

Análise de dinâmica de uso e de desempenho: o caso do *web site* da Embrapa Monitoramento por Satélite*

Ivo Pierozzi Júnior

Pesquisador em Arquitetura da Informação e chefe adjunto de P&D.
E-mail: ivo@cnpn.embrapa.br

Eliane Gonçalves Gomes

Pesquisadora em Métodos Quantitativos em P&D.
E-mail: eliane@cnpn.embrapa.br

Maria de Cléofas Faggion Alencar

Bibliotecária e pesquisadora em Ciência da Informação.
E-mail: cleo@cnpn.embrapa.br

Carlos Alberto de Carvalho

Analista de Sistemas e *Web designer*
E-mail: calberto@cnpn.embrapa.br

Resumo

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*. Com os 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*.

Palavras-chave

Dinâmica de uso de *websites*; Arquivos de *log*; Perfil de usuário.

Usage dynamics and performance analysis: the case study of Embrapa Satellite Monitoring web site

Abstract

The *Embrapa Satellite Monitoring* has used for more than one decade 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, a *Webalizer* software has been used 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 the origin of requests (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.

Keywords

Website usage dynamics; Log files; User's profile.

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. Seus limites estão mais relacionados ao estágio atual de desenvolvimento das tecnologias eletrônicas ou à infra-estrutura de telecomunicações, ainda precária em algumas partes do mundo.

Surgida inicialmente no meio militar por meio da Arpanet e, posteriormente, difundida para o meio acadêmico americano e para as instituições de ensino mundiais (Leiner *et alii*, 2000), atualmente a Internet já foi incorporada por diversos setores e tipos de atividades humanas, tais como a educação, comércio, mídia, lazer, serviços, relações interpessoais e outros. Não seria exagero considerar que se pode encontrar quase tudo na Internet: desde hora certa até psicoterapia *on-line*!

Nesse ponto, destaca-se a importância da representação na economia da informação, na qual a identificação e a avaliação são atividades importantes da utilização de recursos na Internet. "O trabalho informacional de desenvolvimento e sistematização de representações das informações disponíveis na Internet joga um papel decisivo, viabilizando processos de identificação, avaliação e uso, atuando como intermediária entre usuários e suas necessidades de informação e recursos potencialmente importantes." (Marcondes, 2001)

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, em um só veículo, imagens, sons e textos.

* Embrapa Monitoramento por Satélite – Av. Dr. Júlio Soares de Arruda, 803, Parque São Quirino, 13088-300, Campinas, SP, Brasil – Fone: + 55 19 3256-6030, Fax: + 55 19 3254-1100

A Embrapa Monitoramento por Satélite presta assessoria a órgãos públicos e à iniciativa privada em problemas e desafios agropecuários e ambientais, por meio 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 por seus projetos de pesquisa (Pierozzi Jr. *et alii*, 2000). Mas, além disso, tem investido na evolução tecnológica que envolve o setor, acompanhando sua rapidez e 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.

Diante da importância da opinião, expectativas e satisfação desse público em relação ao conteúdo e formato das informações disponibilizadas, algumas iniciativas de avaliação do uso do *web site* foram implementadas. Esses procedimentos visam a aprimoramentos na arquitetura da informação e na gestão do conteúdo que, segundo Nakano (2002), engloba 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).

Nessa verdadeira babel de informações que é hoje a Internet, o estabelecimento e a manutenção de padrões de qualidade constituem o diferencial para que se consolidem os canais de comunicação entre as instituições e seus usuários: ninguém melhor para avaliar um produto do que a pessoa que dele faz uso.

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 a 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.

ANÁLISE DE ARQUIVOS DE LOG

Arquivos de *logs* são arquivos-texto gerados por um *software* que registra os eventos e a hora em que ocorrem em 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 por meio de outras fontes. Se processados apropriadamente, podem ser calculadas estatísticas de uso, que fornecem dados úteis para estudos de usuários e gerenciamento da informação.

Peters (1993, 1998) destaca que essa técnica tem sido usada em estudos de recuperação da informação *on-line* por duas décadas, para uso e gerenciamento de informações eletrônicas em bibliotecas, e ganhou maior popularidade nos anos 90, com a grande difusão da Internet. Na literatura encontram-se diversas experiências referentes ao uso e análise dos arquivos de *log*.

Catledge & Pitkow (1995) afirmam que a análise dos arquivos de *log* pelo lado do cliente incrementa o entendimento das estratégias de navegação do usuário, bem como oferece interface real dos dados de uso. Esses padrões de navegação podem ser *browsing* serendipista ou geral e busca baseada em análise empírica dos arquivos de *log* dos eventos do usuário. As caracterizações foram derivadas da pesquisa de hipertexto e são promissoras para a área de interesse da *web*.

A análise de arquivos de *log* realizada por Lee & Heller (1997) refere-se ao uso e aos padrões de uso de visitantes de um sistema multimídia interativo, instalado no US Holocaust Memorial Museum. As análises dos resultados referem-se à duração da sessão, ao início das escolhas no *menu*, ao meio visto (um evento é uma série de filme ou fotos) e ao acesso ao arquivo de ajuda. Assim como pode ocorrer na avaliação de *logs* para a *Web*, esse estudo sugere refinamentos do compromisso com o sistema técnico e com o conteúdo do sistema. Além disso, fornece guias para adequar o tempo de uso do sistema em um espaço público.

Bertot *et alii* (1997) apresentam diversas técnicas de análise de arquivos de *log* e apontam questões associadas à interpretação e gestão das informações contidas nesses arquivos. Os autores destacam que, devido ao caráter novo da análise dos arquivos de *log*, há falta de consistência nos formatos dos arquivos e necessidade de desenvolver meios adicionais de analisá-los. O objetivo seria minimizar alguns problemas, como avaliação da transferência de arquivos, contagem dupla de *hits* e errônea das páginas de acesso, separação dos agentes internos e externos, entre outros.

A experiência do uso dos arquivos de *log* para entender o comportamento do usuário no projeto *SuperJournal* (consórcio de 19 editores de jornais eletrônicos) é relatada por Yu & Apps (2000). O objetivo dos autores residuiu em dar diretrizes ao processamento de arquivos de *log* e possíveis análises focadas no usuário.

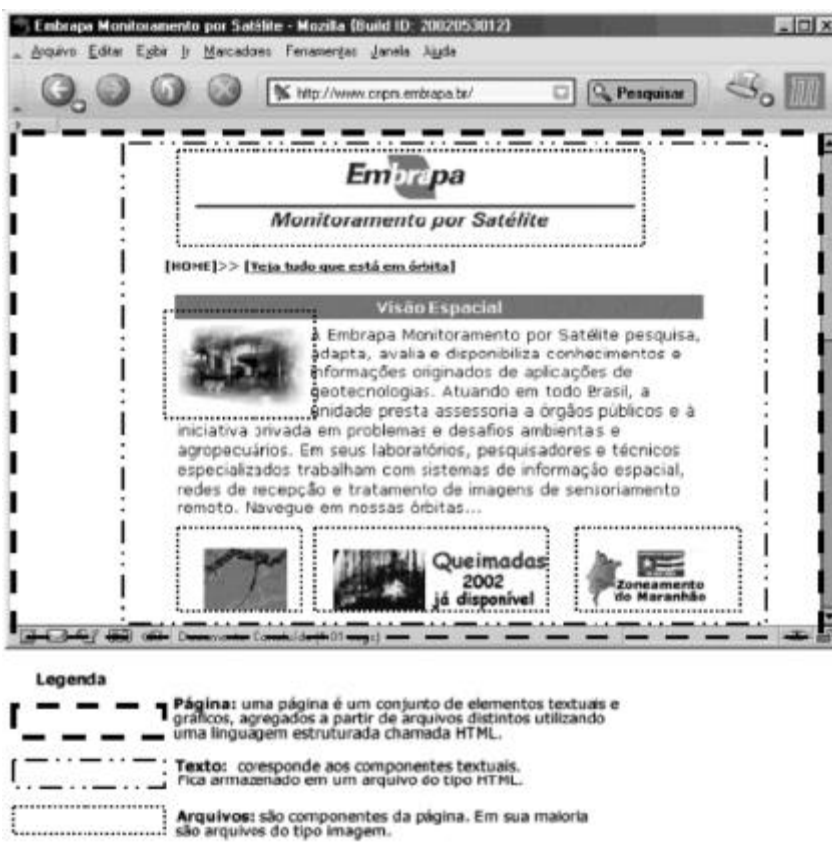
Murphy *et alii* (2001) discutem o caso de análise de *logs* de um *web site* de um restaurante. Os autores mostram como essas análises podem ser usadas para recuperar informações sobre os visitantes. São discutidas as potencialidades e limitações das análises para objetivos de comercialização de produtos ou serviços.

Os resumos apresentados pelos *softwares* de análise de arquivos de *logs* podem, muitas vezes, ocultar informações importantes e restringem os usuários e operadores a uma posição meramente de interpretação das estatísticas disponibilizadas. Hochheiser & Schneiderman (2001) propõem visualizações interativas, que combinam as variáveis analisadas em um espaço bidimensional. Os autores igualmente sugerem que um maior esforço de pesquisa seja dedicado à questão do processamento e análise dos arquivos de *log*.

Dias (2002) avalia o acesso a periódicos eletrônicos disponibilizados na *World Wide Web* por meio de *log* de acesso a partir da ferramenta *Webtrends Log Analyser*. São apresentados os problemas relacionados à utilização de dados de arquivos de *log* de acesso, como, por exemplo, a somatória de *hits* que determinado *site* recebeu ou, então,

FIGURA 1

Conceituação de alguns componentes constitutivos do *web site* da Embrapa Monitoramento por Satélite



o processo de contagem e identificação de sessões de usuários. O autor destaca que as análises com esses instrumentos não podem ser consideradas totalmente inúteis, pois podem auxiliar no melhoramento de forma e de conteúdo apresentados ao internauta.

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 e outros.

As definições desses indicadores foram propostas por Barrett (1997) e são representadas, em conjunto, na figura 1, com a finalidade de garantir o entendimento das análises dos resultados aqui apresentados.

Uma página é formada por vários componentes, cada qual sendo um arquivo. Esses arquivos são agregados por meio de comandos da linguagem HTML. No exemplo da figura 1, a página é composta pelos seguintes arquivos:

- INDEX.HTML: um arquivo que armazena os elementos textuais e os comandos HTML que irão agregar os demais arquivos. A aparência de um arquivo HTML é apresentada na figura 2;

- LOGO_EMBRAPA.GIF, FRENTE_EMBRAPA.JPG, MAPA1.JPG, MAPA2.JPG, MAPA3.JPG: imagens que ilustram a página.

A página da figura 1 é formada por seis arquivos: um arquivo do tipo HTML e cinco arquivos do tipo imagem. Ao ser requisitada por um *browser* ou agente, ela irá gerar seis *hits*.

Outras definições importantes (Eichmann & Wu, 1996; Baccala, 1997; The Web, 2002) são necessárias:

- *web site*: um conjunto de páginas interligadas e tratando do 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*) – códigos de três 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;

FIGURA 2

Exemplificação da estrutura inicial de um arquivo HTML do *web site* da Embrapa Monitoramento por Satélite

```
<html>
<head>
  <title>Embrapa Monitoramento por Satélite</title>
</head>
<body>
  <img src=logo_embrapa.gif>
  <font>[home] >> [Veja tudo que está em órbita]</font>
  <img src=frente_embrapa.jpg>
  <font>A Embrapa Monitoramento por Satélite pesquisa...</font>
  <img src=mapa1.jpg> <img src=mapa2.jpg> <img src=mapa3.jpg>
</body>
</html>
```

- tipo 3xx: *Redirection* – mais ações, além do pedido, são necessárias para que o arquivo seja enviado ao solicitante.

Códigos de erro:

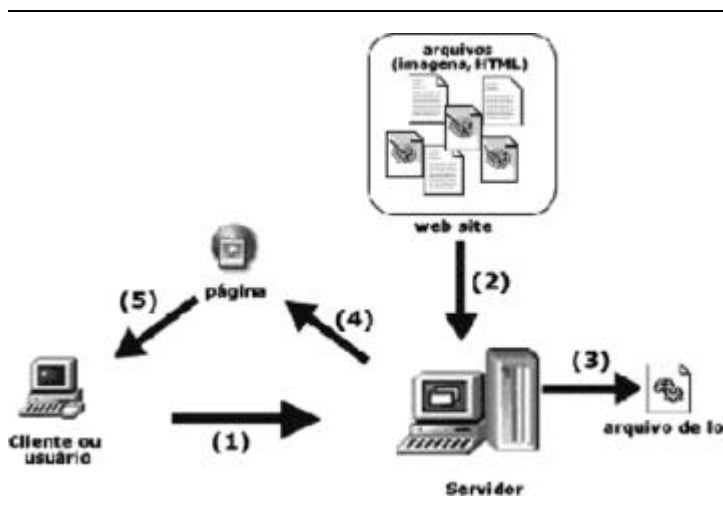
- tipo 4xx: *Client Error* – o pedido não foi feito corretamente devido a algum erro de sintaxe, por exemplo, nome de arquivo errado ou inexistente;
- tipo 5xx: *Server Error* – o pedido não foi aceito devido à falha do servidor que abriga o *web site*.
- Agentes: tecnicamente são conhecidos como *user-agents*. 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*. Existe, ainda, uma outra classe de agentes, principalmente utilizada por mecanismos de busca como *Yahoo* e *Google* para alimentar suas bases de dados, chamados de robôs (*web robots*);
- 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. Os *web browsers* não são robôs porque eles dependem da intervenção humana para buscar informações. Os *web robots* possuem nomes como *web wanderers* (errante), *web crawlers* (rastejantes), *spiders* (aranhas), *worms* (vermes) ou *ants* (formigas) por esses programas darem a impressão de estarem se movendo sozinhos através do hipertexto, como uma aranha caminhando na teia. De fato, eles não fazem isso, apenas vão requisitando os arquivos

conforme encontram-nos na estrutura de hipertexto. Uma divisão interessante de robôs é a do tipo *mirroring* ou *off line browsing*, que realizam a tarefa de transferir todos os arquivos encontrados na estrutura do hipertexto do *web site* para um computador cliente. Esse tipo de robô faz uma cópia de um *web site* para que o usuário navegue no mesmo sem necessidade de estar conectado à Internet. Todos os tipos de robôs, mais intensamente os *mirroring*, causam um comportamento de “inflacionar” o número de *hits* (por consequência, páginas), pois um robô apenas consegue requisitar o *web site* inteiro sem que o fator humano tenha interferência;

- 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. De forma simplificada, o processo de solicitação de uma página em um *web site* é o representado na figura 3. O cliente ou usuário é um computador conectado à Internet. Ele efetua o pedido de uma página (1) ao servidor, que irá verificar no *web site* se aquele pedido pode ser atendido. Esse pedido pode ser um arquivo em particular ou uma página formada por vários arquivos. Uma vez que o servidor identificou os arquivos necessários (2), registram-se no arquivo de *log* algumas informações como data e hora em que o arquivo foi solicitado, qual cliente solicitou (seu endereço IP), que tipo de *web browser* ou agente foi utilizado, qual o tamanho do arquivo solicitado, se o arquivo foi encontrado ou se gerou um erro. Cada solicitação de arquivo para formar a página (os chamados *hits*) é gravada em uma linha do arquivo de *log* (3). O servidor envia para o cliente os arquivos que compõem a página (4) e (5).

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 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*.

FIGURA 3
Esquema de acesso a uma página em um *web site*



Entre o conjunto de indicadores fornecidos pelo *Webalizer*, usaram-se apenas aqueles cujas informações foram úteis à avaliação do uso e do desempenho do *web site*. Os indicadores escolhidos são mostrados 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, para contextualizar a discussão apresentada nesse trabalho, é necessário estabelecer uma definição dos melhores critérios de caracterização da dinâmica de visitação de um *web site*. Adicionalmente, 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).

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, diários e horários, assim como 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).

Dinâmica e origens do uso

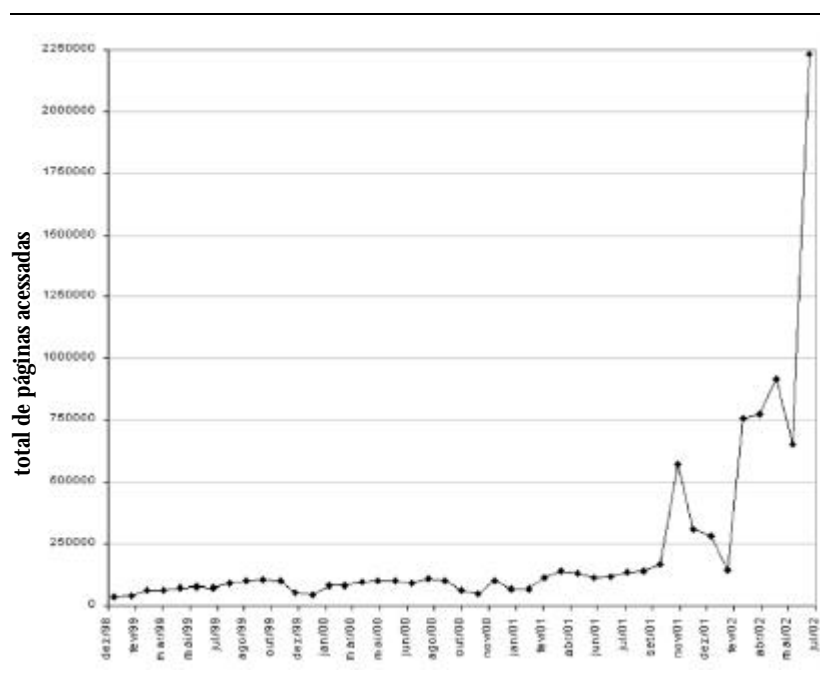
Análise temporal

Analisando-se a dinâmica geral de uso do *web site* da Embrapa Monitoramento por Satélite, durante o período considerado neste trabalho (figura 4), observa-se que nos primeiros dois anos e meio o número de acessos manteve-se em patamar relativamente constante, situado próximo a 100 mil 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. Além disso, 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.

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

FIGURA 4

Dinâmica temporal dos acessos às páginas do *web site* da Embrapa Monitoramento por Satélite, entre 1999 e 2002



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. 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, no qual imagens de satélite foram reunidas e tratadas digitalmente em forma de mosaicos, de forma que o produto representasse a superfície terrestre conforme vista do espaço.

O *web site* do Zoneamento Ecológico e Econômico do Estado do Maranhão foi concebido para abrigar espaços de interação com os internautas. Essa condição pode ter contribuído para o aumento dos acessos mencionados anteriormente.

Outro aspecto observado no padrão temporal de uso é que, claramente, nos meses de dezembro a fevereiro e no mês de julho, os acessos são reduzidos praticamente pela metade. Esses meses correspondem ao período de

férias escolares no Brasil. À primeira vista, isso pareceria indicar um uso predominantemente exercido pelo setor educacional. No entanto, vale a pena levar em consideração que, acompanhando as férias escolares, outros setores da sociedade também diminuem suas atividades nesses meses. Essa concentração de uso nos meses de maior atividade escolar e do setor produtivo em geral parece indicar que o usuário do *web site* não procura lazer e concentra-se, mais especificamente, na busca de informações relacionadas com o perfil de atuação da Embrapa Monitoramento por Satélite e assuntos afins.

A análise da dinâmica diária de usos apóia essa idéia. Verifica-se claramente que os acessos concentram-se nos dias e horários de expediente de trabalho, com nítidas reduções nos horários de almoço e noturno (figura 5) e finais de semana (figura 6).

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, 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, a seguir, 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

FIGURA 5
Número médio de páginas do *web site* da Embrapa Monitoramento por Satélite, acessadas entre 1999 e 2002, em três intervalos diários de horários

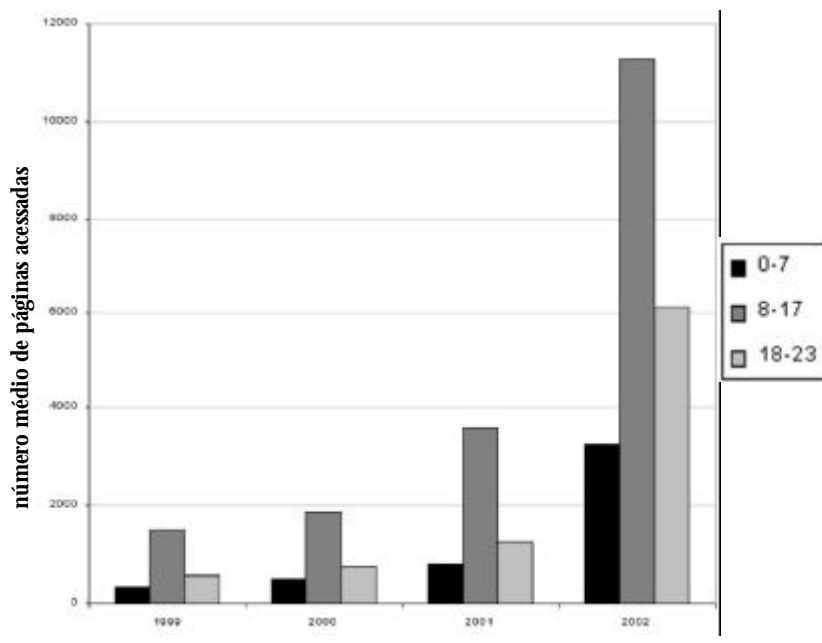
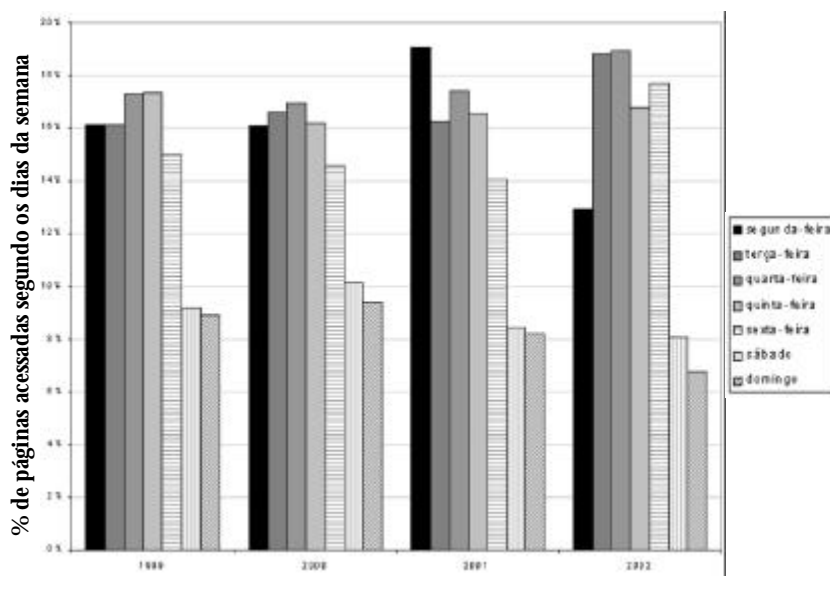


FIGURA 6
Percentual de páginas do *web site* da Embrapa Monitoramento por Satélite, acessadas entre 1999 e 2002, segundo os dias da semana



a atenção nacional e internacional e tem merecido espaço e tratamento (de forma e conteúdo) bastante destacados no conjunto do *web site* da unidade (www.queimadas.cnpm.embrapa.br).

Indicações de coerência dos mecanismos de busca também são reveladas 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” que os coloca entre os dez *strings* mais frequentes. 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 mil!), observa-se que essas duas expressões são títulos de página. Os mecanismos de busca, em geral, indexam as primeiras linhas do arquivo HTML das páginas disponibilizadas na Internet (Cendón, 2001), caso dessas duas expressões, em particular, que figuram como *tags* “<title></title>” (figura 2), nos seus respectivos arquivos (Search Engine Watch, 2002).

TABELA 1

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

String de busca	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%
Jaguatirica	0,58%	0,43%	0,12%	0,23%
Representação Total	65,51%	57,13%	48,75%	54,28%

Esse tipo de análise torna-se importante à medida 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.

País de origem dos acessos

A figura 7 e a tabela 2 mostram alguns países de origem dos acessos ao *web site*. 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 disperso.

Do ponto de vista quantitativo, a figura 7 demonstra padrão bastante nítido, indicando a origem predominantemente nacional dos acessos: mais de 60% do número total de *hits* acessados no período desse estudo. Em seguida, observa-se 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). O endereço IP (por exemplo, 200.136.111.23) está disponível e é necessário saber a que domínio pertence (país, comercial, educacional e outros). Caso o administrador de cada rede não denomine cada computador a ela conectado, gera-se o erro do tipo “não resolvido”.

O país 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. O conjunto restante, abrangendo todos os outros países detectados, nunca ultrapassou 5% desse indicador.

Esse dado demonstra forte inserção nacional do *web site*, indicando necessários cuidados e atenção por parte dos seus idealizadores e realizadores, em termos de adequação

FIGURA 7

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

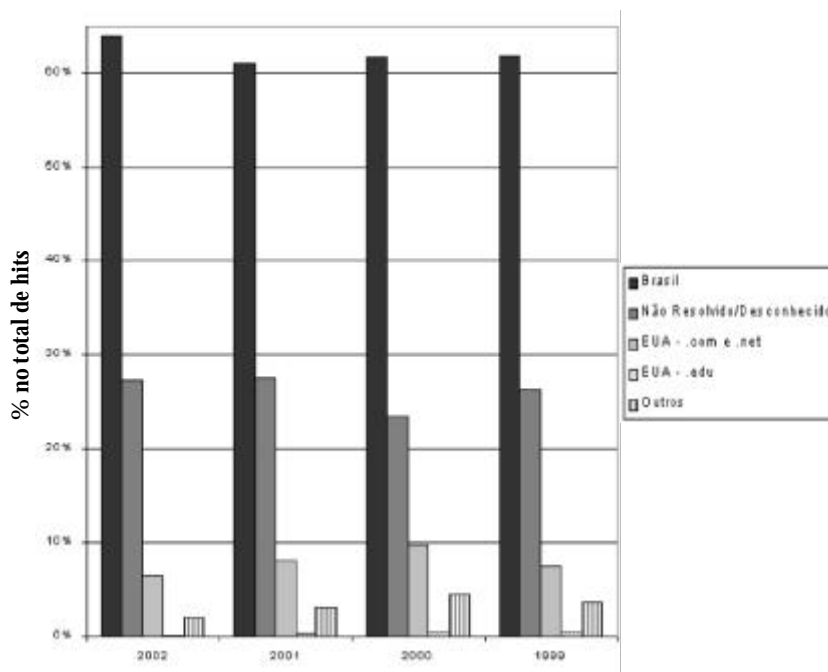


TABELA 2

Percentual de participação de outros países de origem dos acessos ao *web site* da Embrapa Monitoramento por Satélite, entre 1999 e 2002

Origem dos acessos	Percentual no total de hits			
	1999	2000	2001	2002
Alemanha	0,33%	0,25%	0,30%	0,10%
Argentina	0,23%	0,23%	0,13%	0,06%
Canadá	0,10%	0,13%	0,27%	0,30%
Espanha	0,17%	0,21%	0,26%	0,17%
França	0,21%	0,17%	0,15%	0,11%
Holanda	0,10%	0,08%	0,08%	0,06%
Itália	0,12%	0,14%	0,07%	0,07%
Japão	0,13%	0,12%	0,14%	0,10%
México	0,21%	0,26%	0,13%	0,07%
Portugal	0,78%	1,42%	0,68%	0,45%
Reino Unido	0,10%	0,25%	0,09%	0,09%
Suíça	0,09%	0,05%	0,03%	0,03%
Outros	0,98%	0,96%	0,52%	0,25%

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, valida-se, uma vez que, mesmo apresentando uma parte significativa de páginas já traduzidas para o inglês, o acesso internacional ao *web site* ainda é praticamente inexpressivo (tabela 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.

Desse modo, as informações podem ser organizadas e disponibilizadas com objetivos direcionados à captura da atenção do usuário pela sua importância imediata ou momentânea, ou, ao contrário, pelo seu caráter de pertinência permanente (base de dados, por exemplo).

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

Páginas de entrada

A tabela 3 apresenta as 10 páginas de entrada mais acessadas no período de tempo considerado neste trabalho.

A análise desses dados 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.

TABELA 3

Páginas de entrada, em ordem alfabética, e respectivo percentual anual de acessos ao *web site* da Embrapa Monitoramento por Satélite, entre 1999 e 2002

Página de entrada (ordem alfabética)	Percentual de registros no total de meses			
	1999	2000	2001	2002
Cdbrasil	---	---	33,33%	100,00%
Cnpm	100,00%	100,00%	100,00%	100,00%
Demene	75,00%	100,00%	100,00%	100,00%
Fauna	0,00%	91,67%	100,00%	100,00%
gafa_mt	0,00%	75,00%	25,00%	0,00%
Jaguar	25,00%	100,00%	75,00%	100,00%
Machadinho	25,00%	50,00%	91,67%	85,71%
Qmd	100,00%	100,00%	100,00%	100,00%
Satelite	50,00%	100,00%	66,67%	100,00%
Vs	0,00%	8,33%	100,00%	85,71%

Outros temas, como o estudo dos gafanhotos em Mato Grosso (“gafa_mt” – URL: 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 (“satelite” – 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 impacto no interesse público, pois, tão logo disponibilizados, passam a ser sistematicamente acessados, como, por exemplo, o trabalho realizado no Rio Demene, AM/RR (“demene”: URL: <http://www.cnpm.embrapa.br/projetos/demene/index.html>); os estudos sobre fauna em Campinas, SP (“fauna” – URL: <http://www.cnpm.embrapa.br/projetos/fauna/index.html>) e 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.

Agentes

Os agentes foram agrupados em quatro 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 tabela 4 apresenta a participação percentual de cada grupo de agentes utilizados para acesso, no total de *hits*.

A análise da tabela 4 mostra que a quase totalidade de usuários do *web site* utiliza agentes do tipo *browser*, o que implica 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 os usuários mais comuns: membros 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 do tipo robôs, observa-se incremento de sua participação no total de *hits* acessados durante o ano de 2000, o que pode ser explicado pelo aumento de informações incorporadas ao *web site* durante esse ano em particular, como, por exemplo, as informações da campanha nacional sobre alternativas para prática de queimadas na agricultura (http://www.cnpm.embrapa.br/projetos/qmd/qmd_2000/index.htm) e da coleção “Brasil Visto do Espaço”.

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

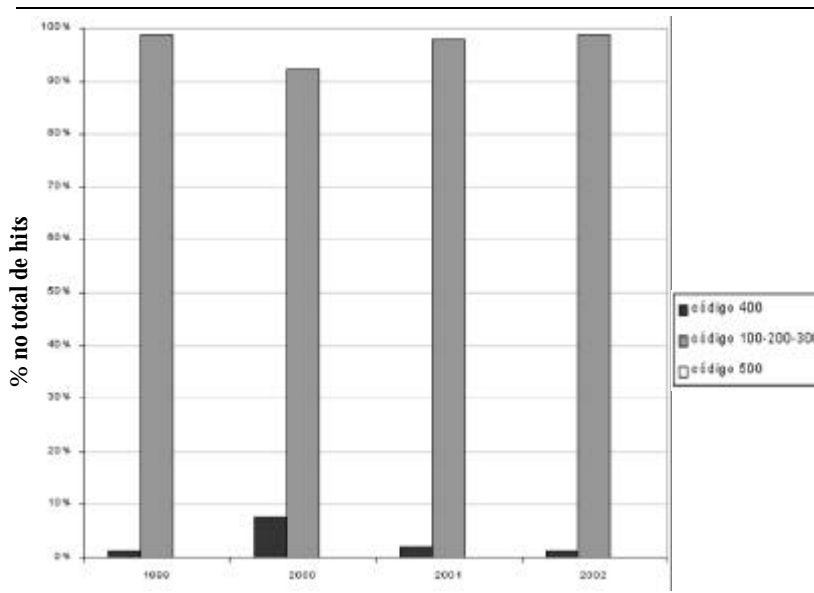
TABELA 4

Percentual de participação no total de *hits* dos principais agentes de acesso ao *web site* da Embrapa Monitoramento por Satélite, entre 1999 e 2002

Agente	Percentual no total de hits			
	1999	2000	2001	2002
Indefinido	0,58%	0,96%	0,32%	0,67%
Browser	92,46%	89,94%	88,50%	83,81%
Robô	6,05%	8,35%	6,11%	3,61%
Mirroring	0,90%	0,75%	5,07%	11,92%

FIGURA 8

Desempenho do sistema servidor do *web site* da Embrapa Monitoramento por Satélite, entre 1999 e 2002, segundo os códigos de sucesso e erro



disponibilização dos dados do projeto “Brasil Visto do Espaço”, que despertou no público o interesse de obtenção das informações, estimulando-o a copiar e transferir 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 se estar conectado à Internet durante o processo.

Desempenho do *web site*

A figura 8 representa a evolução temporal do desempenho do *web site*, segundo os códigos de *status*.

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 se mantém 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. *et alii*, 2000).

O aumento observado em 2000, para o percentual de códigos de erro tipo 400, pode ser explicado pela transformação ocorrida naquele ano no 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, para “Centro Nacional de Pesquisa de Monitoramento por Satélite”, sigla “CNPM” – URL: www.cnpm.embrapa.br. Essa mudança pode ter ocasionado sensível aumento dos erros tipo 400 – erros de sintaxe – quando os usuários, habituados a acessar o *web site* com sintaxes contendo “NMA”, passaram a não ter respostas de sucesso no acesso, mesmo com o mecanismo 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.

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.

Algumas conclusões emergem da análise anteriormente apresentada. 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 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*, relacionada 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 *robôs*;
- identificar, por meio 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*. [S. l. : s. n.], 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*. [S. l. : s. n.], 1997. Disponível em: < <ftp://ftp.mrunix.net/pub/webalizer/README>>. Acesso em: 15 jul. 2002.
- BERTOT, J. C. *et al.* 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 a 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, 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, v. 31, n. 1, p. 7-12, jan./abr. 2002.
- EICHMANN, D.; WU, J. Sulla. *A user agent for the web*. [S. l. : s. n.], 1996. Disponível em: < <http://mingo.info-science.uiowa.edu/eichmann/www-s96/Overview.html>>. Acesso em: 25 set. 2002.
- GNUS' not UNIX!. Disponível em: < <http://www.gnu.org>>. Acesso em: 24 set. 2002.
- HJELM, J. *Log analysis tools*. [S. l. : s. n.], 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.
- LEINER, B. M. *et al.* Brief history of the Internet. [S. l. : s. n.], 2000. Disponível em: <http://www.isoc.org/internet/history/brief.shtml>. Acesso em: 24 set. 2002.
- LOG files: apache http server. 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, 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. The history and development of transaction log analysis. *Library Hi Tech*, v. 11, n. 2, p. 41-50, 1993.
- _____. Remotely familiar: using computerized monitoring to study remote use. *Library Trends*, v. 47, p. 7-20, 1998.
- PIEROZZI JÚNIOR., I.; CAPUTI, E.; FILARDI, A. L. *A Internet como veículo de comunicação, difusão de resultados e imagem institucional na Embrapa Monitoramento por Satélite: I. Infra-estrutura e funcionamento*. Campinas : Embrapa Monitoramento por Satélite, 2000. 25 p.
- 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.
- SEARCH engine watch: tips about Internet search engines & search engine submission. Disponível em < <http://searchenginewatch.com>>. 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. [S. l. : s. n.], 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.