGUILLO, Isidro F. (Coord.) *Factor de impacto y visibilidad de 4.000 sedes web universitarias españolas*. Madrid: CINDOC, InternetLab, 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. <<http://bd.ub.es/pub/rzgairin/altavis.htm>>. Acesso em: 03 out. 2010.
- 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, March 2001.
- THOMSON REUTERS. History of Citation Indexing. *Science*, October 15, 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.