

Revisitando a Lei de *Mooers*: a recuperação de artigos de periódicos científicos eletrônicos

Revisiting Mooers' Law: the retrieval of articles in a scientific electronic journal

Autor

Guilherme Ataíde Dias¹

Resumo

Este artigo discute como o posicionamento físico dos *links* para artigos de periódicos científicos eletrônicos na área da Ciência da Informação, disponibilizados na *World Wide Web* afetam a escolha dos mesmos. O comportamento dos usuários, na recuperação dos artigos foi mapeado através dos arquivos de *log* de acesso gerados pelos servidores *WEB* de três periódicos. Desenvolvemos as conclusões relativas ao acesso dos artigos, à luz de uma proposta alternativa à interpretação da Lei de *Mooers*, a qual apresentamos e discutimos.

Palavras-chave

**LEI DE MOOERS
PERIÓDICOS ELETRÔNICOS
ESTUDOS DE USUÁRIO
SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO ELETRÔNICOS**

¹ Bacharel em Ciência da Computação. Doutorando em Ciência da Informação na ECA/USP. Agradecemos a Profª. Dra. Dinah Población pelas sugestões e revisão.

Introdução

O estudo do comportamento de um usuário de informação quando, este faz uso de um sistema de recuperação de informação eletrônico, é deveras importante, pois contribui para um melhor entendimento das dinâmicas relacionadas à utilização desse tipo de sistema. Estudos relacionados a este tema têm estado em evidência mais recentemente, devido à grande proliferação e exposição a que temos nos submetido aos sistemas de recuperação de informação eletrônicos disponíveis na *WEB*.

Neste estudo procuramos entender como o usuário faz suas escolhas quando apresentado a uma seqüência de opções de acesso a informações em um dispositivo de saída gerado por um sistema de recuperação eletrônico. Mais precisamente, objetivamos entender como os usuários das revistas científicas eletrônicas brasileiras na área da Ciência da Informação fazem as suas escolhas ao acessarem uma página de um periódico onde são disponibilizados diversos *links* para os artigos (Figura 1). Tentaremos responder à seguinte questão: O local onde estão posicionados fisicamente os *links* para artigos dentro de um periódico científico eletrônico tem influência na sua recuperação?

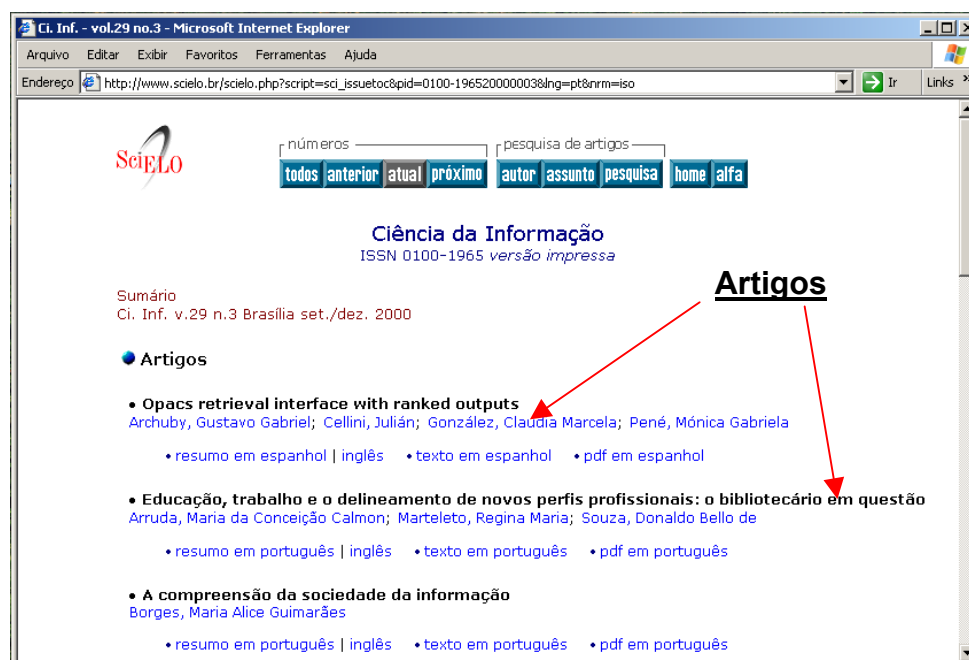


Figura 1 – Periódico eletrônico *Ciência da Informação* – Exemplo da disposição de artigos

Consideramos aqui, que um periódico científico eletrônico disponibilizado na *WEB* não deixa de ser um sistema de recuperação de informações. Conforme Mooers (1951) “*A recuperação de informação engloba os aspectos intelectuais da descrição da informação e de sua especificação para a busca, bem como qualquer sistema, técnica ou máquina que são utilizadas para realizar a operação*”. Dos três periódicos que utilizamos neste estudo - relacionaremos cada um deles mais adiante -, dois disponibilizam ferramentas de busca à informação².

Encontramos na literatura da área estudos anteriores sobre como os usuários acessavam a informação através dos sistemas de recuperação de informação eletrônicos. Veremos a seguir dois exemplos: #1 - De acordo com o estudo realizado por Friedman (1996), quando um agente de reservas de uma determinada empresa aérea solicita um itinerário para um cliente, o sistema de reservas (o SABRE³ seria um exemplo) faz uma busca em uma base de dados e recupera todos os itinerários factíveis, baseado em critérios de busca pré-determinados, tais como rotas mais diretas, menor tempo de voo, etc. Os itinerários apresentados na primeira tela apresentam uma grande probabilidade de serem os escolhidos. Taib (1990) cita que 90% dos voos selecionados estão listados na primeira tela apresentada. Sem levarmos em consideração que um agente de uma determinada empresa aérea só vai selecionar voo para a sua empresa, existirá após algumas telas, os itinerários de outras empresas aéreas, que eventualmente podem apresentar melhores itinerários para o cliente; #2 – Brown (1999) realizou uma pesquisa com estudantes de Biblioteconomia e Ciência da Informação na *University of Washington* sobre como os mesmos se comportavam na escolha de *links* previamente preparados para a pesquisa. Dentre outros achados, esta pesquisa revelou que os estudantes tinham uma tendência de selecionar o primeiro *link* apresentado quando o mesmo era seguido por uma sequência de *links* similares, não importando se os *links* visualmente eram ou não do mesmo tamanho. Nos casos em que existia um *link* cuja escolha era a melhor opção dada uma determinada tarefa de busca, tal como “Ache a lista das próximas palestras disponibilizadas neste *site*”,

² *Informação & Sociedade: estudos e Ciência da Informação*

³ Sistema de reservas da *American Airlines*

os usuários procediam de forma a escolher o *link* correto, independente deste *link* estar apresentado logo no início ou ser necessário passar a tela algumas vezes até achar a informação solicitada.

A lei de Mooers

Calvin N. Mooers foi um dos pioneiros na Ciência da Informação e responsável por diversas inovações nesta área bem como na Ciência da Computação. De acordo com Garfield (1997), Mooers é provavelmente mais conhecido por ter criado o termo *information retrieval* durante o período em que escrevia a sua tese de mestrado no *Massachusetts Institute of Technology*. Mooers (1960) estabeleceu um princípio ao qual atribuiu o seu próprio nome, este princípio chama-se a Lei de Mooers e estabelece que:

“Um sistema de recuperação de informações terá a tendência de não ser usado se é mais irritante e problemático para um usuário obter a informação do que não obtê-la”.

Garfield (1997) propõe um corolário para a lei acima apresentada:

“Quanto mais informação relevante um sistema de recuperação de informação provê, mais usado este sistema será”.

Assim como em Garfield (1997), encontramos muitas citações de outros autores à Lei de Mooers. Baseado nesta lei, estritamente como enunciada acima, podemos tentar explicar os porquês do uso ou não uso dos sistemas de recuperação de informações que atualmente proliferam na *Internet*. Quantas vezes já não tivemos a oportunidade de nos confrontar com um número infindável de *links* com informações similares resultantes de uma pesquisa retornada por um sistema de recuperação de informações e abandonamos a busca na primeira tela, clicando apenas no primeiro ou segundo *link* retornado e perdendo a seguir o interesse!

Encontramos mais recentemente na literatura alguns questionamentos relacionados à Lei de Mooers, principalmente no que condiz a sua interpretação. Segundo Austin (2001), a interpretação da mesma deveria ser feita dentro do contexto do artigo original em que a lei foi apresentada e não dissociada deste artigo como ocorre correntemente. Ainda de acordo com este mesmo autor, a Lei de Mooers tomada com verdadeira pelos profissionais de informação não é a mesma que Mooers propôs.

Conforme Austin (2001), - “Mooers explica que a sua lei foca nos aspectos irritantes e problemáticos de ter a informação em mãos e, portanto sentir-se obrigado a fazer alguma coisa com a mesma”. Em seu artigo, Mooers (1960) explica que um indivíduo de posse da informação primeiro precisa lê-la, precisa entendê-la e conseqüentemente pensar sobre a mesma, tarefas estas que nem sempre são triviais e que podem forçar o indivíduo a tomar ações nem sempre agradáveis, tais como, a mudança na forma pelas quais os processos são realizados. Desta maneira, podemos considerar que a Lei de Mooers está relacionada ao processamento e uso que será feito da informação pelo usuário e não necessariamente voltada para a funcionalidade do sistema de recuperação de informações, e que a posse da informação longe de ser a panacéia universal representa um problema.

Austin (2001) propõe uma extensão à lei original de Mooers, na seguinte forma:

Primeira Lei de Mooers:

“Em um ambiente no qual é mais irritante e problemático para um usuário ter a informação em mãos do que não tê-la, um sistema de informações terá a tendência de não ser usado”;

Segunda Lei de Mooers:

“Em um ambiente no qual é absolutamente crítico para um usuário ter informação, um sistema de informação terá a tendência de ser utilizado, não importa o quanto deficiente o mesmo seja”;

Terceira Lei de Mooers:

“Em um ambiente no qual o problema de ter informação versus o de não tê-la estão razoavelmente equilibrados, o fator decisivo se um sistema de informação será utilizado ou não será o design e a performance do mesmo”.

Utilizaremos estas extensões da lei original para explicar o comportamento do usuário quando do acesso a artigos de um periódico científico eletrônico. Inicialmente detalharemos o caminho desenvolvido para a obtenção dos dados (*logs* de acesso) dos periódicos científicos analisados e, num segundo momento analisaremos os mesmos.

Metodologia

Selecionamos para esta pesquisa os três seguintes periódicos científicos eletrônicos na área da Ciência da Informação: *Informação & Sociedade: Estudos; Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação*, e *Ciência da Informação*. Seria plenamente factível estudar todos os periódicos científicos da área, e este era o nosso intuito inicial, contudo, em virtude de impedimentos alegados por alguns editores em fornecer informações, não conseguimos aumentar o nosso campo de estudo.

Os números dos periódicos utilizados na pesquisa estão relacionados na tabela a seguir:

<i>Informação & Sociedade: Estudos</i>	<i>Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação</i>	<i>Ciência da Informação</i>
Vol. 9, no.1	Set. 1998, n.6	Vol.28, no. 1
Vol. 9, no. 2	Abr. 1999, n.7	Vol.28, no. 2
Vol. 10, no. 1	Set. 1999, n.8	Vol.28, no. 3
Vol. 10, no. 2	Jun. 2000. n.9	Vol.29, no. 1
Vol. 11, no. 1	Out. 2000. n.10	Vol.29, no. 2
	Jun. 2001. n.11	Vol.29, no. 3
		Vol.30, no. 1
		Vol.30, no. 2
		Vol.30, no. 3

Tabela 1 – Número dos periódicos estudados.

Os arquivos de *log* de acesso dos periódicos científicos eletrônicos *Informação & Sociedade: Estudos* e *Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação* foram obtidos através de solicitação pessoal, informal ao Núcleo de Tecnologia da Informação (NTI) da Universidade Federal da Paraíba - UFPB e ao Núcleo de Processamento de dados da Universidade Federal de Santa Catarina – UFSC, respectivamente.

O arquivo de *log* de acesso do periódico científico eletrônico *Informação & Sociedade: Estudos* obtido continha dados de acesso ao periódico durante o período compreendido entre 1 de março de 2000 e 30 de setembro de 2001. Este arquivo estava de acordo com o formato *Common Logfile Format - CLF*.

O arquivo de *log* de acesso do periódico científico eletrônico *Estudos e Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação* obtido continha dados de acesso ao periódico durante o período compreendido entre 3 de maio e 2 de agosto de 2001. O arquivo estava no formato *Extended Logfile Format – ELF*.

A diferença de datas no início da coleta de dados entre os dois periódicos científicos eletrônicos acima mencionados ocorre, pois os administradores de sistema nem sempre possuem todos os arquivos de *log* de acesso armazenados, devido a restrições de espaço em mídia ou mesmo pelo desinteresse em armazenar informações que as organizações não consideram importantes.

Quando solicitamos o arquivo de *log* de acesso do periódico científico eletrônico *Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação* ao núcleo de processamento de dados da Universidade Federal de Santa Catarina, não conseguimos cópias deste arquivo anteriores à data de nossa solicitação inicial (maio de 2001), visto que, após um período de tempo, este arquivo é removido do servidor por restrições de espaço e não é armazenado em nenhum outro local. Neste caso para termos o arquivo de *log* de acesso que cobrisse um período de tempo mais prolongado foi necessário solicitarmos dentro de períodos de tempo pré-estabelecidos uma cópia do arquivo de *log* de acesso e armazená-lo em uma mídia adequada para um posterior processamento.

Após o mês de julho de 2001, paramos de receber o arquivo de *log* de acesso que nos era enviado periodicamente por um técnico do núcleo de processamento de dados da UFSC. A interrupção no envio dos *logs* coincidiu com o início de uma greve de funcionários das instituições federais de ensino superior brasileiras. Tentamos entrar em contato para restabelecer o envio dos *logs*, mas não obtivemos nenhuma resposta, provavelmente devido à ausência dos funcionários envolvidos com o movimento grevista.

Os dados concernentes ao acesso ao periódico científico eletrônico *Ciência da Informação* não foram extraídos diretamente do arquivo de *log* de acesso, mas através de ferramenta automatizada, disponibilizada de forma pública e gratuita no próprio *site* do SciELO - *Scientific Electronic Library Online*. O referido periódico também está hospedado no *site* SciELO, além do *site* próprio IBICT. As informações compiladas nos relatórios gerados pela ferramenta são relativas ao período entre 01 de janeiro de 1999 e 31 de março de 2002.

Antes de optarmos por utilizar os dados disponibilizados através do SciELO, fizemos uma solicitação através de correspondência enviada diretamente ao Presidente do Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT com o intuito de conseguir uma autorização para obtermos uma cópia do arquivo de *log* de acesso do periódico científico *Ciência da Informação*, contudo não fomos atendidos da forma que exatamente desejávamos. Segundo a representante do IBICT, que entrou em contato conosco, o arquivo de *log* de acesso solicitado não estava individualizado, mas misturado com todos os outros registros gerados pelo servidor *WEB* do IBICT, sendo desta forma inviável o repasse destes dados, pois seriam remetidas também todas as outras informações relativas ao *site WEB* do referido instituto.

Os arquivos de *log* dos periódicos *Informação & Sociedade: Estudos e Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação* foram tratados previamente com o auxílio do utilitário *grep* incluído no ambiente *GNU/Linux* e processados posteriormente pelo *software Webtrends 7.0* disponível para a plataforma *Windows*, através do qual geramos uma tabela para cada número de periódico estudado com o número de vezes que cada artigo de periódico foi acessado bem como a posição física que cada artigo ocupava no mesmo. Decidimos não incluir aqui todas as tabelas geradas devido a restrições de espaço, mas apenas os gráficos gerados a partir das tabelas.

Faz-se ainda necessário mencionar que através dos arquivos de *log* de acesso utilizados, não foi possível determinar exatamente a população dos usuários que acessam os periódicos estudados, mas pelo tipo de conteúdo existente nos periódicos científicos eletrônicos na área da Ciência da Informação, podemos supor que os usuários seriam preferencialmente: pesquisadores, professores, profissionais técnicos, estudantes de graduação e de pós-graduação da área.

Interpretação e apresentação dos dados coletados

O gráficos de linha apresentados nas figuras 2, 3 e 4 mostram o número de acessos feitos por usuários a cada número de periódico. No eixo do *y* temos o número de acessos,

no eixo do x os artigos, obedecendo à posição física que cada um deles ocupa nos periódicos.

Os gráficos representados na Figura 2, relativos ao periódico *Informação & Sociedade: Estudos*, não apresentam tendências que determinem a predileção dos usuários por *links* para artigos posicionados em nenhum local determinado do periódico, seja no começo, meio ou fim do mesmo. O gráfico que apresenta um comportamento de menor variação é o relativo a Seqüência 5 (Vol. 11, no. 1), o periódico associado a este gráfico é o mais novo disponibilizado para acesso no *site* hospedeiro, tendo sido exposto menos tempo aos usuários. Desta forma, podemos não ter um comportamento de acesso ao mesmo mais definido.

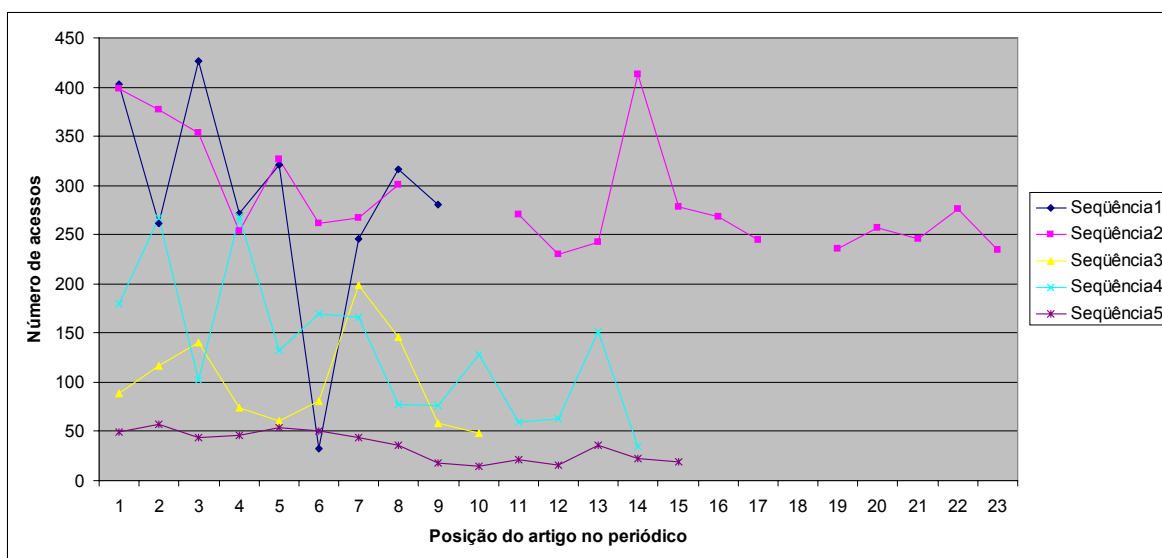


Figura 2 – Informação & Sociedade: Estudos – Número de acessos X artigo

Seqüência 1	Vol. 9, no. 1
Seqüência 2	Vol. 9, no. 2
Seqüência 3	Vol. 10, no. 1
Seqüência 4	Vol. 10, no. 2
Seqüência 5	Vol. 11, no. 1

O periódico *Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação*, cujo comportamento de acesso aos seus artigos está representado nos gráficos da Figura 3, também não apresenta tendências para que afirmemos que os *links* para artigos

posicionados em um determinado local tenham uma maior ou menor tendência de serem mais ou menos acessados. Este periódico é o que apresentou o menor número de artigos por periódico. Dos números estudados, a maior quantidade de artigos em um número foi seis. Dos artigos contidos nos periódicos, o artigo cujo *link* está na última posição (quarto lugar) do periódico representado pela Seqüência 2, (Abr. 1999, n.7) apresentou um número de acessos bem superior a todos os demais.

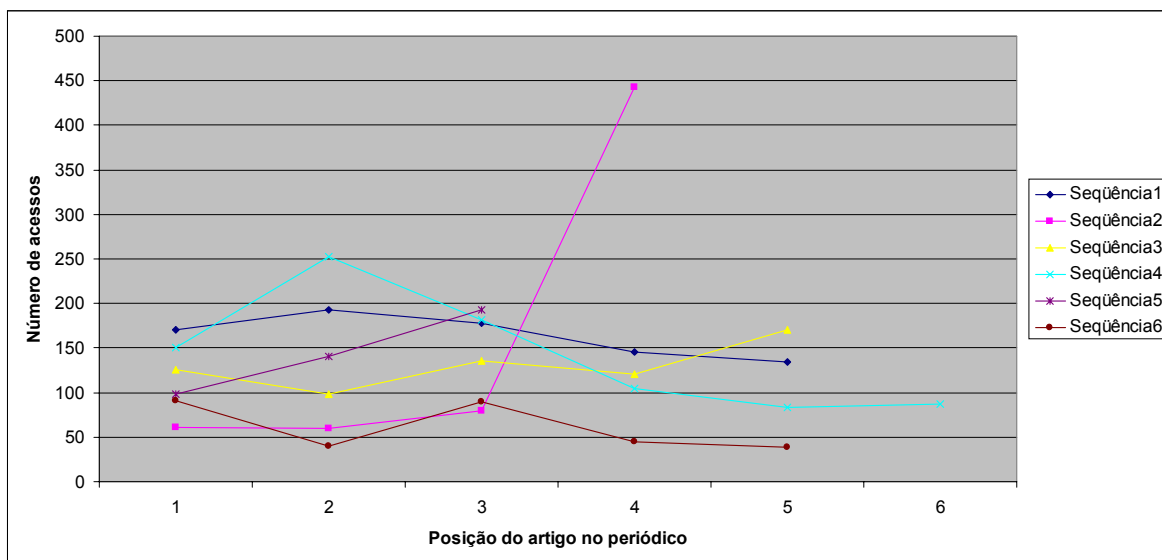


Figura 3– Periódico *Encontros Bibli* - Número de acessos X Artigo

Seqüência 1	Set. 1998, n.6
Seqüência 2	Abr. 1999, n.7
Seqüência 3	Set. 1999, n.8
Seqüência 4	Jun. 2000, n.9
Seqüência 5	Out. 2000, n.10
Seqüência 6	Jun. 2001, n.11

Os gráficos associados ao periódico *Ciência da Informação*, representados na Figura 4, também não indicam tendências que apontem a predileção dos usuários por *links* para artigos posicionados em nenhum local específico do periódico. Da mesma forma que na Figura 2 (Seqüência 5) e na Figura 3 (Seqüência 6) o número do periódico (Vol.30, no. 3) disponibilizado por último, no *site* hospedeiro, também apresenta uma pequena variação numérica na quantidade de acessos em relação aos outros números de periódicos.

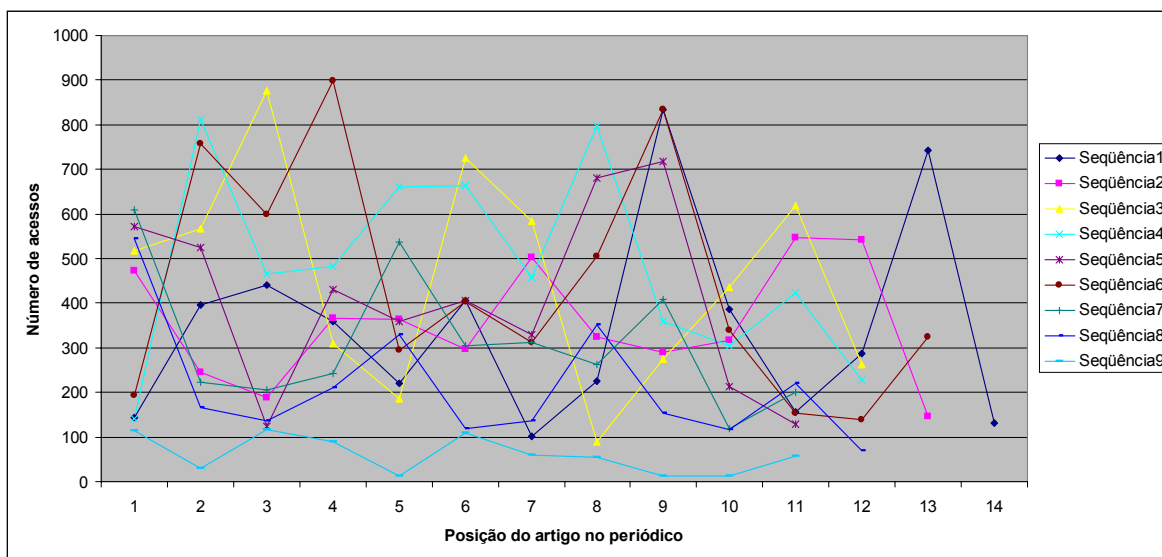


Figura 4– Periódico *Ciência da Informação* - Número de acessos X Artigo

Sequência 1	Vol.28, no. 1
Sequência 2	Vol.28, no. 2
Sequência 3	Vol.28, no. 3
Sequência 4	Vol.29, no. 1
Sequência 5	Vol.29, no. 2
Sequência 6	Vol.29, no. 3
Sequência 7	Vol.30, no. 1
Sequência 8	Vol.30, no. 2
Sequência 9	Vol.30, no. 3

Podemos verificar que todos os gráficos apresentados, de uma maneira geral, não mostram qualquer comportamento que sugira um maior número de acesso aos artigos posicionados no início de cada periódico. Pelo contrário, interpretando os mesmos, encontramos um comportamento completamente errático quando do acesso do usuário aos artigos disponibilizados nos periódicos.

Na hipótese de que os artigos posicionados fisicamente no início do periódico tivessem um maior acesso pelos usuários do que os artigos posicionados posteriormente, os gráficos correspondentes teriam um comportamento decrescente ao longo do eixo dos *x*. Inicialmente os valores de *y* seriam altos, diminuindo à medida que fossemos em direção aos *links* dos artigos mais distantes do início.

Um fato que deparamos e que chamou a atenção foi o baixíssimo acesso a todos os artigos não escritos na língua portuguesa⁴. Artigos em outras línguas, com os respectivos *links* colocados nas mais diversas posições do periódico, foram de forma inquestionável os menos acessados.

Conclusões

A partir dos resultados obtidos dos dados extraídos dos *logs*, demonstrando a não existência de um padrão regular na recuperação dos artigos, podemos levantar a hipótese de que os usuários de periódicos científicos eletrônicos, de uma maneira geral, devido ao seu perfil especializado, buscam informações igualmente especializadas. Assim a posição física dos *links* em um periódico científico eletrônico não afeta a recuperação dos mesmos. O elemento orientador no acesso aos artigos seria o tema e/ou o assunto tratado no artigo.

Considerando a não existência de um padrão regular no acesso aos artigos, podemos inferir que o usuário se dispôs a pesquisar o artigo desejado, seja através de recursos automáticos disponibilizados pela revista ou simplesmente através da realização de uma busca seqüencial visual entre os *links* apresentados, ao invés de simplesmente selecionar os primeiros *links* disponibilizados no periódico. Tomaremos a liberdade de utilizarmos a segunda Lei de Mooers conforme proposto por Austin (2001), para darmos suporte a estas idéias. Ou seja, o usuário vai fazer uso de um sistema de recuperação de informações não importando quão mal projetado ou deficiente o mesmo seja, desde que a informação requerida mostre-se crítica para o usuário.

Acreditamos que se o resultado obtido a partir dos dados nos mostrasse que os usuários acessavam os periódicos e faziam as suas seleções centradas em cima dos primeiros *links* para os artigos, teríamos um comportamento de quem não tem realmente um determinado interesse nas informações lá disponibilizadas, pelo fato das mesmas não serem realmente relevantes ou, possivelmente, pelos problemas causados pela possibilidade

⁴ Encontramos outros artigos escritos em espanhol e inglês.

de ter estas informações em mãos, tal qual explicitado na primeira Lei de Mooers, conforme proposto por Austin (2001).

Finalmente, a terceira Lei de Mooers, conforme proposto por Austin (2001), nos faz refletir sobre a importância do design e da performance dos sistemas de informações como fator decisivo para a utilização efetiva ou não dos mesmos, pois, em última instância, estes serão os fatores determinantes do uso do sistema, sempre que existir um equilíbrio entre o problema de ter ou deixar de ter a informação. Neste artigo, não nos dedicamos ao estudo de questões relativas ao design e a performance dos sistemas aos quais estão associados os periódicos científicos eletrônicos em questão. Sugerimos que outros estudos devem ser conduzidos para desvendar as dinâmicas relacionadas a este tema.

A constatação feita, durante a análise, dos dados da existência de um baixíssimo acesso aos artigos escritos em outras línguas que não a portuguesa, serve para levantar a hipótese de que os usuários dos periódicos analisados possuem dificuldades na compreensão de línguas estrangeiras. Este fato não deveria acontecer, pois embora impossibilitados de identificar exatamente os usuários das revistas através do *log* de acesso, podemos inferir que os mesmos são profissionais da informação com alguma qualificação acadêmica, em virtude dos temas tratados nos periódicos. O domínio de um idioma estrangeiro deveria ser fundamental, levando-se em consideração que grande parte da literatura da área não está em língua portuguesa.

Abstract

This article discusses how the retrieval of articles in an Information Science scientific electronic journal, made available at the Word Wide Web, is affected by the physical position of the links to the articles. The users' behavior while retrieving the articles were studied through the log files generated by the web servers of three journals. The conclusions regarding the analysis of the log files were developed in the light of an alternative interpretation of Mooers' Law, which we present and discuss.

Referências Bibliográficas

AUSTIN, Brice. Mooer's Law: In and Out of Context. *Journal of the American Society for Information Science*, v 52(8):607-609. 2001.

BROWN, Fred , **TIMMONS**, Traci & **WILDER**, Heather, Selection Behavior on the World Wide Web for LIS Students. Artigo apresentado ao ASIS Pacific Northwest Chapter Student Paper Competition, University of Washington, Seattle, Washington, 1999.

FRIEDMAN, Bataya & **NISSENBAUM**, Helen. Bias in Computer Systems. *ACM Transactions on Information Systems*. Vol. 14, No.3, July 1996, Pages 330-347.

GARFIELD. Eugene. A Tribute To Calvin N. Mooers, A Pioneer Of Information Retrieval. *The Scientist*, Vol: 11(4) March 17, 1997.

MOOERS, Calvin N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*. 2(1), p.20-32, 1951.

MOOERS, Calvin N. Mooers' law or, Why Some Retrieval Systems Are Used and Others Are Not. *American Documentation*. 11 (3): ii, July 1960.

TAIB, I. M. Loophole allows bias in displays on computer reservations systems. *Avia. Week Space Tech*. Feb. 1990.