

## USO E DESEMPENHO DE WEB SITES: UM ESTUDO DE CASO

Maria de Cléofas Faggion Alencar<sup>1</sup>

Ivo Pierozzi Júnior<sup>2</sup>

Eliane Gonçalves Gomes<sup>3</sup>

Carlos Alberto de Carvalho<sup>4</sup>

### RESUMO

A Internet é um meio de comunicação remota, inovador, extremamente dinâmico e, cada vez mais, amplamente acessível, que tem adotado uma linguagem própria, inerente à sua natureza multimídia. A Embrapa Monitoramento por Satélite, há mais de uma década, utiliza a Internet como meio de difusão de resultados de pesquisa e de interação com clientes, parceiros e usuários. Com intuito de avaliar o uso do *web site* por esse público e o desempenho do sistema de comunicação eletrônica envolvido, tem-se utilizado o programa *Webalizer* que monitora e elabora estatísticas de entradas e saídas do sistema, com base na análise de arquivos de *log*. Através dos indicadores contabilizados pelo programa foi possível avaliar aspectos sobre a origem dos acessos, as ações executadas pelos usuários e sobre aspectos do desempenho do sistema em termos de respostas enviadas às solicitações dos usuários. Os resultados podem possibilitar a remodelagem do *web site*, de modo a melhorar a dinâmica de interação instituição-usuário e promover o desenvolvimento de uma abordagem própria de análise de *logs*.

---

<sup>1</sup> Doutora em Ciência da Informação, Bibliotecária da Embrapa Monitoramento por Satélite, cleo@cnpm.embrapa.br

<sup>2</sup> Doutor em Ecologia, Chefe de P&D da Embrapa Monitoramento por Satélite, ivo@cnpm.embrapa.br

<sup>3</sup> Mestre em Pesquisa Operacional, Pesquisadora da Embrapa Monitoramento por Satélite, eliane@cnpm.embrapa.br

<sup>4</sup> Analista de Sistemas, Suporte à Pesquisa, Embrapa Monitoramento por Satélite, calberto@cnpm.embrapa.br

## ABSTRACT

The Internet is a remote, innovative, extremely dynamic and widely accessible communication mean. As in all others human communication formats, we observe the development and adoption of a proper language, inherent to its multimedia aspects. The Embrapa Satellite Monitoring uses for more than one decade the Internet as a dissemination mean of its research results and interaction with clients, partners and *web site* users. In order to evaluate the *web site* usage and performance of the e-communication system it has been used Webalizer software to track and to calculate statistics based on web server log file analysis. The objective of the study is to analyse the data and evaluate the indicators related to requests origin (search *string*, country, time), actions performed by users (entry pages, agents) and system performance (error messages). It will help to remodel the *web site* design to improve the interaction dynamics and also develop a customised log file analyser. This tool would retrieve information coherent and real.

## 1. INTRODUÇÃO

A Internet tornou-se, inegavelmente, o mais inovador, democrático e completo meio de comunicação até agora desenvolvido. Nos dias de hoje, praticamente não encontra mais barreiras de expansão criativa e de abrangência temática.

Como em todas as outras formas de relacionamento humano, a Internet desenvolve e adota linguagens e códigos próprios, inerentes à sua natureza multimídia. A interatividade é outra característica singular da Internet, conferindo-lhe uma dinâmica própria e que a difere de outras formas de comunicação remota, com a possibilidade de contatos em tempo real reunindo, num só veículo, imagens, sons e textos.

A Embrapa Monitoramento por Satélite presta assessoria a órgãos públicos e à iniciativa privada em problemas e desafios agropecuários e ambientais, através da pesquisa,

adaptação, avaliação e disponibilização de conhecimentos e informações originados de aplicações de geotecnologias. Desde 1991, tem utilizado a Internet não apenas como um veículo passivo de difusão do conhecimento e das informações geradas através de seus projetos de pesquisa (PIEROZZI Jr; CAPUTI; FILARDI, 2000) porém, além disso, tem investido e acompanhado a rápida evolução tecnológica que envolve o setor, transformando seu *web site*, de mera vitrine de resultados para um canal dinâmico e ativo de contato com usuários, clientes e parceiros, que solicitam produtos, serviços e ações de monitoramento por satélite.

Ciente da importância da opinião, expectativas e satisfação desse público em relação ao conteúdo e forma das informações disponibilizadas, algumas iniciativas de avaliação do uso ao *web site* foram implementadas. Esses procedimentos visam aprimoramentos na arquitetura da informação e na gestão do conteúdo que, segundo NAKANO (2002), envolve desenvolvimento, gerenciamento, manutenção e organização das soluções de conteúdo da *web*. Uma das iniciativas foi o monitoramento dos acessos ao *web site* por *web trackings*, mecanismos de análise de arquivos de *log*. Essas ferramentas geram vários relatórios para análise e compreensão de variáveis relacionadas à dinâmica do uso de *web sites*, podendo ser encontradas na Internet (HJELM, 1999).

Os arquivos de *log* oferecem aos operadores de *web sites* detalhamento substancial dos visitantes sendo, atualmente, bastante diversificados e qualitativamente utilizados.

O presente trabalho apresenta e discute os resultados da análise de três anos e meio de monitoramento do uso do *web site* da Embrapa Monitoramento por Satélite, visando traçar um perfil de sua dinâmica e obter indicadores para ajustes e melhorias na concepção, disponibilização ou manutenção das informações. Para tal, foram usados dados sobre a dinâmica e origens do uso, temporizados e caracterizados pelos *strings* e por países; sobre as

ações de uso, caracterizadas pelas páginas de entrada e pelos agentes; e sobre o desempenho do *web site*, a partir dos códigos de erros.

## **2. ANÁLISE DE ARQUIVOS DE LOG**

Arquivos de *log* são arquivos texto gerados por um *software* que registra os eventos e a hora em que ocorrem em um determinado sistema computacional (YU; APPS, 2002). Um arquivo de *log* típico relata, em um formato predefinido (LOG FILES, 2002), que ação foi executada, quando, onde, que informação é recuperada, em que formato e como. É possível ainda identificar o autor da ação (computador que gerou a solicitação). A geração de arquivos de *log* é feita sem obstrução, de forma mais objetiva do que uma coleta de dados através de outras fontes. Se processados apropriadamente, podem ser calculadas estatísticas de uso, que fornecem dados úteis para estudos de usuários e gerenciamento da informação.

Na literatura são encontrados diversos estudos sobre o uso de arquivos de *log* para a recuperação da informação *on line* (BERTOT ET AL., 1997; CATLEDGE; PITKOW, 1995; DIAS, 2002; HOCHHEISER; SHNEIDERMAN, 2001; LEE; HELLER, 1997; MURPHY ET AL., 2001; PETERS, 1993, 1998; YU; APPS, 2000).

## **3. MATERIAL E MÉTODOS**

Os dados para a avaliação do uso do *web site* da Embrapa Monitoramento por Satélite foram os da série histórica de relatórios gerados pelo *Webalizer* entre janeiro de 1999 a julho de 2002. O *Webalizer* (THE WEBALIZER, 2002) é um programa de uso livre (*free software* – GNU, 2002) de análise de *log files*, que gera estatísticas mensais, diárias e horárias de uso, categorizadas segundo indicadores como, por exemplo, número de visitas, país de origem, *strings* de busca etc.

As definições desses indicadores foram propostos por BARRETT (1997). Uma página é formada por vários componentes, cada qual sendo um arquivo. Esses arquivos são agregados através de comandos da linguagem HTML. Outras definições importantes (BACCALA, 1997; EICHMANN; WU, 1996; THE WEB, 2002) são necessárias:

- *Web site*: um conjunto de páginas interligadas e tratando de um mesmo tema;
- Hipertexto: a ligação entre páginas. O hipertexto está presente na sintaxe do HTML e permite a ligação entre as páginas do *web site* ou de qualquer outro ponto da Internet;
- Servidor: o computador que abriga fisicamente o *web site*;
- Cliente: computador que faz um pedido ao *web site*. No *Webalizer*, os termos servidores e clientes são sinônimos;
- Códigos de Status (*Status Code*): São códigos de 3 dígitos que expressam o resultado de um pedido (*request*) de arquivo ao *web site*:
  - Códigos de Sucesso
    - Tipo 1xx: *Informational* – O pedido ocorreu perfeitamente e será dada continuidade ao processo;
    - Tipo 2xx: *Success* – O pedido ocorreu perfeitamente e foi enviado o arquivo ao solicitante;
    - Tipo 3xx: *Redirection* – Mais ações além do pedido são necessárias para que o arquivo seja enviado ao solicitante.
  - Códigos de erro
    - Tipo 4xx: *Client Error* – O pedido não foi feito corretamente devido a algum erro de sintaxe, por exemplo, nome de arquivo errado ou inexistente;
    - Tipo 5xx: *Server Error* – O pedido não foi aceito devido a falha do servidor que abriga o *web site*;

- Agentes: tecnicamente são conhecidos como *User-agents*. São programas utilizados por um ou mais usuários para facilitar a busca e recuperação de informações na *web*. Os mais conhecidos agentes são os navegadores para *web* (*web user-agents* ou *web browsers*) como o *Netscape Navigator* e *Microsoft Internet Explorer*;
- Robô: é um programa que vasculha automaticamente a estrutura de hipertexto da *web* procurando por um arquivo e continua buscando recursivamente todos os arquivos que são referenciados nessa estrutura;
- Arquivo de *log*: estrutura utilizada para armazenar uma série de informações a respeito do arquivo que foi requisitado ao servidor e qual cliente ou usuário o requisitou.

Quando da decisão de se utilizar uma ferramenta para análise de arquivos de *log*, a ferramenta *Webalizer* já se encontrava ativada no *web site* da Embrapa Monitoramento por Satélite. Para manter a objetividade, homogeneidade e coerência das informações essa ferramenta não foi substituída. Além disso, dentre as ferramentas existentes, o *Webalizer* permite um maior acesso aos dados, quando comparado com outros pacotes que sumarizam e analisam as informações, deixando ao usuário apenas a possibilidade de interpretação passiva. Entretanto, devido às suas particularidades, que limitam a amplitude de conclusões possíveis sobre a dinâmica de uso do *web site*, detalhes específicos somente podem ser obtidos com a análise direta dos arquivos de *log*.

Dentre o conjunto de indicadores fornecidos pelo *Webalizer*, foram usados apenas aqueles cujas informações foram úteis à avaliação do uso e do desempenho do *web site*. Os indicadores escolhidos, são apresentados a seguir, à medida que os resultados desse trabalho também são apresentados e discutidos. Esses indicadores possibilitam avaliar aspectos sobre origem dos acessos, ações executadas pelos usuários e desempenho do sistema em termos de respostas enviadas às solicitações dos usuários. Esses resultados podem possibilitar a

remodelagem do *web site*, de modo a melhorar a dinâmica de interação e promover o desenvolvimento de uma abordagem própria de análise de *logs*.

Destaca-se que os dados não podem ser vistos de forma bruta, mas sim sob a forma de valores estimados ou em ordens de grandeza, para não deixar escapar ao leitor as verdades e mentiras sobre as estatísticas dos *web trackings* (WINETT, 2001).

#### **4. RESULTADOS E DISCUSSÃO**

O uso do *web site* da Embrapa Monitoramento por Satélite foi analisado, no contexto desse trabalho, pela seguinte composição de eventos, cujo conjunto determinou os objetivos específicos do estudo:

- Dinâmica e origens do uso: eventos relacionados à dimensão tempo, expressa pelos acessos mensais; aos *strings* de busca e ao país de origem dos acessos;
- Ações de uso: representadas pelas páginas de entrada de maior acesso e agentes mais utilizados;
- Desempenho do *web site*: inferido pelos códigos de *status* (sucessos e erros de acessos).

##### **4.1. Dinâmica e origens do uso**

As avaliações da dinâmica dos uso do Web site e origem dos acessos relacionam-se à dimensão tempo (expressa pelos acessos mensais) e aos *strings* de busca e país de origem dos acessos.

###### **4.1.1. Análise temporal**

Analisando-se a dinâmica geral de uso do *web site* da Embrapa Monitoramento por Satélite, durante o período considerado nesse trabalho (Figura 1), observa-se que nos primeiros dois anos e meio o número de acessos manteve-se num patamar relativamente

constante, situado próximo a 100.000 páginas acessadas mensalmente. Mesmo assim, uma leve tendência de aumento anual pode ser notada. Isso indica, possivelmente, a resposta dos usuários à própria evolução do *web site* que, no decorrer do período considerado, foi agregando cada vez mais conteúdo de interesse geral ou específico aos internautas. Por outro lado, o aumento sempre crescente da “informatização” de todos os setores da sociedade, impulsionado pelas facilidades e oportunidades atuais de aquisição de equipamentos e meios de conexão, oferecidas comercialmente, além da própria popularização da Internet, podem igualmente ter contribuído para esse aumento progressivo.

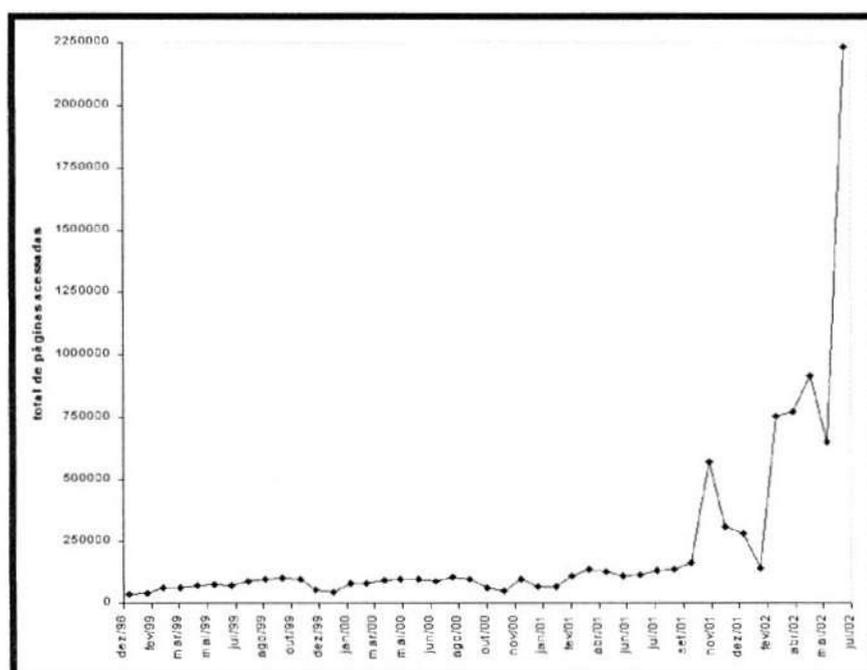


Figura 1: Dinâmica temporal dos acessos às páginas do *web site* da Embrapa Monitoramento por Satélite entre 1999 e 2002.

Verificou-se que os acessos evoluíram de uma média mensal de 71.732 páginas acessadas em 1999 para 586.296, em novembro de 2001, correspondendo a um aumento de acessos bastante distinto do padrão até então observado. Nessa época, estavam sendo

veiculadas, no próprio *web site* e na imprensa em geral, as primeiras informações sobre o Zoneamento Ecológico e Econômico do Estado do Maranhão e o lançamento da coleção de CD ROMs “Brasil Visto do Espaço”, dois trabalhos de grande repercussão nacional e de interesse geral da sociedade. A partir daí, a tendência observada foi de aumento muito expressivo de acessos, avalizando o acerto de estratégia institucional, adotada pela Embrapa Monitoramento por Satélite, de difusão dos resultados de suas pesquisas com a disponibilização de informações na Internet tendo-se detectado, prontamente, uma resposta positiva dos usuários.

O pico no número de páginas acessadas, em julho de 2002 é decorrente do lançamento da coleção de CD ROMs “Brasil Visto do Espaço” e da disponibilização do site desse tema, [www.cdbrasil.cnpm.embrapa.br](http://www.cdbrasil.cnpm.embrapa.br). O assunto despertou enorme interesse no público em geral e nele foi criado espaço para cadastramento dos interessados em adquirir a coleção, além de detalhes técnicos sobre o desenvolvimento, resultados e modo de utilização das informações. A coleção é um trabalho inédito no Brasil, onde imagens de satélite foram reunidas e tratadas digitalmente na forma de mosaicos, de forma que o produto representasse a superfície terrestre conforme vista do espaço.

#### 4.1.2. *Strings* de busca

A entrada dos internautas ao *web site* da Embrapa Monitoramento por Satélite pode ocorrer de maneira direta, se foi utilizada a URL [www.cnpm.embrapa.br](http://www.cnpm.embrapa.br), ou indireta, se utilizados *strings* ou palavras-chave, lançadas em um mecanismo de busca ou *links* a partir de outros sites para o da Embrapa Monitoramento por Satélite.

A Tabela 1 permite uma análise dos *strings* de busca registrados pelo *Webalizer*, como início de acesso ao *web site*. Nessa tabela, os valores referem-se ao percentual de participação dos *hits* para cada *string* no total de *hits*. A análise desses resultados mostra que os

mecanismos de busca devem estar respondendo bem às solicitações dos usuários ao indicar o *web site* da Embrapa Monitoramento por Satélite como referência para informações sobre queimadas, pois esse tema, no contexto dos impactos ambientais mundiais, tem despertado significativamente a atenção nacional e internacional e tem merecido um espaço e um tratamento de forma e conteúdo bastante destacado no conjunto do *web site* da unidade ([www.queimadas.cnpm.embrapa.br](http://www.queimadas.cnpm.embrapa.br)).

Tabela 1: *Strings* de busca mais utilizados como entrada para o *web site* da Embrapa Monitoramento por Satélite, entre 1999 e 2002

<i>String de busca</i>	1999	2000	2001	2002
Queimadas	12,17%	4,84%	8,51%	13,38%
Embrapa	2,20%	2,49%	2,61%	4,88%
Morcego	9,14%	2,79%	4,62%	4,54%
Satélite	5,09%	6,07%	6,19%	4,42%
Mapa rodoviário	5,78%	9,03%	6,08%	4,32%
GPS	0,63%	5,42%	1,52%	3,82%
Casas de madeira	1,08%	2,25%	2,44%	3,35%
Monografias	5,58%	4,79%	3,93%	3,14%
Impacto ambiental	8,36%	4,01%	3,09%	2,57%
Cana de açúcar	2,79%	2,90%	3,68%	2,11%
Gafanhoto	0,30%	0,60%	2,06%	1,37%
Pernilongo	0,12%	0,61%	0,23%	1,29%
Anfíbios	7,63%	3,71%	0,46%	0,99%
Rio Tejo	0,34%	0,75%	0,19%	0,96%
Cartografia	0,34%	0,97%	0,25%	0,87%
Analfabetismo	1,01%	4,21%	2,57%	0,86%
Tocantins	0,70%	0,53%	0,05%	0,69%
Reflorestamento	1,67%	0,73%	0,14%	0,45%
Jagatirica	0,58%	0,43%	0,12%	0,23%
<b>Representação Total</b>	<b>65,51%</b>	<b>57,13%</b>	<b>48,75%</b>	<b>54,26%</b>

Indicações de coerência dos mecanismos de busca também são revelados quando apontam o *web site*, respondendo aos *strings* “embrapa” e “satélite”. Entretanto, no que se

refere ao *string* “morcego”, explica-se a sua significativa ocorrência pela existência no *web site* de páginas destinadas à educação ambiental. Nessas páginas discute-se o impacto ambiental das atividades agrícolas sobre a fauna silvestre. São ainda apresentadas informações referentes a vários animais. Esse conteúdo específico, deve ser bastante acessado por um público, interessado em dados e ilustrações sobre animais silvestres. Isso explicaria também a existência de outros *strings* relacionados à fauna, na lista apresentada.

Deve-se ainda ressaltar a ocorrência, quantitativamente significativa, dos *strings* “mapa rodoviário” e “casas de madeira”. Em relação ao primeiro, alguma coerência poderia ainda ser apontada, por tratar-se de uma expressão que carrega um termo bastante relacionado ao conteúdo abordado no *web site*: “mapa”. Essa consideração, no entanto, não cabe ao segundo. Analisando-se as páginas componentes do *web site* (atualmente próximas de 30.000!), observa-se que essas duas expressões são títulos de página e os mecanismos de busca, via de regra, indexam as primeiras linhas do arquivo HTML das páginas disponibilizadas na Internet (CENDÓN, 2001), caso dessas duas expressões, em particular

A análise dos *strings* de busca torna-se importante na medida em que pode contribuir para que os idealizadores de *web sites* (*web designers*) usem de maneira estratégica o conteúdo das páginas, escolhendo palavras, expressões ou indicadores mais adequados para serem indexados. Conseqüentemente, pode-se ajustar o direcionamento do acesso para públicos específicos ou generalizados, conforme o interesse da instituição e a finalidade do *web site*. Esse tipo de ajuste é considerado também por MARCONDES (2001). Cuidados em monitorar as entradas dos usuários no *web site* podem evitar situações completamente inusitadas e inadequadas, relacionadas com termos ambíguos, imprecisos e até mesmo nomes próprios homônimos.

#### 4.1.3. País de origem dos acessos

Qualitativamente, o *web site* já foi acessado por internautas de 173 países diferentes, o que reitera a idéia de que as informações disponibilizadas na Internet tornam-se de interesse cosmopolita e amplamente dispersas.

Do ponto de vista quantitativo, a Figura 2 demonstra um padrão bastante nítido, indicando a origem dos acessos predominantemente nacional: mais de 60% do número total de *hits* acessados no período desse estudo. Em seguida, observa-se uma quantidade expressiva de acessos cuja origem não pôde ser determinada e que, como os acessos nacionais, no período do estudo, manteve-se sempre entre 20 e 30% do total de *hits*. Esse fato deve-se à má configuração de DNS (COMITÊ GESTOR, 2002) em que o endereço físico IP do computador cliente não foi traduzido para um nome uniforme, ou seja, não foi definido o endereço reverso (PRÁTICAS DE SEGURANÇA, 2002).

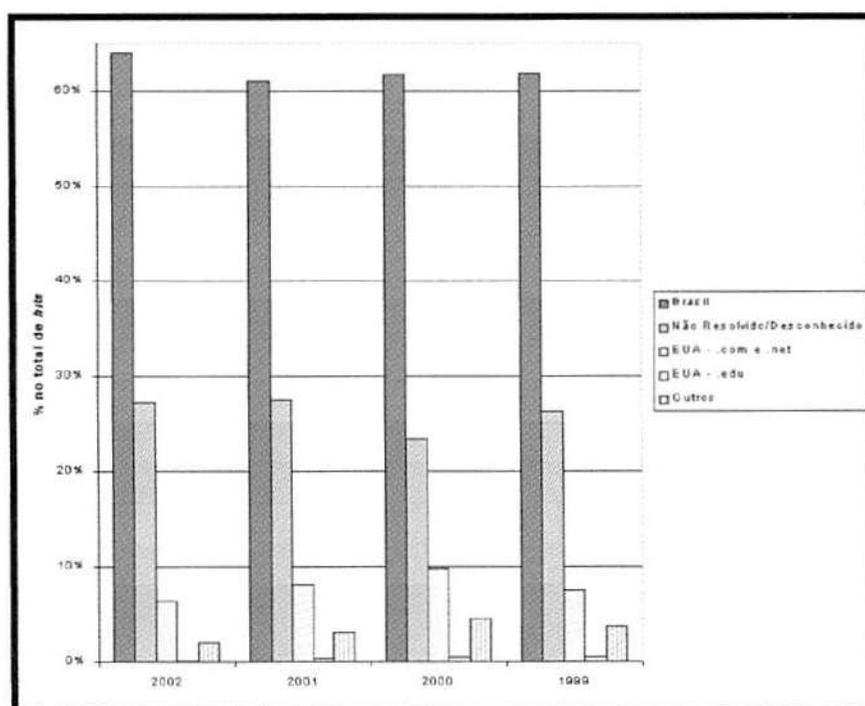


Figura 2: Percentual de participação dos principais países de origem dos acessos ao *web site* da Embrapa Monitoramento por Satélite, entre 1999 e 2002.

O país estrangeiro responsável pela quantidade de acessos mais significativa são os Estados Unidos, que ocupou o terceiro lugar na lista, embora nunca ultrapassando 10% do número total de *hits* acessados. Todo o conjunto restante, abrangendo todos os outros países detectados, nunca ultrapassou 5% desse indicador.

Esse dado demonstra uma forte inserção nacional do *web site*, indicando necessários cuidados e atenção, por parte dos seus idealizadores e realizadores, em termos de adequação da língua portuguesa e menos prioridade para veiculação de informação em outras línguas, pelo menos a curto e médio prazos. Essa última consideração, em particular, se valida uma vez que, mesmo já apresentando uma parte significativa de páginas já traduzidas para o inglês, o acesso internacional ao *web site* ainda é praticamente inexpressivo.

## **4.2. Ações de uso**

Indicações importantes para a gestão de conteúdo de um *web site* podem ser obtidas pelo monitoramento das ações que os usuários executam a partir do momento que o acessam. O itinerário percorrido e os arquivos recuperados remotamente podem fornecer indícios ou mesmo confirmar o grau de interesse despertado junto ao público por determinado conjunto de informações.

No entanto, as estatísticas organizadas pelo *Webalizer* permitem uma análise limitada dessas ações. Entre as situações que puderam ser consideradas no contexto desse trabalho apresentam-se, a seguir, os dados referentes às páginas de entrada ao *web site* e os agentes utilizados nos acessos.

### **4.2.1. Páginas de entrada**

A análise dos dados sobre as páginas mais acessadas demonstra que o conteúdo do *web site* é explorado pelos usuários de forma bem variada. A página inicial do *web site* da

Embrapa Monitoramento por Satélite (“cnpm” – URL: <http://www.cnpm.embrapa.br/>) e temas como o monitoramento orbital das queimadas no Brasil (“qmd” – URL: <http://www.queimadas.cnpm.embrapa.br/>) têm suas páginas constantemente acessadas, durante todos os meses do ano.

Outros temas, como o estudo dos gafanhotos em Mato Grosso (“gafa\_mt” – URL: [http://www.cnpm.embrapa.br/projetos/gafa\\_mt/index.html](http://www.cnpm.embrapa.br/projetos/gafa_mt/index.html)); os trabalhos nos municípios de Jaguariúna, SP (“jaguar” – URL: <http://www.cnpm.embrapa.br/projetos/jaguar/index.html>) e em Machadinho d’Oeste, RO (“machadinho” – URL: <http://www.cnpm.embrapa.br/projetos/machadinho/index.html>); as páginas com informações sobre satélites utilizados em pesquisa agroecológica (“satélite” – URL: <http://www.cnpm.embrapa.br/vp/saibamais/index.html>) e as páginas com a versão digital do informativo da unidade “Via Satélite” (“vs” – URL: <http://www.cnpm.embrapa.br/vp/viasatelite/index.html>) apresentam, de ano para ano, uma variação aleatória nos índices de acesso, provavelmente refletindo acessos de usuários interessados apenas momentaneamente nas informações.

Por último, observam-se temas que geram claramente um impacto no interesse público pois, tão logo disponibilizados, passam a ser sistematicamente acessados como, por exemplo, a coleção de CD ROMs “Brasil Visto do Espaço” (“cdbrasil” – URL: <http://www.cdbrasil.cnpm.embrapa.br/>), cuja importância no contexto do *web site* já foi anteriormente mencionada.

#### 4.2.2. Agentes

Os agentes foram agrupados em 4 categorias, a saber: indefinido, *browser* (Netscape, Internet Explorer, etc.), robôs (*crawlers*, *spiders*, *link checkers*, *proxys*) e *mirroring* (capturam o site para navegação *off line*).

A análise da participação percentual de cada grupo de agentes utilizados para acesso, no total de *hits* mostra que a quase totalidade de usuários do *web site* utilizam agentes do tipo browser, o que implica em assumir um perfil majoritário de usuários que estão em busca de informações específicas sobre os temas de atuação da Embrapa Monitoramento por Satélite, os quais são transferidos para variadas sessões no *web site*. Esses resultados parecem indicar que o usuário mais comum é um membro da comunidade acadêmica (estudante, professor ou pesquisador), funcionários de organizações governamentais ou não governamentais, empresários, representantes de organizações civis (cooperativas agrícolas, comissões e grupos de trabalhos relacionados à temática de agricultura, agronegócio e meio ambiente). Esses usuários, muito provavelmente, utilizam os browsers mais conhecidos e disseminados no mercado pelas facilidades de aquisição, instalação e familiaridade de uso.

No caso dos agentes tipo robôs, observa-se um aumento de sua participação no total de *hits* acessados durante o ano de 2000, o que pode ser explicado pelo aumento de informações incorporadas no *web site* durante esse ano em particular.

O mais interessante, no entanto, é o aumento de participação do tipo de agente *mirroring*, expressivamente crescente, a partir do ano 2001. Esse fato é função da disponibilização dos dados do projeto “Brasil Visto do Espaço”, que despertou no público o interesse de obtenção das informações, estimulando-os a copiarem e transferirem a totalidade dos arquivos para suas próprias máquinas. Isso permitiu ao usuário acesso mais barato e rápido aos dados desse projeto, já que não houve necessidade de estar conectado à Internet durante o processo.

#### **4.3. Desempenho do *web site***

O ponto que mais chama a atenção, em relação ao desempenho do *web site* da Embrapa Monitoramento por Satélite, é a constância do percentual de códigos de sucesso

durante o período considerado, que mantém-se praticamente sempre próxima de 100%; ao mesmo tempo, os códigos de erro tipo 500 não apresentam registros. O grande investimento e a constante preocupação da unidade na atualização e manutenção da infra-estrutura (*hardware* e *software*) e recursos operacionais (humanos, mecanismos de segurança incluídos) têm garantido a disponibilidade das informações ao acesso externo 24 horas por dia, sete dias por semana (PIEROZZI Jr; CAPUTI; FILARDI, 2000).

O aumento observado no ano de 2000, para o percentual de códigos de erro tipo 400, pode ser explicado pela transformação ocorrida naquele ano, do nome da instituição e, conseqüentemente, das URLs das páginas do *web site*. Em 2000, a Embrapa Monitoramento por Satélite transformou-se de “Núcleo de Monitoramento Ambiental e de Recursos Naturais por Satélite”, sigla “NMA” – URL: [www.nma.embrapa.br](http://www.nma.embrapa.br), para “Centro Nacional de Pesquisa de Monitoramento por Satélite”, sigla “CNPM” – URL: [www.cnpm.embrapa.br](http://www.cnpm.embrapa.br). Essa mudança pode ter ocasionado um sensível aumento dos erros tipo 400 – erros de sintaxe. – quando os usuários, habituados a acessarem o *web site* com sintaxes contendo “NMA”, passaram a não ter respostas de sucesso no acesso, mesmo com o mecanismo de automático de correção, implantado e operacional durante o ano todo que redirecionava, para as novas URLs, os acessos originados com a utilização da URL antiga.

## 5. CONCLUSÕES

A análise dos dados reunidos e organizados pelo *Webalizer*, permitiu evidenciar algumas tendências e padrões no uso das informações disponibilizadas, via Internet, pela Embrapa Monitoramento por Satélite. Indicações sobre a funcionalidade do sistema e seu desempenho também puderam ser inferidas.

Os resultados dessa análise podem gerar diretrizes gerenciais no que se refere à melhoria contínua do *web site* (forma e conteúdo) e ao estabelecimento de um padrão de divulgação de informações.

O *Webalizer* mostrou-se útil em relação ao monitoramento bruto da dinâmica de uso, uma vez que identificou padrões temporais de utilização do *web site*, permitindo a caracterização dos períodos com incremento do número de acessos. Permitiu ainda a comprovação de ubiquidade das informações e a identificação de coerência entre o que está sendo procurado pelo internauta e o que está disponível. As discrepâncias observadas no processo de monitoramento (*strings*) também contribuem para o planejamento e desenvolvimento do *web site*.

Embora fosse importante obter-se o número exato de pessoas que acessaram o site nesse período, isso só seria possível com uma prática restritiva de autenticação de acesso por senha que é contrária à filosofia democrática da Internet, incorporada pela Embrapa Monitoramento por Satélite. Resta a alternativa de se utilizar instrumentos que fazem uma abordagem aproximativa do perfil do usuário, sem desconsiderar os cuidados de interpretação dos dados recuperados.

Em relação às ações de uso, percebe-se crescente presença de agentes *mirroring*, relacionadas com o tipo de informação que interessa ao internauta retirar do *web site* para armazenamento e consulta *off line*. Consonante, portanto, com o resultado da análise dos acessos das páginas de entrada.

O desempenho do *web site* mostrou-se constante ao longo do período, oferecendo confiabilidade ao usuário no que tange à concepção da arquitetura da informação.

Essa análise abre uma perspectiva bastante rica na exploração e caracterização do perfil do usuário, exigindo o desenvolvimento de procedimentos customizados de análise arquivos de log conforme interesses predeterminados, consolidando-se como uma das etapas

do processo de concepção, disponibilização e utilização da informação. Dessa maneira algumas idéias se apresentam como cenários de exploração:

- separar seções do web site com a finalidade de determinar a distribuição de acessos;
- interceptar entradas mais acessadas e comparar com o *log*;
- promover alteração da estrutura com base nas regras de indexação nos mecanismos de busca e *robots* ;
- identificar através da ocorrência dos agentes de *mirroring* oportunidades para download de seções completas;
- desenvolver abordagem própria de análise de arquivos de *log*;
- adequar a apresentação das estatísticas de acesso aos diferentes públicos.

## REFERÊNCIAS

- BACCALA, B. Connected: An Internet Encyclopedia. 1997. Disponível em: <<http://www.freesoft.org/CIE/index.htm>>. Acesso em: 25 set. 2002.
- BARRETT, B.L. The Webalizer - A web server log file analysis tool. 1997. Disponível em: <<ftp://ftp.mrunix.net/pub/webalizer/README>>. Acesso em: 15 jul. 2002.
- BERTOT, J. C.; MCCLURE, C. R.; MOEN, W. E.; RUBBIN, J. Web usage statistics: measurement issues and analytical techniques. *Government Information Quarterly*, v. 14, n. 4, p. 375-395, 1997.
- CATLEDGE, L. D.; PITKOW, J. E. Characterizing browsing strategies in the World Wide Web. *Computers Networks and ISDN Systems*, v. 27, p. 1065-1073, 2001.
- CENDÓN, B. V. Ferramentas de busca na Web. *Ciência da Informação*, Brasília, DF, v. 30, n. 1, p. 39-49, jan./abr. 2001.
- COMITÊ GESTOR DA INTERNET NO BRASIL. Disponível em: <<http://www.cg.org.br/faq/basico-01.htm>>. Acesso em: 25 set. 2002.

- DIAS, G.A. Avaliação do acesso a periódicos eletrônicos na web pela análise do arquivo de log de acesso. *Ciência da Informação*, Brasília, DF, v. 31, n. 1, p. 7-12, jan.-abr. 2002.
- EICHMANN, D.; WU, J. Sulla - A User Agent for the Web. 1996. Disponível em: <http://mingo.info-science.uiowa.edu/eichmann/www-s96/Overview.html>. Acesso em: 25 set. 2002.
- GNU'S NOT UNIX!. Disponível em: <http://www.gnu.org>. Acesso em: 24 set. 2002.
- HJELM, J. Log analysis tools. 1999. Disponível em: <http://www.w3.org/WCA/loganalysis-tools.html>. Acesso em: 24 set. 2002.
- HOCHHEISER, H.; SHNEIDERMAN, B. Using interactive visualizations patterns of WWW log data to characterize access patterns and inform site design. *Journal of the American Society for Information Science and Technology*, v. 52, n. 4, p. 331-343, 2001.
- LEE, S.; HELLER, R. S. Use of keystroke log file to evaluate na interactive computer system in a museum setting. *Computers Education*, v. 29, n. 2/3, p. 89-101, 1997.
- LOG FILES - Apache Http Server. 2002. Disponível em: <http://httpd.apache.org/docs/logs.html>. Acesso em: 1 out. 2002.
- MARCONDES, C. H. Representação e economia da informação. *Ciência da informação*, Brasília, DF, v. 30, n. 1, p. 61-70, jan./abr. 2001.
- MURPHY, J.; HOFACKER, C. H.; BENNETT, M. Website-generated market-research data: tracing the tracks behind visitors. *Cornell Hotel and Restaurant Administration Quarterly*, p. 82-91, feb. 2001.
- NAKANO, R. *Web Content Management: a collaborative approach*. Boston: Addison-Wesley, 2002. 238 p.
- PETERS, T. A. Remotely familiar: using computerized monitoring to study remote use. *Library Trends*, v. 47, p. 7-20, 1998.

PETERS, T. A. The history and development of transaction log analysis. *Library Hi Tech*, v. 11, n. 2, p. 41-50, 1993.

PIEROZZI Jr., I.; CAPUTI, E.; FILARDI, A. L. *A Internet como veículo de comunicação, difusão de resultados e imagem institucional na Embrapa Monitoramento por Satélite: 1. Infra-estrutura e funcionamento*. Campinas: Embrapa Monitoramento por Satélite, 2000. 25p.

PRÁTICAS DE SEGURANÇA PARA ADMINISTRADORES DE REDES INTERNET. Disponível em: <<http://www.nbso.nic.br/docs/seg-adm-redes.html>>. Acesso em: 25 set. 2002.

THE WEB ROBOTS PAGES. Disponível em: <<http://www.robotstxt.org/wc/robots.html>>. Acesso em: 24 set. 2002.

THE WEBALIZER: WHAT IS YOUR WE SERVER DOING TODAY? Disponível em: <<http://www.mrunix.net/webalizer/>>. Acesso em: 24 set. 2002.

WINETT, B. Tracking your visitors. 2001. Disponível em: <<http://hotwired.lycos.com/webmonkey/98/16/index2a.html>>. Acesso em: 25 set. 2002.

YU, L.; APPS, A. Studying e-journal user behavior using log files: the experience of SuperJournal. *Library & Information Science Research*, v. 22, n. 3, p. 311-338, 2000.