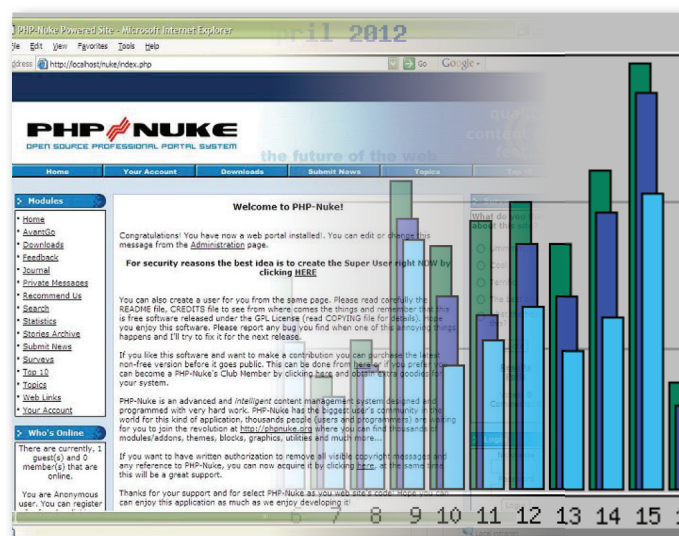




Análise de logs do sistema Agritempo por meio do log do PHPNuke e WebAlizer

Luciana Alvim Santos Romani¹
Adriano Franzoni Otaviano²
Thallita Dias Tomaz³

Nos últimos anos, a tecnologia da informação, em especial a web, tem provido mecanismos capazes de tornar disponível um grande volume de informações a baixo custo para um número expressivo de usuários (VASCONCELLOS-SILVA et al., 2003). Os *websites* fornecem uma ampla gama de conteúdos diferentes para todos os tipos de perfis de usuário. Diferentes formas de avaliação desse público têm sido propostas e usadas pelos administradores a fim de adaptar e aperfeiçoar o conteúdo a seu público-alvo. Uma das formas de monitoramento é a análise de arquivos de *log* (ALENCAR et al., 2003).

Um arquivo de *log* é gerado pelo servidor web e contém informações relativas ao endereço de *Internet Protocol* (IP) do cliente, data e hora do acesso, transcrição da solicitação efetuada pelo *browser* ao servidor, status da solicitação e quantidade de dados transferidos. Dependendo da configuração e da capacidade do servidor, é possível registrar ainda a página que levou o visitante ao *website* e o nome e versão do *browser* utilizado (COI, 2002). Esse tipo de análise é fundamental também para a constante manutenção do sistema,

uma vez que essa análise permite verificar quais são as informações existentes de maior e menor interesse dos usuários. Alguns sistemas gratuitos para análise de logs estão disponíveis como o *Webalizer*, o *Google Analytics*, o *Piwiki*.

O foco deste trabalho é a análise dos *logs* do sistema Agritempo, um sistema de monitoramento que disponibiliza informações meteorológicas e agrometeorológicas de diferentes regiões brasileiras gratuitamente na internet⁴. O Agritempo possui um amplo público-alvo: desde produtores, extensionistas, consultores, agentes do governo, estudantes e professores universitários, além da iniciativa privada.

Materiais e métodos

Nesse trabalho, foram analisados os *logs* armazenados pelo servidor Apache por meio da ferramenta PHP Nuke⁵, que apresenta a quantidade de *Page*

¹ Doutora em Ciência da Computação, Pesquisadora da Embrapa Informática Agropecuária, Campinas, SP, luciana.romani@embrapa.br

² Bacharel em Engenharia de Computação, Técnico de Nível Superior da Embrapa Informática Agropecuária, Campinas, SP, adriano.otaviano@embrapa.br

³ Estagiária da Embrapa Informática Agropecuária, thallita_thomaz@hotmail.com

⁴ Disponível em: <www.agritempo.gov.br>.

⁵ Disponível em: <http://phpnuke.org/>.

Views que houve no site em períodos de anos, meses, dias e horas. *Page View* é uma métrica utilizada para indicar a quantidade de acessos às páginas de um website. As estatísticas referentes a *Page Views* podem auxiliar na análise do perfil de acessos das páginas do website, em um determinado período, em determinadas horas do dia, entre outros critérios.

Outra ferramenta também avaliada nesse artigo é o Webalizer⁶, um aplicativo gratuito, que analisa os logs e gera estatísticas por meio de diversas variáveis como país de acesso, pastas e visitas. Nesse documento, as análises foram feitas considerando os *hits* dos anos de 2004, 2007, 2009, 2010 e 2011 nos meses de setembro e outubro de cada um deles. *Hits* são quaisquer pedidos feitos pelo servidor que está conectado, ou seja, representa o número total de pedidos feitos em um determinado espaço de tempo já estabelecido por ele.

Resultados e discussão

Estatísticas de *Page Views* do PHP Nuke

Os dados gerados pelo PHP Nuke, estão disponíveis na interface principal do sistema, não são armazenados os dados de *Page Views* dos campos “Consultas”, “Sumário” e “Acesso Restrito” do sistema Agritempo, pois foram desenvolvidos em módulos em linguagem Java separados da estrutura do PHPNuke. Com a análise dos logs de acesso do Agritempo, entre 2003 e 2010, pretende-se avaliar a usabilidade do sistema e também seu conteúdo, uma vez que um baixo número de *Page Views* pode indicar problemas no conteúdo do website. A Figura 1

apresenta uma comparação dos acessos anuais, onde a evolução da quantidade de páginas acessadas entre 2003 e 2010 é destacada, indicando um acentuado crescimento.

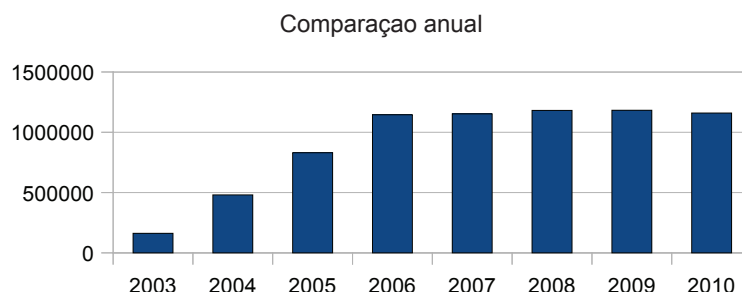


Figura 1. Comparação anual de *Page Views*.

A Figura 2 apresenta uma comparação das páginas visitadas no ano de 2007, indicando que no mês de agosto se dá o maior número de *Page Views*. Uma causa possível para esse perfil de acesso seria pelo fato de ser um mês de baixa taxa de umidade do ar e que precede ao mês de plantio; o que poderia gerar maior interesse nos usuários por consultas a informações agrometeorológicas e à previsão de chuvas.

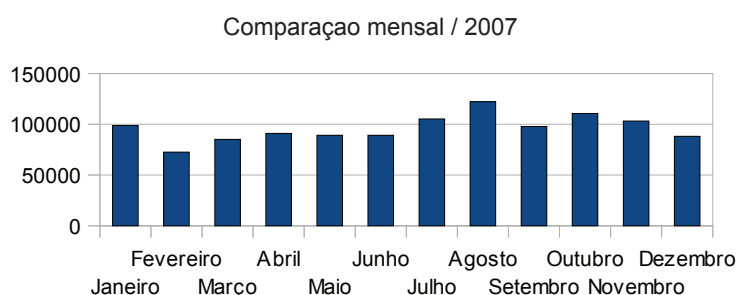


Figura 2. Comparação mensal de *Page Views*.

A análise dos acessos diários do mês de agosto de 2007, apresentada na Figura 3, demonstra constância nos acessos diários ao site durante os dias úteis, com ligeira queda nos fins de semana.

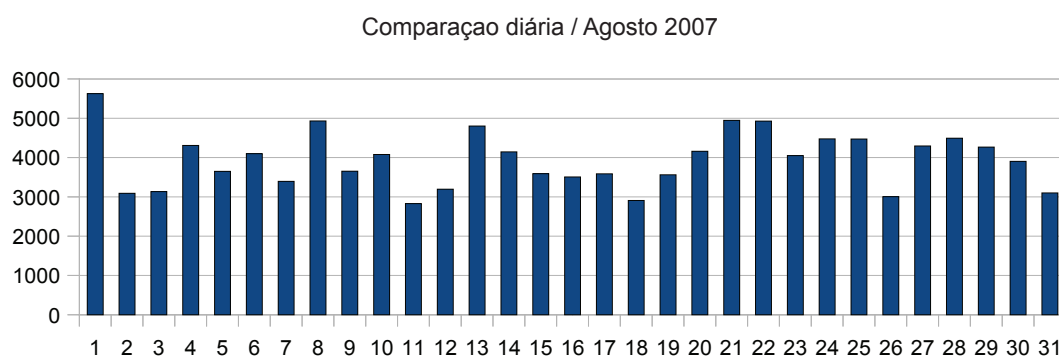


Figura 3. Comparação diária de *Page Views*.

⁶ Disponível em: <<http://www.webalizer.org/>>.

Avaliando detalhadamente os acessos realizados em dias da semana, nota-se que, embora haja acessos em todos os horários, o número de *Page Views* é maior em dias úteis do que nos fins de semana, como pode ser visto nas Figuras 4 e 5.

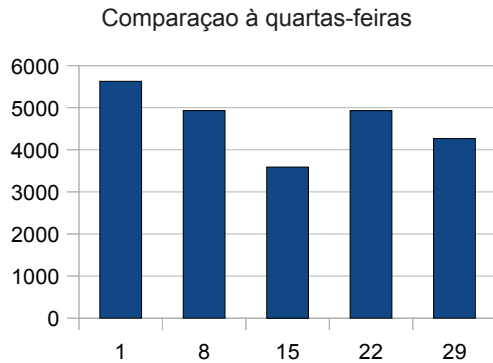


Figura 4. *Page Views* às quartas-feiras.

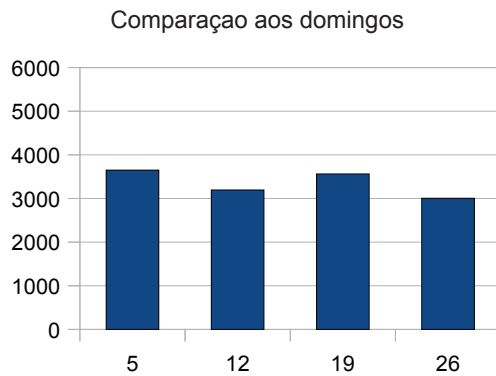


Figura 5. *Page Views* aos domingos.

Relatório de *hits* do Webalizer

Os relatórios gerados pelo *Webalizer* mostram um aumento no número de *hits* de 2004 para 2007, passando de 23500 em setembro de 2004 para 51800 no mesmo mês de 2007, como pode ser visto nas Figuras 6 e 7.

O aumento no número de *hits* (Figuras 6 e 7) mostra um crescimento nos acessos ao site após 4 anos de sistema em funcionamento. Entre os anos de 2009

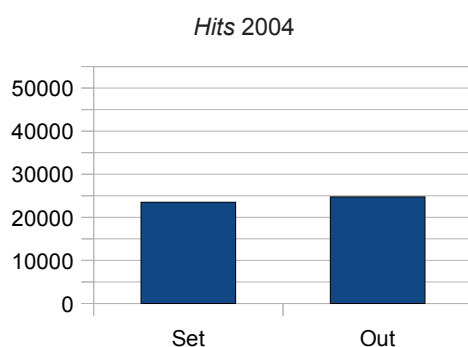


Figura 6. *Hits* do ano de 2004.

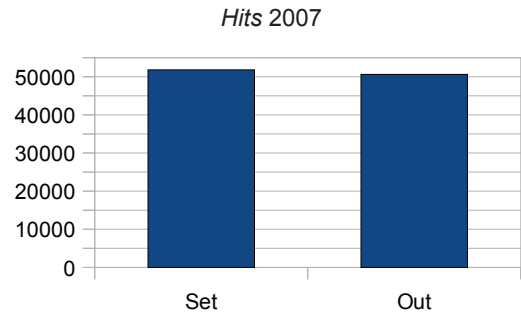


Figura 7. *Hits* do ano de 2007.

e 2011, em contrapartida, a análise feita usando o *Webalizer* indica uma queda no número de *hits* (Figuras 8, 9 e 10). Durante esse período, os arquivos de *log* não foram armazenados de forma integral, o que prejudica uma análise mais detalhada dos acessos. Além disso, uma falta de atualização das páginas do sistema

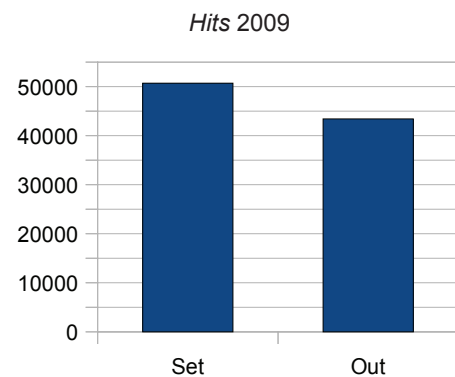


Figura 8. *Hits* do ano de 2009.

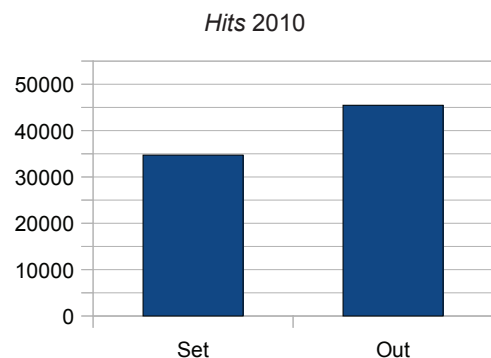


Figura 9. *Hits* do ano de 2010.

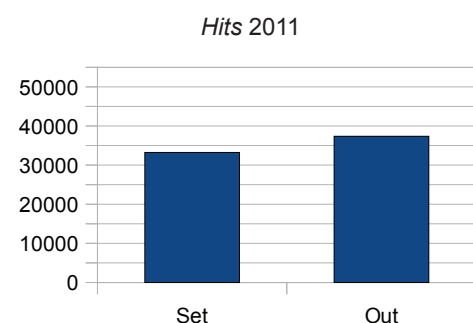


Figura 10. *Hits* do ano de 2011.

Agritempo também pode ter motivado um não aumento no número de visitas ao site.

Com isso, este trabalho de análise dos *logs* mostra que há uma necessidade de armazenamento dos *logs* do sistema, de forma a possibilitar uma análise mais adequada e detalhada das consultas feitas pelo usuário. Além disso, corrobora a necessidade de atualização do sistema para uma versão mais moderna e alinhada com conceitos da web 2.0.

Outra variável importante que o *Webalizer* permite analisar é o *ranking* de páginas acessadas por mês. Por meio de uma análise comparativa para três anos (2004, 2007 e 2010), foi possível verificar quais páginas foram mais acessadas e o que mudou nesse período em relação a elas, como pode ser visto nas Tabelas 1, 2 e 3. Nessas tabelas são apresentadas análises comparativas do mês de setembro de cada ano indicado, para que, assim, haja uma base em relação a outros meses

e anos. Nelas pode-se verificar que o conteúdo das páginas acessadas pouco se altera com o passar dos anos. Uma justificativa coerente para isso é que elas são os produtos do Agritempo de maior interesse dos usuários.

Para que melhore cada vez mais o sistema em termos de atualizações e internacionalização, saber a origem do acesso é fundamental. No ano de 2010 durante os meses de setembro e outubro, foram identificados os países que acessaram o sistema como mostra a Tabela 4.

Assim como todos os outros meses e anos, a predominância de acessos é do Brasil e dos Estados Unidos, respectivamente. Os números são dados em porcentagem, porém na tabela acima eles foram transformados em relação ao número total de acessos de cada mês. O *Webalizer* não especifica alguns países agrupando-os em “Desconhecido” e “Outros”.

Tabela 1. Ranking das páginas mais acessadas no mês de setembro/2004.

Ranking	2004	Ranking	2004
1	Interface principal do sistema	9	Mapas de produção do IBGE sidra
2	Pesquisa de dados de estações	10	Mapa de previsão de SP
3	Sumário do sistema	11	Página de acesso restrito
4	Geração de gráficos do sistema	12	Mapa de previsão de MG
5	Página de boletins agrometeorológicos	13	Mapa de previsão de PR
6	Visualização de mapas diários	14	Consulta para o RS do El Niño/La Niña
7	Página de mapa de probabilidade de chuva	15	Mapa de previsão de MS
8	Árvore hiperbólica do zoneamento		

Tabela 2. Ranking das páginas mais acessadas no mês de setembro/2007.

Ranking	2007
1	Interface principal do sistema
2	Download do livro Meteorologia e Climatologia de Mario Varejão
3	Pesquisas de dados de estações
4	Sumário do sistema
5	Visualização de mapas diários
6	Sumário de séries históricas de chuvas
7	Geração de gráficos do sistema
8	Visualização de mapas de séries históricas de chuvas
9	Página de boletins meteorológicos
10	Boletim regional (Sudeste)
11	Mapa de previsão de SP
12	Boletim de zoneamento da região Sudeste
13	Boletim regional (Sul)
14	Mapa de previsão de MG
15	Mapa de previsão de PR

Tabela 3. Ranking das páginas mais acessadas no mês de setembro/2010.

Ranking	2010
1	Interface principal do sistema
2	Download do livro Meteorologia e Climatologia de Mario Varejão
3	Sumário do sistema
4	Pesquisa de dados de estações
5	Mapas diários
6	Manual do sistema
7	Sumário de séries históricas de chuvas
8	Visualização de mapas de séries históricas de chuvas
9	Geração de gráficos do sistema
10	Mapa de previsão de SP
11	Mapa diário de PR
12	Mapa com todas as estações referenciadas
13	Mapa de probabilidade de chuva
14	Boletim regional (Sudeste)
15	Mapa diário de MG

Tabela 4. Países que mais acessaram o Agritempo em set/out/2010.

2010	Setembro	Outubro	Total
Brasil	19788,69	27284,40	47073,09
EUA	17011,33	16825,38	33836,71
Desconhecido	347,17	454,74	801,91
Outros	1041,51	909,48	1950,99
Rússia	0,00	454,74	454,74

Conclusão

A análise dos *logs* do Agritempo, por meio da ferramenta de análise do sistema PHP-Nuke e *Webalizer*, permitiu identificar algumas tendências e padrões de uso do sistema pelos usuários. Os resultados desse tipo de análise podem auxiliar na definição de melhorias a serem incorporadas ao sistema e na proposição de novos recursos para melhor atender às necessidades do seu público-alvo.

Devido ao armazenamento incorreto dos *logs* no servidor, podemos concluir que a ferramenta mais eficaz na análise dos *logs* do sistema Agritempo, atualmente, é a realizada pelo PHP Nuke.

Referências

- ALENCAR, M. C. F.; GOMES, E. G.; PIEROZZI JUNIOR, I. Uso do website da Embrapa Monitoramento por Satélite para educação: perfil de usuário e análise de conteúdo. **Contrapontos**, v. 3, n. 1, p. 27-38, 2003.
- BARRET, B. L. **The Webalizer** – A web server log file analysis tool. 1997. Disponível em: <ftp://ftp.mrunix.net/pub/webalizer/README>. Acesso em: 15 out. 2012.
- COI **Website evaluation, usage and analysis**, Cabinet Office UK: UK, 2002.
- VASCONCELLOS-SILVA, P. R.; CASTIEL, L. D.; RIVERA, F. J. U. Assessing an internet health information site by using log analysis: the experience of the National Cancer Institute of Brazil. **Revista Panamericana de Salud Publica**, v. 14, n. 2, Aug. 2003.

Comunicado Técnico, 114

Embrapa Informática Agropecuária
Endereço: Caixa Postal 6041 - Barão Geraldo
13083-886 - Campinas, SP
Fone: (19) 3211-5700
Fax: (19) 3211-5754
http://www.cnptia.embrapa.br
e-mail: cnptia.sac@embrapa.com.br



Ministério da
Agricultura, Pecuária
e Abastecimento



1ª edição on-line - 2012

Todos os direitos reservados.

Comitê de Publicações

Presidente: Sílvia Maria Fonseca Silveira Massruhá
Membros: Adhemar Zerlotini Neto, Stanley Robson de Medeiros Oliveira, Thiago Teixeira Santos, Maria Goretti Gurgel Praxedes, Adriana Farah Gonzalez, Neide Makiko Furukawa, Carla Cristiane Osawa (Secretária)
Suplentes: Felipe Rodrigues da Silva, José Ruy Porto de Carvalho, Eduardo Delgado Assad, Fábio César da Silva

Expediente

Supervisão editorial: Stanley Robson de Medeiros Oliveira, Neide Makiko Furukawa
Normalização bibliográfica: Maria Goretti Gurgel Praxedes
Revisão de texto: Adriana Farah Gonzalez
Editoração eletrônica: Neide Makiko Furukawa