

Novos caminhos e alternativas para a Webometria

Fábio Castro Gouveia

RESUMO

Com cerca de 15 anos de existência a webometria vivencia hoje um cenário de crise e novos desafios diante das restrições de acesso a informação impostos pelos mecanismos de busca comerciais, colocando em cheque as tradicionais análises de *hiperlinks*. Historicamente, o link foi a unidade central de informação da webometria por ser e a ligação direta entre os atores (páginas) que compõem Web. Diante deste cenário, como obter daqui para frente dados webométricos? Neste momento crítico, novas oportunidades são geradas pelo desenvolvimento de metodologias para o campo. Neste artigo apontamos algumas reflexões e exploramos possibilidades que vem sendo discutidas entre os pesquisadores, como retorno ao uso de *web crawlers* e a citação da URL e de palavras em mecanismos de busca comerciais, além de trazer uma abordagem ainda pouco explorada, por meio do uso de um operador do Google, o *allinanchor*.

PALAVRAS-CHAVE: Webometria. Cibermetria. Mecanismo de busca. Aspector metodológicos. Allinanchor.

1 Introdução

Criada há cerca de 15 anos, em um campo fértil da informática, que surge no desenvolvimento das Tecnologias de Informação e Comunicação (TICs), a webometria vivencia hoje um cenário de crise e novos desafios diante das restrições de acesso a informação impostos pelos mecanismos de busca comerciais. Estas restrições colocam em cheque as tradicionais análises de *hyperlinks* e dão início a um debate crucial entre os pesquisadores: como trabalhar e obter daqui para frente dados webométricos? Quais os caminhos e as alternativas possíveis e mais indicados para obtenção de métricas na web?

Historicamente, o principal objeto de estudo da webometria é o *link*, considerado como a unidade central de informação e a ligação direta entre os atores (páginas) que compõem essa vasta rede virtual. Em sua origem, os estudos webométricos se basearam na obtenção desta rede virtual por intermédio de *web crawlers*, *softwares* especializados que levantavam o conteúdo interligado dos *websites* estudados e faziam o mapeamento dos *links* existentes. Mudanças tecnológicas que tornavam os links inacessíveis (ex: conteúdos em *Flash* etc.), bem como o acesso restrito apenas à *Web* visível sempre foram desafios. A instabilidade da manutenção dos conteúdos na *Web*, parte da dinâmica natural de seu desenvolvimento, também já foi motivo de debate e questionamento (BAR-ILAN, 1999; ROUSSEAU, 1999). O próprio significado do link e suas motivações, bem como a analogia com a citação suscitou, ao longo dos anos, estudos voltados para o melhor entendimento da dinâmica de interligação entre páginas na *Web* (PRIME; BASSECOULARD; ZITT, 2002; SMITH, 2004; ZUCCALA, 2006).

Como um campo de pesquisa ainda recente e diretamente relacionado ao avanço das tecnologias e da ciência da computação, a webometria tem se debruçado sobre estudos de cunho exploratório que visam testar ferramentas e novas metodologias (LANG; GOUVEIA; LETA, 2010), e investigar possibilidades de aplicação na cientometria (HERRERO-SOLANA; MORALES-DEL-CASTILLO, 2004; ORTEGA et al., 2008; VAUGHAN, 2006), e para além dela, como na área de competitividade comercial (VAUGHAN; YOU, 2006), análise institucional (LANG; GOUVEIA; LETA, 2008) e de museus e centros de ciência (GOUVEIA; KURTENBACH, 2009).

Com o passar do tempo, os mecanismos de busca comerciais se estabeleceram como fonte frequente de informação para os estudos webométricos. Apesar do fenômeno de isolamento de determinados conteúdos levando a sua não indexação pelos mecanismos de busca, o uso de entradas múltiplas no processo

de formação da base de dados destas ferramentas levava a um melhor mapeamento de web e a formação da assim chamada “web visível”. No entanto, suas funcionalidades variavam e a escolha por determinado mecanismo dependia, prioritariamente, do tipo de busca pretendida.

O cenário de opções nos últimos cinco anos se tornou instável com exemplos de estudos que necessitaram utilizar diferentes mecanismos de busca em um intervalo de seis meses devido a mudanças nas permissões de acesso aos levantamentos webométricos (VAUGHAN; TANG; DU, 2009). Mais recentemente, apenas o Yahoo! ainda permitia o uso de buscas combinadas por termos booleanos¹ (fundamentais para os estudos de *interlink*² e *co-inlinks*³) e automatizadas por meio de *Application Programming Interface* (API).

Mas a atual crise tem início em 2009, quando o Yahoo! é adquirido pela Microsoft, que já havia interrompido a possibilidade de buscas avançadas no seu mecanismo de busca, o Bing, e desejava migrar a base de dados do Yahoo!. Em abril de 2011, o uso de APIs foi totalmente bloqueado pelo Yahoo!, impossibilitando buscas automatizadas em larga escala (YAHOO!, 2012). No início de 2012 já não era mais possível realizar uma série de tipos diferentes de busca por *hiperlinks*.

Se por um lado o momento atual representa uma crise, por outro, gera oportunidades de desenvolvimento de novas metodologias para o campo. Diante deste cenário, apontaremos a seguir algumas reflexões e exploraremos possibilidades que vem sendo discutidas entre os pesquisadores, como retorno ao uso de *web crawlers* e a citação da URL e de palavras em mecanismos de busca comerciais, além de trazer uma abordagem ainda pouco explorada, por meio do uso de um operador do Google, o *allinanchor*.

2 A webometria e os seus desafios ao longo do tempo

A webometria – termo cunhado por Almind e Ingwersen (1997) – trata dos aspectos quantitativos tanto da construção, quanto do uso da Web. Compreende quatro áreas de pesquisa no campo: análise de conteúdo das páginas web; análise da estrutura dos links; análise do uso da Web; e análise de tecnologias da Web (THELWALL; VAUGHAN; BJÖRNEBORN, 2005).

No final da década de 1990, os principais mecanismos de busca utilizados eram o AltaVista e o Infoseek, em razão das funcionalidades de busca oferecidas como o uso de operadores booleanos. No entanto, à época, ainda havia poucos estudos sobre o processo de indexação por esses mecanismos, o que levantou questionamentos sobre a confiabilidade dos dados e a opção de pesquisadores pelo uso de *web crawlers* (THELWALL, 2001).

¹ Operadores booleanos, no contexto dos mecanismos de busca, são palavras ou símbolos que definem como o mesmo deve combinar os termos de busca incluídos em uma pesquisa.

² *Inlinks* são os *links* recebidos por uma página ou sítio de uma outra página. É o análogo webométrico da citação.

³ *Co-inlink* são *inlinks* simultaneamente fornecidos por uma mesma página a páginas ou sítios em estudo. É o análogo webométrico da co-citação.

De fato, deve-se considerar que a indexação dos mecanismos de busca não é completa. Gulli e Signorini (2005), em um estudo de cobertura dos mesmos verificaram que da web visível o Google indexava 79%, o Yahoo/AltaVista, 69%, o MSN, 62% e o Ask/Teoma, 58% sendo que por terem conjuntos distintos, apenas 28,85% das páginas eram indexadas simultaneamente pelos quatro mecanismos de busca.

Vanti (2002) alertou para a natureza dinâmica e em tempo real da Internet, de sua enorme massa de dados não padronizados, bem como das inconsistências entre os mecanismos de busca. A preocupação com a estabilidade dos dados coletados por essas ferramentas foi debatida entre os pesquisadores em busca de possíveis soluções (BAR-ILAN, 1999; ROUSSEAU, 1999; METTROP; NIEUWENHUYSEN, 2001). Dez anos após a criação formal da webometria, Payne e Thelwall (2007) realizaram um estudo longitudinal (2000 a 2005) sobre universidades apontando para uma tendência de estabilização da rede acadêmica.

Ainda com relação aos mecanismos de busca, a estimativa do número de páginas que satisfazem uma determinada consulta também pode variar. Com o intuito de padronizar a metodologia aplicada à pesquisa no AltaVista, foi conduzido um estudo para verificar qual a melhor abordagem para se obter valores mais estáveis (GOUVEIA, 2007). Para tal, um conjunto de consultas realizado com uma mesma cláusula, foi efetuado o levantamento de dados partindo de registros iniciais 1, 6, 11, 21, 31, 41, 51, 61, 71, 81, 91 e 101, e com número de resultados por página de 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 e 100. Os resultados para o número total de *inlinks* estimados pelo AltaVista para cada uma destas consultas foram colocados em uma matriz para posterior representação gráfica com o objetivo de verificar a oscilação dos valores obtidos. Os resultados apontaram para uma tendência à estabilidade, quando o número de resultados por página era ampliado para 100 ou o registro inicial era superior a 100.

Em 2005, no entanto, o AltaVista foi comprado pelo Yahoo! e houve gradativamente uma fusão das bases de dados de ambos os mecanismos. Os pesquisadores passaram então a optar pelo uso do Yahoo! para condução de seus estudos. Quatro anos mais tarde a ferramenta foi comprada pela Microsoft, anunciando uma crise para grande parte dos estudos com *hiperlinks* por definir um cronograma para fechamento do acesso às buscas com cláusulas webométricas.

3 Web crawlers e mecanismos de busca diante do cenário atual

Em relação aos estudos clássicos do campo da webometria, diferenças sempre existiram quanto à aplicabilidade dos métodos de coleta por mecanismos de busca comerciais ou por *web crawlers*. Para o Fator de Impacto na Web (FIW) e Fator de Impacto na Web externo (eFIW) (INGWERSEN, 1998), necessita-se do levantamento do número de páginas com links para um determinado sítio (excluídos os links provenientes de páginas do próprio sítio no caso do eFIW) e o número de páginas de um determinado sítio. Apenas os mecanismos de busca seriam capazes de fazer um levantamento amplo que leva em consideração todas as páginas encontradas em sua vasta base de dados⁴. Hoje, esse tipo de análise só seria possível por meio dos *web crawlers*. Por outro lado, esse tipo de estudo estaria limitado a universos controlados, nos quais todos os sítios de uma determinada área de conhecimento e região são levantados. Assim, o FIW passaria a ser estimado pelo impacto que estes sítios têm em seus pares.

Para as análises de *co-inlink*, que se baseiam no levantamento do número de páginas que simultaneamente fornecem links para duas páginas individuais de diferentes domínios da Web estudados, as mesmas regras se aplicam. Os mecanismos de busca que permitiam consultas booleanas não são mais acessíveis para esse tipo de estudo. Novamente, a opção se daria pelo uso de *web crawlers*, mas funcionais somente dentro de uma perspectiva de levantamento de um campo fechado e restrito a priori.

Para estudos de *co-inlink*, resta apenas, no momento, a opção de co-citação das URLs estudadas, conforme veremos mais adiante, ou o levantamento por *web crawlers*, com as mesmas restrições já explicitadas.

Um recurso pouco explorado, mas ainda presente dentro da ferramenta Bing é o levantamento de links provenientes de um domínio e o estudo em conjunto de *co-outlinks* (limitado ao uso da API) a partir do uso da cláusula de busca *linkfromdomain*:. Estudos usando subdomínios podem não funcionar e estas cláusulas têm apresentado comportamento instável (STATISTICAL CYBERMETRICS RESEARCH GROUP, 2012). Porém, mediante controles no levantamento, pode-se obter resultados satisfatórios para campos da web visível com boa cobertura e que tenham uma densidade de *links* moderada, o que tornaria possível a verificação dos resultados.

Já no caso dos estudos de *interlink*, que focam no intercâmbio de links entre domínios da Web estudados, tanto os mecanismos de busca como os *web crawlers* apresentam condições de serem aplicados, sendo o diferencial estabelecido pelas características destes dois métodos quanto ao levantamento de dados webo-

⁴ Deve-se ter em conta que a vasta base de dados de um mecanismo de busca é na realidade uma amostra incompleta da web possivelmente indexável. Além disso, segundo Hu *et al.* (2001), cerca de 85% do tráfego na Internet é direcionado pelos mecanismos de busca. Dessa forma, se eles contêm vieses em direção às páginas de maior acesso e visibilidade mundial, também contribuirão para a manutenção desse *status quo*.

métricos.

Do ponto de vista do pesquisador do campo da webometria, a abordagem usando os *web crawlers* especializados ou os mecanismos de busca comerciais apresenta diferentes questões a serem consideradas. Para os *web crawlers*, temos a vantagem de um algoritmo de busca conhecido, desenvolvido especificamente com propósitos webométricos, mas que, por outro lado, leva à necessidade de se criar uma base de dados para cada pesquisa e com custos operacionais e de tempo computacional elevados para um laboratório de pesquisa em caso de levantamentos de grande porte. Já para os mecanismos de busca, temos um algoritmo de busca proprietário, com prioridade em fornecer resultados satisfatórios e não necessariamente exatos ou amostralmente fidedignos, mas conta-se com uma ampla base de dados disponível a todo o momento.

Além disso, no caso dos *web crawlers*, temos o problema referente ao levantamento de páginas do domínio estudado, onde subdiretórios ou páginas e sítios que não estejam diretamente ligados ao ponto de partida para a coleta do *web crawler*, apesar de fazerem parte e poder ter papel significativo na instituição analisada, podem ficar de fora da amostra coletada (THELWALL; TANG; PRICE, 2003). A única forma de se garantir a cobertura completa de todas as páginas estáticas de um sítio é ter acesso privilegiado ao servidor onde as páginas estão hospedadas. Neste processo, executa-se um *software* que extrai os links diretamente dos arquivos em vez de segui-los como um *web crawler*. Entretanto, este tipo de abordagem costuma ser limitada pela dificuldade de acesso aos servidores que se deseja estudar. (THELWALL, VAUGHAN; BJÖRNEBORN, 2005). Neste sentido, os levantamentos feitos por mecanismos de busca comerciais podem ser mais abrangentes e fidedignos dependendo da estrutura e conectividade dos conteúdos disponibilizados *online* em um domínio.

Analisando as cláusulas clássicas da webometria, temos ainda acesso aos dados de número de páginas indexadas (cláusula *site* ou similar) tanto para o Bing quanto para o Google. Informações quanto ao número de páginas com links para um domínio estão disponíveis por ferramentas voltadas para webmasters oferecidas por mecanismos de busca específicos, porém a custos elevados (ex: *Bleeko SEO tools* - US\$ 99.00 por mês). A opção que vem sendo recomendada pelos autores do campo e que será mais detalhada a frente é a citação da URL, que pode ainda ser usada com a cláusula de exclusão das citações da URL pelo próprio sítio estudado, evitando assim as “auto-citações” ou *links* internos. O número de páginas com *links* para um determinado URL ainda é fornecido pelo Google, porém obtém-se acesso a apenas uma amostra dos resultados (STATISTICAL CYBERMETRICS

RESEARCH GROUP, 2012).

A questão que ameaça os estudos webométricos nos dias de hoje se situa no crescente interesse comercial pelos dados referentes à rede de links que os sítios recebem e, por conseguinte, a estruturação destes tipos de levantamentos como negócio para empresas de análise estratégica de posicionamento na Web. Vaughan e You (2006) foram pioneiros na utilização de *co-inlink* para análise competitiva. Esse tipo de estudo foi posteriormente explorado por outros autores (VAUGHAN; GAO; KIPP; 2006; VAUGHAN; KIPP; GAO, 2007; VAUGHAN; TANG; DU, 2009). O campo da webometria esteve focado ao longo dos anos no entendimento das interrelações institucionais em diversas temáticas e é de se lamentar que uma ciência que teve papel pioneiro no uso destes dados esteja hoje ameaçada diante de interesses comerciais.

Infelizmente, o cenário atual dos mecanismos de busca nos oferece dados limitados e com restrições ao volume de buscas por mês. A comercialização do acesso aos dados de interesse em pesquisas webométricas culmina com a limitação do Bing para o número de resultados por mês para um determinado usuário cadastrado (5 mil páginas por mês, cada uma com até 50 resultados). Isto na prática limita o pesquisador a 250 buscas completas por mês, já que uma busca completa com 1 mil ou mais resultados (dos quais só se tem acesso aos primeiros 1 mil) consome 20 páginas. Para buscas adicionais é necessário comprar créditos, e as cláusulas disponíveis são bastante limitadas em comparação as previamente acessíveis.

4 Novos caminhos e alternativas para a webometria

Os novos caminhos para a webometria podem se iniciar nas velhas trilhas deixadas pelos pioneiros do campo, com os *web crawlers*, ou seguir adiante, utilizando os mecanismos de busca, mas com abordagens diferentes, por meio da citação da URL ou de termos específicos, e de um operador ainda pouco explorado do Google, o *allinanchor*.

O “retorno” dos *web crawlers* especializados serve de ponto de partida para a manutenção das análises que têm o link como unidade central de estudo. A funcionalidade desta metodologia de levantamento de dados se manteve, ao longo do tempo, tendo sido utilizada mesmo durante o apogeu dos estudos via mecanismos de busca comerciais. A principal crítica a este método está no fato de só serem encontradas novas páginas na web seguindo os links existentes nelas. Por isso, é possível que páginas ou coleções de páginas que estejam isoladas, com links quebrados ou ainda em JavaScript ou Flash não sejam rastreadas.

No contexto das alternativas frente ao desafio colocado pelos mecanismos de busca, as abordagens que se referem à citação da URL e de termos específicos, vêm sendo amplamente discutidas (THELWALL; SUD, 2011; THELWALL; SUD, WILKINSON, 2012; VAUGHAN; YOU, 2010). Ambas exploram as possibilidades de se continuar utilizando mecanismos de busca comerciais a partir de uma nova perspectiva, que não tem mais o link como unidade central da informação.

A citação da URL se refere à busca textual pelo endereço de uma dada página na web (ex. www.brasil.gov.br). Se, por um lado ela garante o foco está no endereço, por outro, é um tipo de citação pouco usual e não tem vínculo obrigatório com o *link*. Ou seja, embora a citação da URL se refira ao *link* e, nesse sentido, mantenha uma relação direta entre as páginas, não segue a lógica da construção de conteúdo na web. Não é comum que a URL seja textualmente escrita nas páginas web, mas sim inserida no marcador de destino para um texto âncora, não deixando claro se ela poderia ser considerada, de fato, uma boa fonte ou evidência de impacto na Internet (THELWALL; SUD, 2011). Neste sentido, a principal desvantagem seria conceitual. A citação da URL pode levar ao mesmo tempo a uma subamostragem do total de *links* fornecidos ou mesmo uma superestimativa em casos nos quais não houve a intenção de dar o *link* para o sítio, mas este endereço foi colocado no texto. Por outro lado, em documentos PDF que não tiveram a inclusão destes endereços como *link*, tem-se uma situação especial. A partir da crise estabelecida, talvez estejamos nos aproximando mais de levantamentos de dados referentes a documentos com provável importância acadêmica e nos distanciando do acesso a links meramente estruturais existentes em páginas HTML. Assim sendo, diante da crise gerada pelos mecanismos de busca atuais, abre-se uma oportunidade para o desenvolvimento de novas pesquisas que têm o endereço *Web* institucional como foco.

Já a citação da palavra ou de termos em páginas na web corresponde à menção textual de um nome ou termo em uma dada página, sem necessariamente estar associado a um *link*. Este é um tipo de busca que sempre existiu, mas que até então não era comumente objeto de estudos webométricos. A principal desvantagem está justamente no fato de não fazer um vínculo obrigatório com o *link* e a necessidade de termos de busca únicos.

Em um estudo recente, Vaughan e You (2010) compararam os resultados de *co-inlink* de uma pesquisa prévia com a busca textual pelos nomes das instituições estudadas, no intuito de investigar a validade do método como medida de similaridade e proximidade na web. A similaridade observada entre os dois diagramas de escalonamento multidimensional (*Multidimensional*

Scalling – MDS) gerados a partir dos *co-inlinks* e da co-ocorrência de palavras apontou para uma correlação positiva entre as metodologias.

No entanto, apesar das correlações encontradas, a citação da palavra não estabelece uma relação direta entre as páginas e não representa uma intencionalidade do produtor de conteúdo na indicação do outro ente virtual. Perde-se, portanto, a premissa de legitimidade conferida por intermédio do *link*. Outra desvantagem reside no fato de que um mesmo termo pode ter significados distintos ou variações, complicando a análise dos dados. Assim, a questão se situa na necessidade de se encontrar termos específicos e representativos. Buscas por USP podem retornar resultados para a Universidade de São Paulo ou para a *University of the South Pacific* das ilhas Fiji, dentre outras hipóteses, e a filtragem destes dados pode ser impraticável. Além disso, formas múltiplas de se mencionar um determinado objeto de estudo, como o nome por extenso e/ou a sigla da instituição, podem levar também a necessidade de fusão de resultados e gerar mais ruído nas estimativas de busca obtidas.

No âmbito do debate sobre a eficiência dessas duas abordagens, um estudo conduzido por Thelwall e Sud (2011) comparou os métodos de citação da URL e da palavra, utilizando dois mecanismos de busca (Yahoo! e Bing). Os resultados obtidos para universidades britânicas validaram ambos os métodos como medida de impacto e apontaram para uma correlação positiva com a contagem de *inlinks*. Vale ressaltar, no entanto, que todo indicador tem suas limitações, e não deve ser utilizado isoladamente como medida de impacto ou qualidade.

Por último, temos uma abordagem pouco explorada até o momento, mas que merece destaque por manter a relação de análise com foco no link. Ela consiste no levantamento de dados a partir do operador do Google *allinanchor*: (GOOGLE..., 2010), que permite a busca por páginas que recebem links cujo texto âncora corresponde aos termos da consulta. Em outras palavras, ao clicar em um texto âncora você é levado ao destino deste link, e estas páginas de destino são as listadas pela busca usando este operador. Para um termo único, não composto, é possível ainda utilizar a cláusula *inanchor*: o que permite combinar a busca com outros termos que estariam presentes na página de destino.

Ao mesmo tempo em que fazem um vínculo com o *link*, estas abordagens são focadas no sentido oposto aos tradicionais levantamentos de *inlinks* (ou *links* recebidos por uma determinada página ou sítio) e tornam possíveis os levantamentos de *websites* que abordam questões temáticas que podem estar vinculadas a termos específicos presentes nos links distribuídos pela *web*.

Em um estudo recente, Friedrich, Gouveia e Leta (2009;

2012) aplicaram esta metodologia para o levantamento de *websites* relacionados à área de educação ambiental considerando-a uma abordagem eficaz para o mapeamento de áreas temáticas na *web* por meio dos links como indicadores de conexão dessas informações. Nesse sentido, a vinculação efetuada pela delimitação da busca ao termo âncora traz uma vantagem em relação aos métodos tradicionais de busca por palavra-chave dentro dos mecanismos de busca, pois a necessidade de tratamento dos dados diminui e a questão conceitual relativa à validação do vínculo entre as páginas é resolvida pelo *link*.

Embora este tipo de consulta tenha apresentado instabilidade desde outubro de 2011, fazendo-se necessário o acompanhamento e a verificação cuidadosa dos resultados obtidos, a proposta do *allinanchor*: permite um tipo de análise muito diferenciado, onde os endereços apresentados pelo resultado de busca têm como mecanismo de seleção a indicação de sites terceiros que identificaram o conteúdo destas páginas como sendo associado aos termos escolhidos na busca. Este tipo de estudo permite analisar o que é considerado pelos produtores de conteúdo como referente ao tema de estudo e desta forma tendo uma valoração externa ao produtor do conteúdo da página destino.

5 Considerações finais

Diante de um cenário de crise, novas oportunidades sempre se apresentam. O campo de estudo da webometria sempre esteve diante das mudanças que o desenvolvimento das TICs apresentam a cada dia. Mensurar, avaliar e mapear as redes de relações na web e o impacto de seus conteúdos estará sempre entre as necessidades deste campo. Assim sendo, o constante desenvolvimento de metodologias e a validação das mesmas fazem parte da dinâmica deste campo de estudo, e devemos considerá-los como oportunidades de propor novas saídas criativas para as barreiras impostas.

6 Agradecimentos

O autor gostaria de agradecer à leitura crítica do artigo por Pamela Barreto Lang e o apoio da Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

New paths and alternatives to webometrics

ABSTRACT

With nearly 15 years of existence, today webometrics experiences a crisis scenario and new challenges facing the restrictions on access to information imposed by commercial search engines, putting into question the traditional hyperlink analyzes. Historically, the link was the central unit of information to webometrics for being the direct link between the actors (pages) that make up the Web. Given this scenario, how can we get webometric data from now on? At this critical time, new opportunities are created by the development of methodologies for the field. In this paper we point out some thoughts and explore possibilities that have been discussed among researchers, such as the return to the use of web crawlers and the search for the URL and words citation in commercial search engines, and presents a fairly unexplored approach through the use of the Google's "allinanchor:" operator.

KEYWORDS: Webometric. Cybermetric. Search engine. Methodological Aspect. Allinanchor.

Referências

ALMIND, Tomas C.; INGWERSEN, Peter. Informetric analyses on the world wide web : methodological approaches to 'Webometrics'. **Journal of Documentation**, London, v. 53, n. 4, p. 404-426, 1997.

BAR-ILAN, Judit. Search engine results over time – a case study on search engine stability. **Cybermetrics**, Madrid, v. 2/3, 1998/1999. Disponível em: <<http://cybermetrics.cindoc.csic.es/articles/v2i1p1.html>>. Acesso em: 30 set. 2012.

FRIEDRICH, Margarete Pereira; GOUVEIA, Fábio Castro ; LETA, Jacqueline. O Produtor e o conteúdo da informação na internet: um estudo sobre o tema educação ambiental em páginas e sites brasileiros. **Encontros Bibli**, Florianópolis, v. 17, p. 157-170, 2012.

_____. How is the term environmental education linked on the Internet? A study of its representation on the web environment. In: INTERNATIONAL CONFERENCE OF THE INTERNATIONAL SOCIETY IN SCIENTOMETRICS AND INFORMETRICS (ISSI), 12., 2009, Rio de Janeiro. **Proceedings...** Rio de Janeiro, 2009. v. 2, p. 930-931.

GOOGLE Guide making searching even easier. 2010. Disponível em: <http://www.googleguide.com/advanced_operators.html>. Acesso em: 30 set. 2010.

GOUVEIA, Fábio Castro. Estudos webométricos de associações de museus e centros de ciência. 2007. Tese (Doutorado em Química Biológica) – Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2007.

GOUVEIA, Fábio Castro; KURTENBACH, Eleonora. Mapping the web relations of science centres and museums from Latin America. **Scientometrics**, Budapest, v. 79, n. 3. p. 491-505, 2009.

GULLI, A.; SIGNORINI, A. The Indexable Web is more than 11.5 billion pages. INTERNATIONAL WORLD WIDE WEB CONFERENCE, 14., 10-14 May 2005, Chiba, Japan. [**Proceedings...**] Chiba, Japan, 2005. p. 902-903 . Disponível em: <http://www.di.unipi.it/~gulli/papers/f692-gulli_signorini.pdf> Acesso: 30 set. 2012.

HERRERO-SOLANA, Victor; MORALES-DEL-CASTILLO, Jose. Mapas “geopolíticos” de internet: aplicación de las nuevas técnicas de representación de la información. **Ciência da Informação**, Brasília, DF, v. 33, n. 3, p. 69-75, 2004.

HU, Wen-chen.; CHEN, Yining; SCHMALZ, Mark .S.; RITTER, Gerhard X. An overview of the World Wide Web search technologies, In: WORLD MULTI-CONFERENCE ON SYSTEM, CYBERNETICS AND INFORMATICS, SCI, 5., 2001 July 22-25, Orlando, Florida. **Proceedings...** Orlando, 2001.

INGWERSEN, Peter. The Calculation of Web Impact Factors. **Journal of Documentation**, London, v. 54, n. 2, p. 236-243, 1998.

LANG, Pamela Barreto ; GOUVEIA, Fábio Castro ; LETA, Jacqueline. Relações intra-institucionais na Internet: um estudo exploratório com base em metodologias webométricas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 13, p. 137-150, 2008.

_____. Site co-link analysis applied to small networks: a new methodological approach. **Scientometrics**, Budapest, v. 83, n. 1, p. 157-166, 2010.

METTROP, Wouter ; NIEUWENHUYSEN, Paul. Internet search engines – fluctuations in document accessibility, **Journal of Documentation**, London, v. 57, n. 5, p. 623-651, 2001.

ORTEGA, Jose Luis et al. Maps of the academic web in the European Higher Education Area: an exploration of visual web indicators. **Scientometrics**, Budapest, v. 74, n. 2, p. 295-308, 2008.

PAYNE, Nigel ; THELWALL, Mike. A Longitudinal study of academic webs: growth and stabilization. **Scientometrics**, Budapest, v. 71, n. 3, p. 523-539, 2007.

PRIME, Camille ; BASSECOULARD, Elise ; ZITT, Michel. Co-citations and co-sitations: a cautionary view on an analogy. **Scientometrics**, Budapest, v. 54, n. 2, p. 291-308, 2002.

ROUSSEAU, Ronald. Daily time series of common single word searches in AltaVista and Northern Light. **Cybermetrics**, Madrid, v. 2, n. 1, 1999. Disponível em: <<http://cybermetrics.cindoc.csic.es/articles/v2i1p2.html>> Acesso: 30 set. 2012.

SMITH, Alastair G. Web links as analogues of citations. **Information Research**. Sweden, v.9, n.4, artigo 188, 2004. Disponível em: <<http://informationr.net/ir/9-4/paper188.html>> Acesso: 30 set. 2012

STATISTICAL CYBERMETRICS RESEARCH GROUP. Search Engine Queries for Webometrics. 2012. Disponível em: <<http://cybermetrics.wlv.ac.uk/QueriesForWebometrics.htm>> Acesso: 30 set. 2012.

THELWALL, Mike. A Web crawler design for data mining. **Journal of Information Science**, London, v. 27, n. 5, p. 319-325, 2001.

THELWALL, Mike ; SUD, Pardeep. A Comparison of methods for collecting web citation data for academic organisations. **Journal of the American Society for Information Science and Technology**, New York, v. 62, n. 8, p. 1488–1497, 2011.

- THELWALL, Mike; SUD, Pardeep; WILKINSON, David. Link and co-inlink network diagrams with URL citations or title mentions. **Journal of the American Society for Information Science and Technology**, New York, v. 63, n. 4, 805-816, 2012.
- THELWALL, Mike ; TANG, Rong ; PRICE, Liz. Linguistic patterns of academic Web use in Western Europe. **Scientometrics**, Budapest, v. 56, n. 3, p. 417-432, 2003.
- THELWALL, Mike ; VAUGHAN, Liwen ; BJÖRNEBORN, Lennart. Webometrics. **Annual Review of Information Science and Technology**, Medford, NJ, v. 39, n. 1, p. 81-135, 2005.
- VANTI, Nadia Aurora Peres. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, Brasília, DF, v. 31, n. 2, p.152-162, 2002.
- VAUGHAN, Liwen. Visualizing linguistic and cultural differences using Web co-link data. **Journal of the American Society for Information Science and Technology**, Medford, NJ, v. 57, n. 9, p. 1178-1193, 2006.
- VAUGHAN, Liwen ; GAO, Yijun ; KIPP, Margaret. Why are hyperlinks to business websites created? a content analysis. **Scientometrics**, Budapest, v. 67 n. 2, p. 291-300, 2006.
- VAUGHAN, Liwen; KIPP, Margaret; GAO, Yijun. Are co-linked business websites really related? a link classification study. **Online Information Review**, Bradford, v. 31 v. 4, p. 440-450, 2007.
- VAUGHAN, Liwen ; TANG, Juan ; DU, Jian. Examining the robustness of web co-link analysis. **Online Information Review**, Bradford, v. 33 n: 5 p. 956-972, 2009.
- VAUGHAN, Liwen ; YOU, Justin. Comparing business competition positions based on Web co-link data: The global market vs. the Chinese market. **Scientometrics**, Budapest, v. 68, n. 3, p. 611-628, 2006.
- _____. Word co-occurrences on Webpages as a measure of the relatedness of organizations: a new Webometrics concept. **Journal of Informetrics**, New York, v. 4, n. 4, p. 483-491, Oct. 2010.
- YAHOO! Web Search APIs from Yahoo! Search. 2012. Disponível em: <<http://developer.yahoo.com/search/web/webSearch.html>> Acesso: 30 set.2012.
- ZUCCALA, Alesia. Author cocitation analysis is to intellectual structure as web colink analysis is to... ? **Journal of the American Society for Information Science and Technology**, Medford, NJ, v. 57, n. 11, p. 1487-1502, 2006.

Fábio Castro Gouveia

Doutor em Ciências (Educação, Gestão e Difusão em Biociências) pela Universidade Federal do Rio de Janeiro (UFRJ).

Tecnologista em Saúde Pública da Fundação Oswaldo Cruz (Fiocruz).

E-mail: fgouveia@fiocruz.br

Recebido em: 30/09/2012

Aceito em: 01/11/2012