

# Augusto Campos

São 2 da manhã. Você sabe o que o seu website está fazendo?

Um aspecto muito importante, embora às vezes negligenciado, da manutenção de servidores Web – seja na Internet, em uma Intranet ou no seu pequeno servidor doméstico – é a análise dos logs. É impressionante quanta informação útil e interessante pode ser obtida a partir dos registros de acesso, tanto no que diz respeito a análises de tráfego e segurança, quanto em aspectos relacionados ao conteúdo.

A partir dos logs você pode descobrir o que seus novos usuários estavam procurando em páginas de busca quando chegaram ao seu site, ou que links de outros sites eles seguiram para chegar ao seu. Você pode descobrir quais as suas páginas mais populares, com que navegadores e sistemas operacionais o seu site é acessado, e cruzar todas essas informações entre si, com os dias da semana, com os diversos horários ou com muitos outros dados ao seu alcance.

Claro que é possível, para um administrador de sistemas experiente, desenvolver sua própria ferramenta de análise personalizada, usando linguagens e ferramentas comuns para manipulação de texto estruturado (como AWK ou Perl). Mas é muito mais fácil e rápido adotar uma das muitas soluções prontas para essa finalidade, disponíveis como software livre.

A minha ferramenta livre preferida para análise de logs do Apache é o AWStats [1], capaz de gerar estatísticas e gráficos bastante avançados de servidores Web, FTP, streaming e e-mail. Operando a partir do próprio servidor Web (como um script CGI) ou diretamente a partir da linha de comando, ele tem recursos bastante interessantes, incluindo o armazenamento de dados históricos para facilitar o processamento freqüente de arquivos volumosos, suporte a servidores virtuais, suporte ao nosso idioma e possibilidade de rodar sem uso de bibliotecas Perl raramente instaladas por default em servidores.

Além das informações mais básicas, como número de páginas exibidas e número de visitantes únicos, o AWStats também calcula e exibe vários dados derivados ou cumulativos, como o tempo médio de duração das visitas ao seu site, análise por horário e dia da semana, análise por domínio, país e localização geográfica (GeoIP) do usuário, páginas mais acessadas, quais as páginas pelas quais os usuários entram ou saem mais freqüentemente do seu site e visitas de robôs, indexadores e worms. Também é possível identificar pesquisas usadas em sites de busca para localizar seu site, inclusão de seu site nos favoritos de

um determinado navegador, tamanho da tela do usuário, percentual de navegadores de seus usuários com suporte a Java, Flash e PDF. Alguns destes cálculos, entretanto, exigem configurações específicas – que constam na documentação do sistema.

Como você pode ver, uma boa ferramenta de análise de logs dá ao webmaster e ao webdesigner informações importantíssimas para melhor desenvolver e direcionar seus serviços e conteúdo. Entretanto, felizmente o AWStats não

---

*Uma boa ferramenta de análise de logs dá ao webmaster informações importantíssimas.*

---

é a única opção; há diversas outras alternativas em software livre. Investigue o popular (mas pouco atualizado) *webalizer* [2], o seu irmão mais novo *webalizer xtended* [3] e o veteraníssimo *analog* [4], que completou 10 anos em 2005.

Se você prefere uma solução rápida e não se preocupa com o fato de informações sobre o tráfego de seu site ficarem na posse de outras empresas, uma alternativa adicional é o *Google Analytics* [5], que exibe praticamente todas as informações relevantes sobre o tráfego do seu site a partir de uma configuração especial (e simples) em seu conteúdo. Para aplicações pessoais, pode ser uma alternativa interessante, mas para uso corporativo as desvantagens superam a vantagem da praticidade.

Não fique no escuro, sempre saiba o que o seu servidor Web fez na madrugada passada ;-). ■

## Mais Informações

[1] <http://awstats.sourceforge.net/>

[2] <http://www.mrunix.net/webalizer/>

[3] <http://www.patrickfrei.ch/webalizer>

[4] <http://www.analog.cx/>

[5] <http://www.google.com/analytics/>

## O autor

**Augusto César Campos** é administrador de TI e, desde 1996, mantém o site [BR-linux.org](http://BR-linux.org), que cobre a cena do Software Livre no Brasil e no mundo.

