

**UNIVERSIDADE DO EXTREMO SUL CATARINENSE
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

MARCELO PEREIRA CARRADORE

**ANÁLISE DE *LOGS* EM SERVIDORES WEB: ESTUDO DE CASO NA
PREFEITURA MUNICIPAL DE CRICIÚMA**

CRICIÚMA, JUNHO DE 2006.

MARCELO PEREIRA CARRADORE

**ANÁLISE DE LOGS EM SERVIDORES WEB: ESTUDO DE CASO NA
PREFEITURA MUNICIPAL DE CRICIÚMA**

Trabalho de Conclusão de Curso para a
Obtenção do Grau de Bacharel em Ciência da
Computação da Universidade do Extremo Sul
Catarinense.

Orientador: Prof. M.Sc. Paulo João Martins

CRICIÚMA, JUNHO DE 2006.

Dedico esse trabalho ao meu avô, Mario Romão Pereira (*In Memoriam*), por ter trabalhado a vida inteira para proporcionar o melhor para as suas filhas, e especialmente aos meus pais, pela compreensão e estímulo dado ao longo dessa caminhada.

AGRADECIMENTOS

Em primeiro lugar gostaria de agradecer a Deus por ter iluminado o meu caminho, além de me dar saúde e sabedoria.

Aos meus pais por estarem sempre ao meu lado.

A minha namorada por toda sua atenção e carinho por mim.

Ao meu orientador pela atenção dada ao longo desse trabalho e também aos outros professores do curso, pelo conhecimento que me foi passado.

Aos meus amigos, que de alguma forma me ajudaram a superar as dificuldades ao longo do curso e aos familiares que auxiliaram na minha formação pessoal.

*“A vida só pode ser comprendida olhando-se
para trás, mas só pode ser vivida olhando-se
para frente”.*

Soren Kierkegaard

RESUMO

A cada ano, aumenta o número de pessoas utilizando a Internet. Essas pessoas por sua vez, enviam e-mails, conversam com o uso de programas de mensagens instantâneas, visualizam vários *sites*, além de outros serviços e produtos que são oferecidos por este mecanismo. Um servidor Web é o responsável por manter os *sites* com seus serviços 24 horas por dia disponível e em responder as requisições dos *browsers* dos usuários quando estes estão visualizando uma página que esteja sendo hospedada nele. Esse tipo de servidor recebe por dia, diversas requisições e nem sempre a pessoa que o administra possui um controle de quem está acessando e quais recursos estão sendo usados. Por esse motivo, o presente trabalho realizou uma análise dos arquivos de *logs* do servidor Web da Prefeitura Municipal de Criciúma, para que, com isso, possa definir o perfil dos acessos ao site da mesma.

Palavras-Chave: Análise de *log*, Servidor Web, Monitoramento.

ABSTRACT

Each year, it increases the number of people using the Internet. These people in turn, send e-mails, talk with programs of instantaneous messages, visualize some sites, beyond other services and products that are offered by this mechanism. A Web server is the responsible for keeping the sites with its services 24 hours a day available and in answering the solicitations of browsers of the users when these are visualizing a page that is being housed in it. This type of server receives per day, several solicitations and not always the person who manages it possesss a control of who is having access and which resources are being used. For this reason, the present work carried through an analysis of the archives of logs of the Web server of the City Hall of Criciúma, so that, with this, it can define the profile of the accesses to the site of City of Hall Criciúma.

Key-Words: Analysis of log, Web Server, Monitoring.

LISTA DE ILUSTRAÇÕES

Figura 1. Evolução do Número de <i>Hosts</i> no Brasil.....	18
Figura 2. <i>Hosts</i> na América do Sul.....	24
Figura 3. Página da Prefeitura Municipal de Criciúma.....	29
Figura 4. <i>Blog</i> da Secretaria de Comunicação.....	30
Figura 5. Servidor de Internet.....	31
Figura 6. Servidor Web.....	32
Figura 7. Página Inicial do Servidor Web Apache.....	34
Figura 8. Arquivo AccessLog.....	41
Figura 9. Arquivo AccessLog.....	55
Figura 10. Informação Mensal dos Acessos.....	59
Figura 11. Informações dos Acessos no Apache.....	60
Figura 12. Informações Mostradas pelo Webalizer.....	61
Figura 13. Códigos de Resposta ao Browsers.....	61
Figura 14. Informações do Webalizer em Gráfico Separadas por Dia.....	62
Figura 15. Informações do Webalizer em Texto Separadas por Dia.....	63
Figura 16. Informações do Webalizer em Gráfico Separadas por Hora.....	63
Figura 17. Informações do Webalizer em Texto Separadas por Hora.....	64
Figura 18. <i>Links</i> mais Visualizados.....	65
Figura 19. <i>Links</i> que mais Trafegam.....	66
Figura 20. Páginas de Entrada do mês de Maio.....	66
Figura 21. Páginas de Saída do mês de Maio.....	67
Figura 22. IP dos Visitantes ao <i>Site</i>	68
Figura 23. <i>Sites</i> da Internet que referenciam a página da PMC.....	68
Figura 24. Palavras que foram utilizadas para encontrar o <i>site</i> em Maio.....	69
Figura 25. Informações Exibidas pelo http-análise em Gráfico.....	70
Figura 26. Informações Exibidas pelo http-análise em Texto.....	71
Figura 27. Informações Exibidas pelo http-analyze.....	72
Figura 28. Outras Informações Exibidas pelo http-analyze.....	72
Figura 29. Mais Algumas Informações Exibidas pelo http-analyze.....	73
Figura 30. Informações Gerais do Analog.....	74
Figura 31. Informações Classificada por Dias da Semana.....	75
Figura 32. Palavras Usadas para Encontrar o Site no Servidor Web.....	75
Figura 33. <i>Browsers</i> Utilizados pelos Visitantes.....	76
Figura 34. Informações Mostradas pelo Awstats.....	77
Figura 35. Informações Separadas por Dia da Semana do Mês de Maio.....	78
Figura 36. Informações Separadas por Dia da Semana.....	78
Figura 37. Visitas Únicas em Março.....	81
Figura 38. Visitas Únicas em Abril.....	81
Figura 39. Conectados a Partir De.....	83
Figura 40. Páginas de Entrada ao <i>Site</i> no Mês de Março.....	83
Figura 41. Páginas de Entrada ao <i>Site</i> no Mês de Abril.....	84
Figura 42. Páginas de Saída ao <i>Site</i> no Mês de Março.....	84
Figura 43. Páginas de Saída ao <i>Site</i> no Mês de Abril.....	85
Figura 44. Palavras que foram utilizadas para encontrar o <i>site</i> em Março.....	85
Figura 45. Palavras que foram utilizadas para encontrar o <i>site</i> em Abril.....	86
Figura 46. <i>Browser</i> utilizado pelos visitantes ao <i>site</i> no mês de Março.....	86

Figura 47. <i>Browser</i> utilizado pelos visitantes ao <i>site</i> no mês de Abril.....	87
Figura 48. <i>Browser</i> utilizado pelos visitantes ao <i>site</i> no mês de Maio	87
Figura 49. Sistema Operacional utilizado pelos visitantes em Maio.....	88
Figura 50. Versão dos Sistemas Operacionais utilizados em Maio.....	88
Figura 51. Tempo em que um visitante fica no <i>site</i> da PMC no mês de Maio ..	89
Figura 52. Informações Separadas por Hora do Mês de Maio	90
Figura 53. Página inicial do servidor Web Apache.....	97

LISTA DE TABELAS

Tabela 1. Campos de Entrada de um <i>Log</i> CLF	37
Tabela 2. Campos Adicionais do ECLF	37
Tabela 3. Códigos de Estado Retornados pelo Servidor	38
Tabela 4. Opções de configuração do servidor Web Apache	39
Tabela 5. Informações sobre os softwares instalados	58
Tabela 6. Para onde serão disseminadas as informações	92

LISTA DE SIGLAS

ARPA	<i>Advanced Research Projects Agency</i>
B2B	<i>Business to Business</i>
B2C	<i>Business to Consumer</i>
CEGE	Comitê Executivo de Governo Eletrônico
CGI	<i>Common Gateway Interface</i>
CLF	<i>Common Log Format</i>
E-Governo	Governo Eletrônico
ECLF	<i>Extended Common Log Format</i>
EUA	Estados Unidos da América
G2G	Governo para Governo
GTTI	Grupo de Trabalho em Tecnologia da Informação
HTML	<i>HyperText Markup Language</i>
HTTP	<i>HyperText Transfer Protocol</i>
HTTPD	<i>HyperText Transfer Protocol Daemon</i>
HTTPS	<i>HyperText Transfer Protocol Secure</i>
IP	<i>Internet Protocol</i>
IPTU	Imposto Sobre a Propriedade Predial ou Territorial Urbana
NCSA	<i>National Center for Supercomputing Applications</i>
PMC	Prefeitura Municipal de Criciúma
RFC	<i>Request for Comments</i>
SSL	<i>Secure Sockets Layer</i>
URI	<i>Universal Resource Identifier</i>
URL	<i>Universal Resource Locator</i>

TCP	<i>Transmission Control Protocol</i>
URSS	União das Repúblicas Socialistas Soviéticas
WWW	<i>World Wide Web</i>

SUMÁRIO

1 INTRODUÇÃO	14
1.1 OBJETIVO GERAL.....	15
1.2 OBJETIVOS ESPECÍFICOS	15
1.3 JUSTIFICATIVA.....	15
1.4 ESTRUTURA DO TRABALHO.....	16
2 A IMPORTÂNCIA DA INTERNET NOS DIAS DE HOJE.....	18
2.1 COMÉRCIO ELETRÔNICO.....	20
2.1.1 E-Commerce Seguro	22
2.2 E-GOVERNO	23
2.2.1 E-Governo do Brasil.....	25
2.2.2 Os Benefícios.....	28
2.2.3 O E-Governo na Prefeitura Municipal de Criciúma.....	28
3 SERVIDORES.....	31
3.1 SERVIDORES WEB.....	31
3.2 SERVIDOR WEB APACHE	33
3.3 MONITORAMENTO	36
3.4 ARQUIVOS DE LOGS	36
3.5 A IMPORTÂNCIA DO MONITORAMENTO E ANÁLISE DOS LOGS.....	41
4 MÉTRICAS ELETRÔNICAS	43
4.1 MEDIÇÕES CENTRADAS NO <i>SITE</i>	44
4.2 MEDIÇÃO CENTRADA NO USUÁRIO	48
4.3 TÉCNICAS DE MEDIÇÃO	50
4.3.1 <i>Logging</i> no Servidor	50
4.3.2 <i>Logging</i> no <i>Proxy</i>	51
4.3.3 <i>Logging</i> no Cliente	52
4.3.4 Monitoração de Pacotes	53
4.3.5 Medição Ativa	54
4.4 ANTES DA ANÁLISE PROPRIAMENTE DITA.....	55
5 ANÁLISE NOS ARQUIVOS DE LOG DA PMC	57
5.1 SOFTWARES UTILIZADOS NO DESENVOLVIMENTO	58
5.1.1 Webalizer	59
5.1.2 http-analyze.....	69
5.1.3 analog.....	74
5.1.4 Awstats	76
5.2 MÉTRICAS UTILIZADAS E RESULTADOS OBTIDOS	79
CONCLUSÃO.....	93
REFERÊNCIAS.....	94
APÊNDICE A – INSTALAÇÃO E CONFIGURAÇÃO DO APACHE	97

1 INTRODUÇÃO

A Internet está cada vez mais se popularizando, devido ao crescente uso dos serviços disponíveis tais como os jogos on-line, e-mails, listas de discussões, entre outros. Desta forma, surge a necessidade de obter um diagnóstico sobre os serviços, como por exemplo, o tráfego de utilização do servidor Web em uma determinada empresa.

O governo e o comércio eletrônico têm contribuído para aumentar a utilização da Internet. Como se sabe, o Brasil é o segundo país do mundo a utilizar a conta bancária por este meio, além também de que os Estados estão sempre fazendo com que a população cada vez mais se habitue às novas tecnologias, como por exemplo, o Programa Poupa Tempo do Governo do Estado de São Paulo (FERRER; SANTOS, 2004).

Nos dias atuais, uma organização que deseja estar com o foco para o cliente, deve ao menos tentar possuir uma página na Internet. Isso é necessário, pois é preciso possuir informações que fiquem 24 horas à disposição de seus clientes ou de quem se possa interessar (NAKAMURA, 2001).

Os arquivos de *logs* mantêm as informações sobre os processos pertinentes aos acontecimentos do sistema. Eles estão para servir o administrador do sistema, pois o ajudam a compreender melhor o que aconteceu no servidor. Em servidores Web, todas as informações que trafegaram por ele, como os erros e as informações gerais, são notificados nesses arquivos.

Quando se fala em servidores Web, a quantidade de acesso é intensa e esses arquivos se tornam enormes, dificultando uma análise manual.

Dessa forma, este trabalho contempla o estudo dos arquivos de *log* do servidor Web da Prefeitura Municipal de Criciúma (PMC), com o uso de alguma ferramenta computacional, de forma a descrever as métricas que foram analisadas a fim de compreender sobre os serviços e os acessos às informações disponíveis no *site*.

1.1 OBJETIVO GERAL

Analisar os arquivos de *logs* em servidores Web, para conseqüentemente definir o perfil dos acessos ao *site* da Prefeitura Municipal de Criciúma.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos consistem em:

- a) descrever sobre o governo e o comércio eletrônico;
- b) análise de arquivos de *log* em servidores Web;
- c) descrever o estado de caso a ser analisado;
- d) definir a métrica a ser analisada no servidor Web.

1.3 JUSTIFICATIVA

Os arquivos de *logs* possuem um grande volume de informações, portanto monitorar estes manualmente é uma tarefa dispendiosa. Para isso, existem ferramentas computacionais que podem auxiliar o responsável pela administração do servidor Web, na forma de monitorar os acessos ao mesmo, pois faz com que os dados inseridos nesse arquivo sejam transformados em dados estatísticos.

A Prefeitura Municipal de Criciúma disponibiliza informações e serviços em seu servidor Web, e esse tipo de análise atualmente não é realizado e não existe documentação a respeito dos acessos ao *site*. A fim de se obter algumas informações, este trabalho se propõe a realizar a análise e a documentação dos *logs*, de forma a auxiliar o administrador de rede a conhecer o perfil dos acessos e de outra forma auxiliar na tomada de decisão sobre quais serviços estão sendo realmente utilizados e o quanto deles. A forma escolhida foi a utilização de uma ferramenta computacional com o propósito de obter a real utilização do *site* e, a partir disso, propor alguma mudança, ou justificar a existência dos serviços diante da demanda da utilização dos serviços, e quem sabe, uma readequação na forma de estabelecer uma otimização das informações no atendimento aos usuários.

A partir dos usos dessas ferramentas, o administrador poderá obter estatísticas referentes ao servidor Web, e assim utilizá-las para saber qual o perfil dos acessos ao *site* da PMC.

1.4 ESTRUTURA DO TRABALHO

Essa pesquisa de Análise de *Logs* de Servidores Web – Estudo de caso na Prefeitura Municipal de Criciúma é apresentado em quatro capítulos de fundamentação teórica, um capítulo sobre o trabalho desenvolvido, as considerações finais e apêndice.

O capítulo dois fala sobre como a Internet teve seu início e qual é a sua importância para o mundo atual. Explica também sobre o comércio eletrônico e o governo eletrônico.

No capítulo três tem-se o estudo de servidores, abordando-se mais detalhadamente os servidores Web. Nesse capítulo também se comenta sobre

monitoramento de servidores, arquivos de *logs* e a importância do administrador em monitorar o servidor.

A fundamentação sobre métricas eletrônicas é realizada no quarto capítulo, destacando-se seus tipos e técnicas de medições.

O trabalho desenvolvido, que envolve a análise dos *logs*, é apresentado no capítulo 5. Neste capítulo são descritas as etapas do trabalho realizado, bem como os resultados atingidos.

A conclusão apresenta os comentários dos resultados obtidos com esta pesquisa e as sugestões para futuros trabalhos.

O Apêndice A (Instalação e Configuração do Apache) apresenta como instalar e configurar o servidor Web Apache, dando ênfase nas configurações de *logging*.

2 A IMPORTÂNCIA DA INTERNET NOS DIAS DE HOJE

Conforme Demétrio (2001), a Internet é uma grande rede mundial de computadores interligados, onde é permitido que usuários de qualquer lugar do mundo troquem informações. Essa rede de computadores é composta de diversas redes menores, como as dos bancos, das universidades, das escolas, dos hospitais, entre outras. Faz parte da Internet, todo e qualquer computador ligado a ela, e, conforme a Figura 1, a cada ano, mais e mais computadores são conectados a ela.

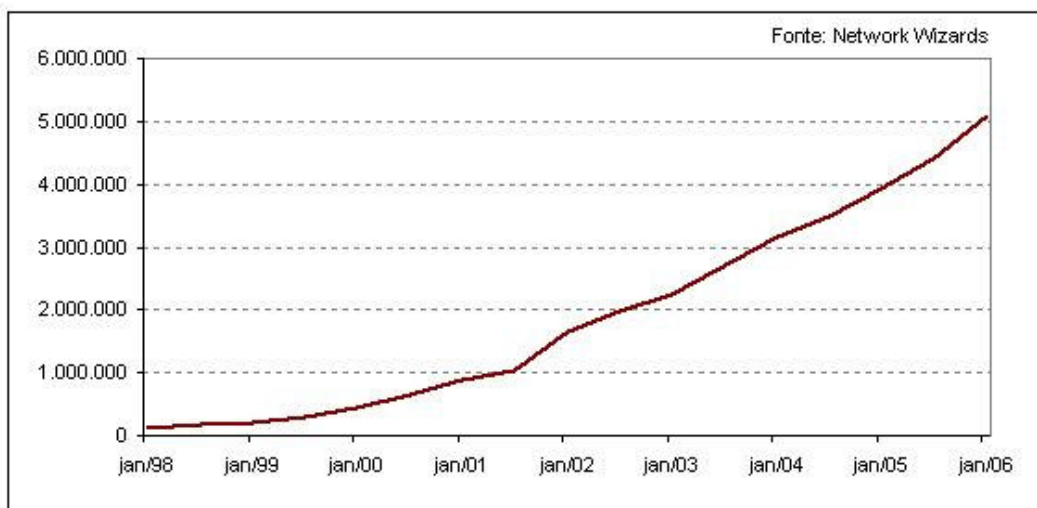


Figura 1. Evolução do Número de *Hosts* no Brasil
Fonte: NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO (2006)

O início da Internet se deve à guerra fria. Na década de 60, no tempo de guerra fria entre os Estados Unidos da América (EUA) e a União das Repúblicas Socialistas Soviéticas (URSS), esses dois países estavam em uma corrida para provar quem tinha a maior indústria de tecnologia. Os Russos estavam na frente por alguns fatos, como por exemplo: ter o primeiro homem a ir ao espaço e colocar em funcionamento o primeiro satélite. Para contra-atacar e tentar virar o jogo, os EUA criaram um órgão que seria responsável por fazer com que o país se tornasse o número

um quando o assunto fosse tecnologia militar de ponta. Com isso foi criado um departamento chamado de *Advanced Research Projects Agency* (ARPA). Nela ficavam as melhores cabeças pensantes do estado. No decorrer do trabalho, alguns problemas foram ocorrendo, como, por exemplo, não era possível manter sempre em funcionamento a comunicação entre os cientistas de diversas partes do país (DEMÉTRIO, 2001).

Para resolver esse problema, foi dada a idéia de fazer com que os computadores estivessem interligados por algum meio, e com isso, tornar possível a troca de informações entre eles. Com algum tempo passado, um cientista criou o conceito de troca de informações entre os computadores: informações divididas e armazenadas em pacotes. Depois de diversos estudos, foi descoberto que para ter os computadores interligados seria necessário utilizar as redes telefônicas, que na época eram consideradas precárias, desenvolver um programa que fizesse a comunicação entre os diversos tipos de computadores e que seria necessário também de um modem¹ que fizesse com que os dados trafegassem através da linha telefônica (DEMÉTRIO, 2001).

Com algumas modificações na rede telefônica e com a criação e utilização de um protocolo, o TCP/IP², que foi o responsável por fazer os computadores se comunicarem, foi feito o primeiro teste, que envolveu a conexão entre duas universidades. Nesse teste foi enviada uma palavra entre os computadores. A partir desse teste, outros computadores foram sendo agregados a essa nova rede, que, no começo, era conhecida como Arpanet.

¹ De modulador e demodulador; É um dispositivo eletrônico que converte a informação digital dos computadores em sinais adequados para trafegar através de linhas telefônicas (ZELENOVSKY; MENDONÇA, 2001).

² TCP/IP: *Transmission Control Protocol/ Internet Protocol*. Os dois protocolos básicos da Internet, usados para viabilizar a transmissão e troca de dados referentes, permitindo assim que os computadores se comuniquem.

Ainda de acordo com Demétrio (2001), com o passar do tempo, o nome da rede foi se modificando, tendo-se no final o nome Internet.

Hoje em dia, qualquer pessoa que tenha acesso a um computador conectado à Internet, pode acessar recursos em centenas de outros computadores (HEIDE; STILBORNE, 1999).

O número de pessoas no mundo que moram em domicílios com acesso à Internet via computador doméstico, já ultrapassa os 400 milhões (NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO, 2006).

A Internet passou a ser, nos últimos anos, muito mais do que diversão. Cada vez mais e mais empresas estão com suas vitrines à mostra e as pessoas que antes tinham medo de comprar ou utilizar esse meio, o estão perdendo. Muita coisa pode ser encontrada na rede mundial de computadores, como grandes empresas, gestões governamentais a até bancas de jornal.

Apesar desta já existir há algum um tempo, somente nos últimos anos, com a sua comercialização, é que ela conseguiu influenciar mais as pessoas. Isso porque, com esse tipo de acesso, os usuários podem fazer teleconferências por meio do computador. Conforme Mello (2004), na área de marketing a Internet é muito importante, pelo motivo de que os *sites* possuem o papel fundamental de informar, entreter e interagir com o usuário.

2.1 COMÉRCIO ELETRÔNICO

Comércio eletrônico é toda atividade de compra e venda realizada com o auxílio de recursos eletrônicos. Uma venda concretizada por fax, negociada por e-mail

ou mesmo uma ligação telefônica é uma atividade de comércio eletrônico (NAKAMURA, 2001).

De acordo com Ferrer e Santos (2004), a partir de 1990, o Brasil começou a entender que era necessário desenvolver uma estratégia para expandir o seu comércio mundial. Com isso, fez com que o comércio eletrônico começasse a fazer parte do dia-a-dia da população. Para fazer com que o povo brasileiro se adaptasse com o comércio eletrônico, a política brasileira se baseou em dois pontos: a infra-estrutura e o arcabouço legal.

A universalização das telecomunicações foi o ponto principal sobre a infra-estrutura, pois, em médio prazo, pode-se disponibilizar o acesso à Internet para toda a população. Após ter a infra-estrutura moldada, o Brasil teve que pensar em como fazer com que as transações eletrônicas tornassem legais perante a lei. Isso foi feito reformulando os contratos, documentos e inúmeros outros itens utilizados na compra e venda convencional. Isso foi necessário para fazer com que o Brasil se incluísse realmente na sociedade da informação. Para tornar realidade, foi criado um Comitê Especial de Segurança da Internet, que assessorou tecnicamente o Governo e o Congresso a regulamentar e modificar as leis (FERRER; SANTOS, 2004).

Depois de ter definido e entrado em prática os pontos acima, foi criado o Comitê de Comércio Eletrônico, que ficou responsável em duas linhas de ação: *Bussiness to Bussiness (B2B)*³ e *Bussines to Consumer (B2C)*⁴.

O início do B2B foi tímido, apenas utilizando o uso de leilões por intermédio de banco de dados dos fornecedores que já existiam. Com o passar do tempo e com o sucesso dos resultados obtidos, a mídia especializada começou a divulgar, fazendo com que novos participantes aderissem ao B2B.

³ Comércio eletrônico (transações on-line) entre empresas.

⁴ Comércio eletrônico (transações on-line) entre empresas e consumidores finais.

Conforme Nakamura (2001), a Internet, como ferramenta de apoio ou como a principal estratégia para as empresas, pode trazer várias vantagens, pois ela melhora a comunicação e se bem planejada orienta melhor o cliente, gerando novos negócios.

Conforme Jarvis [200-?], o comércio eletrônico não é 100% seguro, e por esse motivo, muitos consumidores possuem o receio de efetuar alguma compra. No entanto, é mais seguro comprar na Internet do que andar pelas ruas e correr o risco de ser assaltado, e, por exemplo, ter o cartão de crédito roubado. Mas o problema do comércio eletrônico, de não ser 100% seguro, é um dos motivos que fazem com que as vendas de produtos on-line para consumidores finais no Brasil ainda sejam baixas. Pensa-se que no futuro essa demanda irá crescer, pois se as novas tecnologias forem bem definidas e aproveitadas, poderão fazer com que a Internet seja uma alternativa ao B2C, fazendo com que mais pessoas tenham acesso.

2.1.1 E-Commerce Seguro

As ameaças à segurança de um indivíduo é quase que uma rotina na vida real hoje em dia. Na Internet o problema acontece igualmente, a diferença é que ao invés de se ter que trancar as portas, colocar grades e alarmes, no mundo da Internet, o usuário tem que tomar algumas precauções para não ser surpreendido com algum *hacker*⁵ roubando informações suas para que possa usar, por exemplo, para efetuar compras em *sites de e-commerce*.

Conforme Felipini (2005), para proteger o usuário do mundo virtual, existem algumas ferramentas como, por exemplo:

⁵ Especialistas em violar sistemas computacionais.

- a) certificação: a certificação serve para que o usuário tenha alguma segurança no que se refere se ele está visualizando uma página correta e não uma clonada;
- b) encriptação de dados: a encriptação dos dados é a tecnologia que faz com que as informações que trafegam entre o computador do visitante e o servidor Web não sejam vistas por pessoas que não possuem acesso permitido;
- c) segurança nas transações com cartões de crédito: as seguranças com os cartões de créditos evoluíram bastante. Antigamente quando um usuário colocava o número de seu cartão em uma página de *e-commerce*, esses dados ficavam sob posse da própria loja, onde fazia com que os dados fossem mais facilmente acessados por pessoas indevidas. Nos dias atuais, o número do cartão já não fica mais com as lojas. Quando os clientes digitam seus dados, ele utiliza uma interface segura com a empresa que administra o cartão de crédito e não mais com a loja que está vendendo.

A Internet ainda não é 100% segura e não se sabe quando isso irá ocorrer.

Por esse motivo, cada empresa que utiliza a Internet como forma de negócios, devem se preocupar com a segurança da sua loja virtual e de seus clientes, tomando as medidas cabíveis.

2.2 E-GOVERNO

E-governo aplica os conceitos de comércio eletrônico, como por exemplo, informação e marketing por meio de *sites* da Internet, às operações do governo.

Com a tecnologia que temos hoje e o uso da Internet como um meio de comunicação em massa, surge a necessidade dos órgãos públicos utilizarem, para reunir em apenas um lugar, diversos serviços para os contribuintes, tentando com isso, diminuir a burocracia (BRASIL, 2001).

A Internet possui o objetivo de fornecer e tornar disponíveis informações, serviços e produtos pelo meio eletrônico, a partir de órgãos públicos a qualquer momento e local a todos os cidadãos, de modo a agregar valor a todos os envolvidos com o meio público (BRASIL, 2001).

O governo eletrônico no Brasil está cada vez mais crescendo, sendo que segundo Knight (2005), é o líder da América Latina. Isso acontece, pois como se pode verificar a Figura 2, o Brasil é o país da América do Sul que possui a maior quantidade de *hosts*. Para obter todo o sucesso esperado, o governo deve se aproximar das demandas da sociedade e permitir que a relação entre governo e o mundo empresarial troquem informações, fazendo com isso que ambos cresçam.

HOSTS NA AMÉRICA DO SUL		
Fonte: Network Wizards 2006		
	País	Julho 05
1º	Brasil (.br)	5.094.730
2º	Argentina (.ar)	1.464.719
3º	Colômbia (.co)	462.420
4º	Chile (.cl)	440.585
5º	Peru (.pe)	263.786
6º	Uruguai (.uy)	125.775
7º	Venezuela (.ve)	47.614
8º	Equador (.ec)	17.701
9º	Bolívia (.bo)	16.274
10º	Paraguai (.py)	10.840

Figura 2. *Hosts* na América do Sul
Fonte: NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO (2006)

Com o uso, espera-se que os cidadãos tenham acessos mais rápido e conveniente aos serviços governamentais, pois estes estarão abertos para o público, esteja onde o contribuinte estiver.

De acordo com Sampaio e Campos (2002), por meio do e-governo, é proporcionado uma maior transparência na administração pública, pois é demonstrado e interagido com a comunidade, estando assim, menos sujeito à corrupção, pois a população sempre que preciso for, terá informações referentes ao governo.

O e-governo brasileiro, trouxe uma consequência importante, que é o *e-commerce*. O sucesso do governo eletrônico espalha uma base de confiança nas novas tecnologias, fazendo com que novas empresas e consumidores finais reduzam o receio com essas tecnologias.

O governo eletrônico sempre estimulou o comércio eletrônico, mas ainda é necessário aprimorar e facilitar a entrada de pequenas e médias empresas neste ramo, fazendo com isso, que o mercado continue a crescer.

2.2.1 E-Governo do Brasil

De acordo com Siqueira (2004), a cada novo dia, o Brasil procura informatizar mais a administração pública.

Na década de 70, existiam apenas os números telefônicos que serviam de atendimento a população, como por exemplo, o 190. Com o passar do tempo, a administração do país foi evoluindo e, utilizando as novas tecnologias, cada vez mais foram procuradas novas soluções para poder disponibilizar informações à população referente aos acontecimentos do Brasil. Na década de 90, vários serviços já poderiam

ser feitos do mesmo lugar, como por exemplo, conseguir cópias de documentos, certidões de nascimento e de casamento, entre outros.

O Brasil já aprendeu a usar as novas tecnologias e delas, conseguiu aumentar a eficiência no atendimento à comunidade.

Para os países que estão em desenvolvimento, o e-governo ajuda a acelerar o desenvolvimento econômico, político, social e administrativo, pois contribuem para passar por cima de algumas das ineficiências hoje existentes. Por mais burocracia que exista hoje no Brasil, o governo eletrônico é uma das alternativas que se pode estar utilizando para contribuir com a sua diminuição, ou seja: a economia digital não é apenas uma opção, mas sim, uma tarefa indiscutível dos governos (SIQUEIRA, 2004).

No ano de 2000, o governo brasileiro lançou as bases para a criação de uma sociedade digital ao criar um grupo de trabalho com o propósito de examinar e propôr políticas, diretrizes e normas relacionadas com as novas formas eletrônicas de interação. As ações deste Grupo de Trabalho em Tecnologia da Informação (GTTI) juntaram-se com as metas do programa sociedade da informação, coordenada pelo ministério da ciência e tecnologia (BRASIL, 2005).

Com a orientação do governo nacional, o trabalho do GTTI juntou forças e concentrou em três linhas de ação do programa sociedade da informação (BRASIL, 2005):

- a) universalização de serviços;
- b) governo ao alcance de todos;
- c) infra-estrutura avançada.

Em meados de julho de 2000, o GTTI propôs uma nova política de interação eletrônica do governo com a sociedade, apresentando um relatório preliminar, onde continha um diagnóstico da situação da infra-estrutura dos serviços do governo federal,

das aplicações existentes até o momento, das desejadas e da situação da legislação de interação eletrônica.

O estabelecimento do comitê executivo do governo eletrônico pode ser considerado um dos grandes marcos do compromisso do conselho do governo em prol da evolução da prestação de serviços e informações ao cidadão.

O Comitê Executivo de Governo Eletrônico (CEGE) tem o objetivo de formular as políticas, estabelecer diretrizes, coordenar e articular as ações de implantações. Atendendo o plano de metas, apresentou um documento chamado de política de governo eletrônico.

Em setembro do ano de 2002, foi publicado um documento possuindo o balanço de todas as atividades desenvolvidas em dois anos de governo eletrônico, com capítulos dedicados à política de e-governo, avaliação da implantação, dos resultados obtidos, dos principais avanços, as limitações e desafios futuros do programa. Esse documento constituiu uma base de informações para poder dar continuidade ao programa no ano de 2003 (BRASIL, 2005).

No final do ano, a presidência da república publicou um decreto criando oito comitês técnicos de governo eletrônico:

- a) implementação do software livre;
- b) inclusão digital;
- c) integração de sistemas;
- d) sistemas legados e licenças de software;
- e) gestão de *sites* e serviços on-line;
- f) infra-estrutura de rede;
- g) governo para governo (G2G);
- h) gestão de conhecimentos e informação estratégica.

O ministério do planejamento, orçamento e gestão, por meio de sua secretaria, garante o apoio técnico-administrativo necessário ao funcionamento do CEGE e supervisiona os trabalhos dos comitês, interagindo quando necessário com os seus coordenadores.

2.2.2 Os Benefícios

Um dos benefícios que se tem utilizando o e-governo, são as simplificações de processos administrativos. Por uma página na Internet, ou através de um sistema on-line, é possível conseguir eliminar as formalidades que distânciam a população do governo, e assim fazer com que os contribuintes interagem com o que está acontecendo nos cofres públicos, tornando a gestão mais transparente. Tenta-se assim, evitar a corrupção e aumentar a confiança da população ao governo em vigor. Por meio de *sites*, é possível também diminuir a burocracia, pois utilizando o meio eletrônico, é possível imprimir o seu boleto bancário e efetuar o pagamento sem ter que enfrentar inúmeras filas (SAMPAIO; CAMPOS, 2002).

2.2.3 O E-Governo na Prefeitura Municipal de Criciúma

A Prefeitura Municipal de Criciúma (PMC) hospeda em seu servidor Web, uma página de Internet que pode ser visualizada através do endereço www.criciuma.sc.gov.br, onde disponibiliza aos visitantes, notícias gerais sobre a cidade, entre outros *links*⁶. O *site* ainda está em desenvolvimento, como pode ser visto na Figura 3.

⁶ São palavras ou imagens, que quando clicadas, provocam a transferência para outro assunto ou outra página da Web.



Figura 3. Página da Prefeitura Municipal de Criciúma
 Fonte: CRICIÚMA (2006)

Neste *site*, algumas seções já estão em funcionamento, como o *Blog*⁷ da Secretaria de Comunicação, conforme a Figura 4, que é atualizado diariamente resgatando para a população, notícias do dia-a-dia da cidade, a agenda do prefeito, informações referentes sobre as secretarias da prefeitura, entre outras informações gerais. A seção Licitações PMC, indicam quais são as que estão em andamento na prefeitura. Possui também um *link* do Imposto sobre a Propriedade Predial ou Territorial Urbana (IPTU), onde é possível imprimir os boletos bancários para efetuar o pagamento de 2006, além também de outros que são importantes para a população.

⁷ Serviço de informação diário oferecido pela internet para o público em geral.

Home Agenda do Prefeito Quem Somos Secretarias Fotos Ensaio Fotográfico Emigrante Links

SECOM
Secretaria de Comunicação

Uma Publicação da Secretaria Adjunta de Comunicação da Prefeitura Municipal de Criciúma

Criciúma(SC), terça-feira, 15 de novembro de 2005 11:34:28

Emigrante...

Seguem até 25 de novembro

14/11/05 15:34
Criciúma participa do Dia Nacional de Combate à Tuberculose

Pessoas com tosse por mais de quatro semanas devem procurar, imediatamente, a unidade de saúde, pois pode ser sintoma de Tuberculose. Doença contagiosa, causada pelo Bacilo de Kock com preferência pelo pulmão e torna-se grave se não for tratada.

Como se trata de enfermidade que merece cuidados especiais, a doença ganha destaque, nesta quinta-feira (17/11), no Dia Nacional de Combate à Tuberculose e na 5ª Mobilização Municipal de Tuberculose. Das 8 às 16 horas, na praça Nereu Ramos, serão distribuídos materiais informativos, além de ser realizado trabalho intensivo de nas unidades de saúde. Todas estarão equipadas com folders, cartazes e camisetas. Também os ônibus e terminais urbanos receberão cartazes informativos. Nos 30 dias seguintes, o tema será trabalhado ainda nas escolas municipais por enfermeiras e nos grupos de saúde de cada bairro.

O Programa Municipal de Combate à Tuberculose em Criciúma conta com uma enfermeira, pneumologista e auxiliar de enfermagem. Segundo a enfermeira Sandra Winkler, dos pacientes atendidos em 2005 no município, 21% é feito por meio do Dot's - Tratamento Diretamente Observado. "O Dot's é uma estratégia utilizada para aumentar a adesão ao tratamento, na qual um profissional é responsável em medicar o paciente onde ele estiver. Assim, o doente não tem como abandonar o processo", explica Sandra. O centro de atendimento aos portadores fica na rua São José - 414 - Centro, e funciona das 8 às 14h30min. O médico está à disposição, diariamente, a partir das 13 horas e não é preciso agendar consulta previamente.

Contágio

- de pessoa a pessoa do ar que se respira (tosse e espirro)
- as pessoas que têm maior risco de adoecer são: as que vivem com doente de tuberculose pulmonar em lugares fechados com pouca ventilação e aquelas em condições de alimentação, habitação e saúde precárias



Representantes do Setransc reuniram-se hoje (10/11) com o prefeito Anderlei Antonelli pedindo atenção ao Porto Seco. Eles buscam parceria para o nivelamento do terreno que integra o projeto Porto Seco - Cidade dos Transportes. O prefeito, por sua vez, vai buscar apoio junto ao Governo do Estado, que deve entrar com as máquinas, enquanto que ao Setransc cabe o operador. Segundo o presidente Miro Manique Barreto, a operação vai facilitar os serviços no local.



Antonelli recebeu hoje em seu gabinete a visita das crianças na escola maternal do bairro Demboski. (Foto: Maurício Vieira).

Figura 4. *Blog da Secretaria de Comunicação*
Fonte: CRICIÚMA (2005)

O site está em fase de desenvolvimento, e em breve, muitos outros recursos estarão implementados, como por exemplo: o contribuinte poderá fazer algumas consultas no sistema da prefeitura por meio da página (informação verbal)⁸.

Para ocorrer o funcionamento da Internet com todos os seus benefícios como os sites e outros serviços oferecidos, estes ficam hospedados em diversos servidores ao redor do mundo.

⁸ Informação fornecida por Henrique Sabino, responsável pelo setor de informática da PMC, em dezembro de 2005.

3 SERVIDORES

Um servidor é um computador que, em uma rede, fornece um ou mais recursos aos outros computadores conforme a Figura 5. Esse serviço pode ser compartilhamento de arquivos, compartilhamento da Internet, entre outros serviços. Também pode estar servindo recursos para o restante da rede, como por exemplo, o servidor Web, que disponibiliza arquivos para outros computadores clientes que se conectarem para poderem visualizar os arquivos compartilhados (JÚNIOR; HELVÉCIO, 1999).

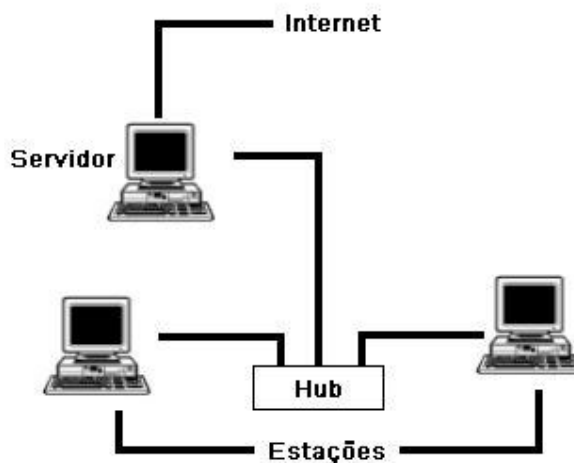


Figura 5. Servidor de Internet
Fonte: ADAPTADA DE GUIA DO HARDWARE (2006)

3.1 SERVIDORES WEB

Um Servidor Web é um computador que serve para outros, páginas da Web. Conectado, os computadores clientes podem visualizar em tempo real, as páginas de Internet hospedadas por ele.

Quando um micro cliente faz a requisição para verificar a página hospedada, o servidor verifica se o micro cliente possui privilégios para acessar o *site*, caso for aceito, a conexão é feita e os dados são transmitidos. Caso não for aceito, o computador cliente é notificado com um erro. Como a qualquer hora do dia pode ter uma requisição para visualizar uma página hospedada no servidor, estes devem permanecer ligados 24 horas.

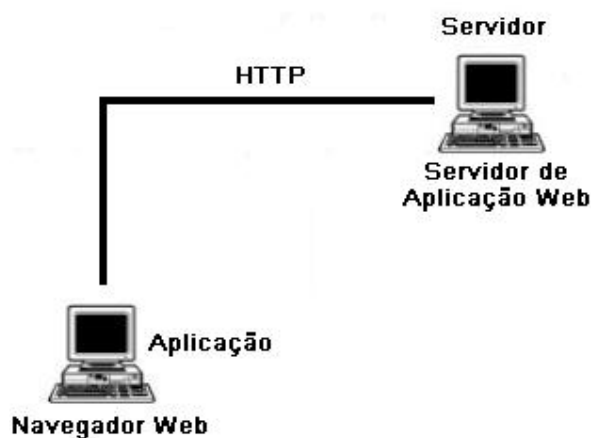


Figura 6. Servidor Web
Fonte: ADAPTADA DE GUIA DO HARDWARE (2006)

Como mostrado na Figura 6, servidores Web interagem com os outros computadores por meio do *HyperText Transfer Protocol* (HTTP) que, de acordo com o *site* A.I.S.A. (1999), é o protocolo padrão que permite que os dados na Internet sejam trafegados entre os servidores e os *browsers* dos visitantes, além também de poder trabalhar com o *HyperText Transfer Protocol Secure* (HTTPS), que é o HTTP sobre uma camada SSL, onde permite que os dados trafeguem por meio de uma conexão segura.

Além de transmitir as páginas na Internet, os servidores Web podem transferir imagens, aplicativos java⁹, disponibilizar banco de dados, além de executar programas e scripts para interagir com o usuário.

3.2 SERVIDOR WEB APACHE

De acordo com Hunt (2000), a *World Wide Web* (WWW) tornou-se sinônimo de Internet. Todo assunto que trata de Internet, é comentado sobre Servidor Web. Estes são serviços utilizados para publicações de produtos, ofertas de serviços, além de poder também disseminar informações para fora das empresas. Utilizando outras palavras, servidores Web são usados para distribuir informações.

Conforme Wadlow (2000), os sistemas Linux geram excelentes servidores Web. Por padrão, a maioria das distribuições Linux, como por exemplo, Redhat, Slackware e Mandriva, trazem consigo a instalação do Servidor Web Apache. Conforme Leite (2005), este é um dos servidores mais utilizado no mundo. O Apache foi criado quando um grupo de programadores foi corrigir alguns problemas existentes e incluir novas funcionalidades ao antigo NCSA HTTPD, um dos primeiros servidores Web.

Um dos pontos positivos do Apache, perante outros servidores Web, é que quando é iniciado, carrega uma quantidade de processos no servidor onde compartilham a carga de trabalho e, caso todos os processos ficarem ocupados, novos processos são iniciados para dividirem os trabalhos, aumentando assim o seu desempenho.

O Apache está disponível na Internet com o seu código fonte e é desenvolvido coletivamente, e sua configuração utilizada encontra-se no Apêndice A.

⁹ Linguagem de programação.

Surgiu como um servidor não comercial para UNIX, mas foi também traduzido para ser utilizado em servidores com o sistema operacional Windows. A Figura 7 mostra a página inicial, recém instalada de um Servidor Web Apache.

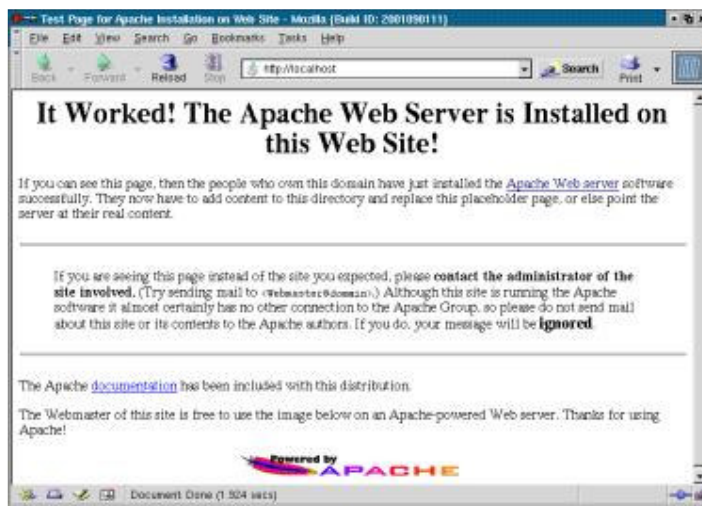


Figura 7. Página Inicial do Servidor Web Apache
Fonte: THE APACHE SOFTWARE FOUNDATION (2005)

Outros, mas não menos importante, são as vastas configurações que se podem fazer neste servidor, como por exemplo:

- a) o número máximo de clientes que estarão conectados no servidor Web;
- b) o tempo em que o servidor ficará esperando por uma nova conexão de um computador visitante para garantir que este não possui mais nenhum pedido adicional;
- c) opções de *logging*;
- d) ter a possibilidade de fazer com que o servidor disponibilize diversos *sites* utilizando o mesmo IP¹⁰, entre outros.

O Apache possui em seu arquivo de configuração, a opção de onde ficará e quais serão os arquivos de *log*. Estes identificam quais foram às falhas do servidor, as atividades de transferência dos arquivos entre outras informações.

¹⁰ Internet Protocol é um dos mais importantes protocolos em que a Internet é baseada.

A maior tarefa não é fazer com que o servidor Web Apache funcione, pois a maioria das opções necessárias para o servidor entrar em funcionamento, vem por padrão. A maior tarefa é fazer com que ele tenha um maior desempenho e assegurar que tudo o que está acontecendo no servidor, seja registrado nos *logs*. Para isso ocorrer, é necessário configurar algumas opções no arquivo de configuração do Apache. Monitorar o uso, o desempenho e assegurar a segurança do servidor são tarefas diárias.

Esses arquivos de *log* oferecem muitas informações importantes do servidor Web. Basicamente, as falhas do sistema são registradas em um arquivo chamado *errorlog* e os acessos com as informações dos clientes ficam em *aceslog*. Esses fornecem informações vitais para várias funções, incluindo a depuração de scripts *Common Gateway Interface*¹¹ (CGI) defeituosos e a produção de relatórios de acesso ao *site*. O *log* deve ser visto ao menos uma vez ao dia, para verificar os problemas e conseqüentemente resolvê-los.

Tão importante como os erros, os *logs* oferecem informações sobre quem conectou ao servidor, quando que foi esse acesso, o quanto e o quê o visitante utilizou e, de que maneira ele esteve servindo os computadores clientes.

No Apache, os *logs* são customizáveis isto é: é possível definir quais são as opções que se deseja registrar, como por exemplo: o endereço IP do cliente, a data e hora que ocorreu o acesso, o que foi solicitado, entre outros.

Servidores Web são vulneráveis a todos os problemas normais de segurança, mas possuem também suas próprias considerações especiais. Ele é responsável por proteger a integridade das informações que está disponibilizando, ou seja, um dos riscos de segurança que um servidor Web pode possuir, é um invasor ter conseguido mudar as informações de suas páginas hospedadas.

¹¹ Protocolo básico para a integração entre um servidor HTTP e um programa gerador de páginas dinâmicas, por exemplo, um *script* CGI liga um Servidor Web em uma página em linguagem PHP.

3.3 MONITORAMENTO

Na perspectiva de Júnior e Helvécio (1999), o monitoramento de sistemas é necessário para planejar o que será preciso no futuro em ambientes computacionais. A tarefa de monitoramento é mais fácil quando é efetuada em apenas um servidor. Monitorando um sistema, pode-se identificar, por exemplo, se existe algum problema de desempenho na rede. Monitoramento é a leitura junto com a observação e análise do estado e do comportamento dos serviços e do sistema.

O monitoramento define quais serão as informações observadas e qual a maneira de se obter essa informação. Também é possível definir a melhor estratégia para obter informações dos recursos gerenciados e como as informações recolhidas do monitoramento pode ser utilizada para permitir a análise e o diagnóstico dos problemas nas várias áreas funcionais.

As informações que são avaliadas no monitoramento da rede são classificadas em três tipos: estáticas, dinâmicas e estatísticas.

As informações estáticas são as configurações atuais da rede e seus componentes, mudam com pouca frequência. As dinâmicas são os eventos das redes, como por exemplo, a transmissão de um pacote pela rede. Já as estatísticas, são derivadas das informações dinâmicas. Indicam por exemplo, a quantidade de pacotes que são enviadas em média por hora em uma rede.

3.4 ARQUIVOS DE LOGS

Em um arquivo de *log* são registrados as atividades e os erros de um sistema. Este possui informações de quando ocorreu o aviso, quem está avisando, qual é o erro

que está sendo avisado, entre outras informações. Esses arquivos ajudam para prevenir ou evitar problemas futuros.

Os servidores Web geralmente criam seus arquivos de *logs* em dois formatos: *Common Log Format* (CLF) e *Extended Common Log Format* (ECLF). De acordo com Hallam-Baker e Behlendorf [199-?], o CLF mantém as requisições que ocorreram em uma página da Web onde contém os campos conforme a Tabela 1.

Tabela 1. Campos de Entrada de um *Log* CLF

Campo de Entrada	Descrição do Campo de Entrada
<i>Remote host</i>	Nome de <i>host</i> ou endereço IP do cliente solicitante.
<i>Remote identity</i>	Conta associada à conexão na máquina do cliente.
<i>Authenticated user</i>	Nome fornecido pelo usuário para autenticação.
<i>Time</i>	Data/ Hora associada com o pedido.
<i>Request</i>	Método de pedido, URI ¹² do pedido e versão do protocolo.
<i>Response Code</i>	Código de resposta HTTP. Esse código possui números com três dígitos, conforme a Tabela 3.
<i>Content length</i>	Número de bytes associados com a resposta.

Fonte: REXFORD, J.; KRISHNAMURTHY, B. (2001)

O ECLF é uma variante do CLF, onde adiciona mais dois campos aos já existentes, conforme a Tabela 2.

Tabela 2. Campos Adicionais do ECLF

Campo de Entrada	Descrição do Campo de Entrada
<i>Referer</i>	O endereço URL que o visitante estava antes de requisitar o <i>site</i> hospedado no servidor Web.
<i>User_agent</i>	Software que o cliente está utilizando.

Fonte: HALLAM-BAKER, P; BEHLENDORF, B. [199-?]

¹² Qualquer recurso disponível na Internet como um documento HTML, imagem, videoclipe ou programa.

Tabela 3. Códigos de Estado Retornados pelo Servidor

Código	Descrição do Código
200	OK, a solicitação é válida.
302	<i>Found 0</i> , o documento requisitado não foi encontrado.
304	<i>Not modified</i> , o documento requisitado não foi modificado.
400	<i>Bad request</i> , a solicitação é inválida.
401	<i>Unauthorized</i> , o cliente ou usuário tem acesso negado.
403	<i>Forbidden</i> , o acesso à solicitação não é permitido.
404	<i>Not found</i> , o documento requisitado não existe.
500	<i>Server error</i> , o servidor ocorreu um erro não especificado.
501	<i>Not implemented</i> , A ferramenta solicitada do servidor não está disponível.
502	<i>Bad gateway</i> ¹³ , o cliente especificou um <i>gateway</i> inválido.
503	<i>Out of resources</i> , o servidor não possui recursos suficientes para responder as solicitações.

Fonte: HUNT. (2000)

Os *logs* comuns estendidos mostram mais informações, porém, necessitam de mais processamento para serem analisados (FRIEDLEIN, 2003).

Quando se utilizam *logs* com esses dois formatos, significa que o *log* gerado por esse servidor, pode ser processado por qualquer ferramenta de análise de *logs* que obedeçam ao padrão.

De acordo com Friedlein (2003), os *browsers*¹⁴ e os servidores Web comunicam utilizando o protocolo HTTP. Quando um navegador Web envia uma solicitação ao servidor, ele inclui um cabeçalho que contém algumas informações que são registradas pelo servidor. A capacidade de analisar a atividade do *site* é confinada às informações que são registradas dessa forma.

¹³ Máquina intermediária, geralmente destinada a interligar redes, separar domínios de colisão, ou mesmo traduzir colisões.

¹⁴ Programa que habilita os usuários a interagirem com documentos HTML hospedados em um servidor Web.

O servidor Apache possui os seus arquivos de *logs* geralmente situados em */var/log/apache*, mais pode ser definido em outro diretório, mudando a configuração referente no arquivo de configuração.

O principal arquivo de *log* do servidor Web Apache é o arquivo *AccessLog*, pois é nele que ficam as informações dos visitantes sobre os acessos às páginas. Por padrão, dois arquivos são criados para serem salvas as informações sobre o servidor: o *accesslog* que como comentado anteriormente é o arquivo onde serão registradas todas as informações sobre o acesso dos visitantes ao servidor Web como data e hora, o que foi solicitado, entre outros e o *errorlog*, que registra os erros de acesso nas páginas e os erros do próprio servidor.

A Tabela 4 mostra as opções em que é possível configurar no Apache para que ele fique monitorando.

Tabela 4. Opções de configuração do servidor Web Apache

Opção de Configuração	Descrição da Opção
AgentLog	Registra o navegador que está acessando o servidor.
ErrorLog	Registra os erros do <i>site</i> e do servidor Web.
RefererLog	Registra os campos <i>referer</i> do cabeçalho HTTP, ou seja: a URL ¹⁵ que o cliente estava visualizando antes de fazer o pedido para o servidor Web.
RewriteLog	Registra as regravações de URL que foram feitas pelo Servidor.
ScriptLog	Registra os erros gerados pelos scripts CGI que foram executados no servidor.
AccessLog	Registram todas as transferências dos arquivos entre o servidor e os micros clientes.

Fonte: Apache HTTP Server. [200-?]

¹⁵ Endereço de um recurso disponível na Internet.

O Apache instalado no servidor Web da Prefeitura utiliza o padrão ECLF. A Figura 8 tem-se uma linha do arquivo de *log* pelo formato CLF, e a seguir se tem a explicação dos itens:

- a) 200.215.36.17: endereço IP do computador que conectou no servidor;
- b) -: o primeiro hífen indica que parte da informação pedida não está disponível. Essa informação é a identidade do cliente RFC¹⁶ 1413¹⁷. Essa informação só é usada em redes internas altamente controladas;
- c) -: o segundo hífen é o *userid* do usuário, o valor é fornecido pelos certificados CGI na variável de ambiente do usuário remoto. Quando o usuário ainda não foi autenticado ou o servidor não registra essa informação, esta terá a entrada como “-”;
- d) [24/May/2006:23:06:32 +0000]: mostra a data e hora da requisição.
- e) “GET /secom/Imagens/fotolegenda/santosdumont.jpg HTTP/1.0”: primeiro o cliente pediu o recurso GET (esse recurso é usado quando um *browser* faz um pedido a uma página de Internet normal do servidor). Em seguida o programa cliente pediu o recurso /secom/Imagens/fotolegenda/santosdumont.jpg e em terceiro o cliente usou o protocolo HTTP em sua versão 1.0;
- f) 304: esse é o código de status registrado. Como comentado anteriormente, a Tabela 3 define qual são as condições que podem ser encontradas;

¹⁶ Documento que descreve os padrões de cada protocolo da Internet previamente a serem considerados um padrão.

¹⁷ Protocolo de identificação. Fornece meio para determinar a identidade de um usuário por meio de uma conexão TCP/IP particular.

- g) -: essa última entrada significa o tamanho do pedido retornado ao cliente. Quando essa entrada está como “-”, significa que nenhum índice foi retornado ao cliente;
- h) “Mozilla/3.01 (*compatible*;)” : Nome e versão do *browser* que está fazendo o pedido ao servidor Web.

```
200.215.36.17 - - [24/May/2006:23:06:32 +0000]
"GET /secom/Imagens/fotolegenda/santosdumont.jpg HTTP/1.0"
304 - "Mozilla/3.01 (compatible;)"
```

Figura 8. Arquivo AccessLog
Fonte: THE APACHE SOFTWARE FOUNDATION (2006)

Os arquivos de *logs* foram projetados no princípio para permitir que os administradores do servidor tivessem a possibilidade de medir o desempenho. Por esse motivo, há muita informação que pode ser considerada redundante, como por exemplo, as solicitações de arquivos. Antes de analisar os *logs* e coloca-los em um banco de dados para ter as informações mais importantes, é necessário usar um software para filtrar grande parte das informações.

3.5 A IMPORTÂNCIA DO MONITORAMENTO E ANÁLISE DOS LOGS

Conforme Friedlein (2003), a medição e a análise do tráfego da Web, desempenham um papel importante no projeto de *sites* da Web. Os arquivos de *log* ferramentas poderosa para ajudar a entender a atividade do usuário em uma página.

O valor real da análise destes arquivos vem com o tempo. O mais útil é a capacidade de ver as tendências e, portanto, as anomalias. A análise não é certamente 100% precisa, mas, pelo menos, os níveis de falta de precisão são coerentes, de modo que as observações com o tempo são validas entre si.

A quantidade de dados do arquivo em uma semana dá um conhecimento instantâneo do volume da atividade em um *site*, mais é necessário por volta de no mínimo três meses para se obter as reais tendências. Quando essas forem entendidas, os picos e anomalias das páginas estarão em evidência e, geralmente, a sua causa poderá ser rastreada e avaliada (FRIEDLEIN, 2003).

Os arquivos de *logs* não informam o porquê das coisas que estão acontecendo, e sim o que ocorreu. Caso um *site* seja analisado antes e depois de algumas atualizações, pode-se ter uma boa idéia do sucesso ou não dessas mudanças.

Também, em vários casos, é preciso analisar e relatar o desempenho de um *site* da Web. Isso se torna necessário, pois em um *site* podem existir diversos sistemas, arquivos e aplicações que aos poucos vão crescendo para atender à demanda de diferentes partes da empresa. Neste caso, é possível relatar sobre certas coisas específicas, mas é muito difícil conseguir dar uma boa idéia de como o *site* inteiro está contribuindo para a empresa.

Para poder informar medições coerentes, é preciso inserir ordem e coerência na estrutura de medição. Podem-se também inserir diretrizes de projeto de navegação, além de outras medidas de padronização, que facilitam a análise.

Para facilitar a criação do relatório, quando se irá montar uma página para uma determinada empresa, é necessário visar o *site* para futuras necessidades de análise, ou seja: deve-se tentar garantir que o *site* fique estruturado, para que quando for preciso montar novas seções, cada elemento contenha informações que o relacionem à finalidade básica da página. Desse modo, pode-se medir cada elemento individual de um modo que reflita realmente como está contribuindo para o valor geral.

Para poder analisar as informações que são visualizadas no arquivo de *log* do Servidor Web, deve-se analisar tendo como comparação uma métrica.

4 MÉTRICAS ELETRÔNICAS

As métricas são as unidades fundamentais que podem ser usadas para medir o sucesso de algo com determinados objetivos (FRIEDLEIN, 2003).

Para definir os padrões, Friedlein (2003), comenta que algumas agências lideram a definição de padrões, protocolos e métricas eletrônicas na Internet, entre eles:

- a) *internet advertising bureau*: associação de propaganda interativa, líder no setor;
- b) *international organization for standardization*: organização responsável em promover o desenvolvimento da padronização e as atividades relacionadas no mundo;
- c) *world wide web consortium*: responsável em desenvolver novas tecnologias para fazer com que a Web chegue ao seu potencial pleno, utilizando fóruns de informações, comércio, comunicação e conhecimento;
- d) *audit bureau of circulation electronic*: fonte de dados de auditoria independente na medição de tráfego da Internet nos seguintes países: Reino Unido e Irlanda. Utiliza padrões internacionais.

Ainda de acordo com Friedlein (2003), com o passar do tempo, a Internet tem crescido, e com isso cria novos modelos e idéias de negócios. Por esse motivo, novas métricas foram procuradas para tentar medir o *e-bussiness*¹⁸. As novas métricas eletrônicas como as visitas, aderências, usuários exclusivos, estavam com pouca clareza

¹⁸ Acrônimo do inglês *Electronic Business*. É o termo que se utiliza para identificar os negócios efetuados por meio eletrônico, geralmente na Internet.

em relação ao seu significado, pois as visitas eram usadas como sinônimos de visões de páginas e as seções eram confundidos com usuários exclusivos.

Com esses problemas, as métricas eletrônicas foram transformadas e baseadas em receita e lucratividade, como por exemplo, o tamanho médio da cesta de compras de usuário cadastrado no *site*. Mesmo com essas mudanças, a métrica eletrônica continua a evoluir na busca por uma inteligência comercial para ajudar a tomar melhores decisões contra a concorrência.

A evolução da métrica eletrônica está ocorrendo porque há um impulso crescente no setor de definir um acordo quanto aos padrões de métrica eletrônica, fazer também que os setores específicos utilizem métricas exclusivas ao que eles fazem integrar essas métricas com o restante da empresa e utilizar as métricas também para medir o sucesso da empresa (FRIDLEIN, 2003).

Muitas empresas possuem pelo menos algumas estatísticas baseadas no tráfego de seu *site* utilizando ferramentas de análise de *logs* gratuitas. Geralmente essas estatísticas não possuem muito valor para os gerentes, pois não possuem a previsão de receitas ou acréscimo a lucratividade, mesmo assim, esses relatórios oferecem mais valor do que é percebido, como por exemplo, quem é o cliente e quais são as preferências dele.

4.1 MEDIÇÕES CENTRADAS NO *SITE*

A medição centrada no *site* é uma categoria da métrica eletrônica, onde os usuários são constantemente referenciados. Não se consegue ter conhecimento individualizado, pois os relatórios só existem como pessoas não identificadas, que interagiram com o *site*.

Este tipo de medição responde a quatro perguntas básicas, sobre um *site* Web e como ele está sendo usado:

- a) o que o usuário está fazendo no site?
- b) com o que ele está visualizando o site?
- c) de onde ele veio para chegar ao site?
- d) como o *site* está funcionando tecnicamente?

Para conseguir obter as respostas a essas perguntas, é necessário verificar os seguintes itens:

- a) *hits*: hit é uma solicitação de um arquivo no servidor. Não é considerada válida, pois se você visualizar uma página que possui cinco figuras, você receberá seis *hits*, um pela página HTML estática e cinco pelas figuras;
- b) exibições de páginas e as páginas mais acessadas: isso é uma solicitação ao servidor para uma página HTML. Muito útil para se medir a quantidade de visitas do site por horário ou por data. Para ser preciso, o tráfego não humano, como, por exemplo, os *spiders*¹⁹ das páginas de buscas, precisam ser filtrados;
- c) principais páginas de entrada e saída: estas são as páginas com que a maioria dos usuários entra e saem do site. É interessante para saber se existe algum *link* que está fazendo com que os usuários saiam da página ou se os usuários possuem um *link* da página em seu *bookmark*²⁰, que não seja a página principal do site;
- d) visitantes exclusivos: usuários distinguidos pelo endereço de IP. Não é confiável, pois os endereços de IP dinâmico que é usado pelos provedores de Internet podem dar um número de usuários

¹⁹ Programas usados em *sites* de buscas para visitar páginas da Web e coletar informações. Também são conhecidos como *bots*.

²⁰ Recurso dos programas de navegação na Web que permite o armazenamento dos endereços favoritos.

enganosamente alto e os servidores *proxy*²¹ podem fazer com que os números de usuários sejam menores que o valor real. Também, se o computador possuir um endereço IP fixo e várias pessoas o utilizam à mesma máquina, irá parecer que é o mesmo usuário que está acessando o *site*;

- e) *seções*: uma seção é o tempo gasto em uma página sem interrupções;
- f) *número médio de impressões de página por visita*: essa é uma medida baseada na divisão do número total de impressões da página pelo número total de seções em um determinado período;
- g) *abandono*: é a medida de quantas pessoas não completam o que você queria que elas fizessem, como por exemplo, o abandono do carrinho de compras;
- h) *conversão*: é a medida do número de pessoas que estão fazendo algo que você deseja acompanhar;
- i) *download*²² de arquivos: o número de vezes que um determinado arquivo foi baixado. Pode-se com isso, verificar quais são os arquivos mais populares de um *Web site*;
- j) *browser*: detalha qual é a versão e qual é o programa de navegação que o usuário está utilizando;
- k) *sistema operacional*: mostra qual está instalado no computador do usuário cliente;
- l) *suporte para Java/ CGI/ Cookies*²³: detalha quais são as tecnologias que o usuário oferece suporte;

²¹ Software responsável em fazer *cache* em redes de computadores.

²² Ato de um usuário transferir para o computador o que está sendo disponibilizado em outro computador.

²³ Permite que servidores gravem informações de seu interesse no *browser* do cliente.

- m) velocidade de conexão: define qual é a velocidade que o usuário está utilizando a Internet. Para descobrir esse ponto, é necessário fazer pesquisas de mercado. É um ponto importante, pois dependendo de qual a velocidade os usuários utilizam para freqüentar um *site*, este pode possuir mais efeitos visuais que não irão afetar a navegação;
- n) IPs de referência: detalhes de onde vieram os usuários para chegar ao site da empresa;
- o) país de acesso: detalha o país de onde o usuário originou a requisição;
- p) erros: detalha os erros que ocorreram no servidor e no cliente, como por exemplo: Erro 500: *Internal Server Error* ou erro 404: *File Not Found*;
- q) tempo de resposta: é a velocidade em que a página é carregada e os dados transferidos;
- r) tempo de funcionamento do servidor: Tempo em que o site ficou disponível para acesso.

Algumas visitas à página não são de fato geradas por humanos, mas por *spiders* de mecanismos de busca e outros processos automatizados. Segundo Friedlein (2003), não é raro que o tráfego não-humano seja responsável por 10 a 20% dos acessos ao site.

Assim como os *spiders* são enviados por motores de busca para indexar *sites*, existem outros tipos de *bots*, que são programas de computadores, que fazem funções específicas: copiam páginas inteiras da Web para a navegação off-line, recolhem endereços de e-mails ou verificam se há novos *links* em um *site*.

Infelizmente, todos estes possuem diferentes comportamentos em termos de caminho e número de páginas visitadas. Alguns *spiders* são enviados por motores de busca conhecidos e, com isso, podem ser reconhecidos pelo endereço IP e

posteriormente filtrados. Contudo, muitos programas são executados por indivíduos desconhecidos ou preparados para não serem localizados.

Para filtrar o máximo de atividade não humana possível, existe várias listas de *bots* on-line, que podem referenciar no servidor Web para filtrá-los.

4.2 MEDIÇÃO CENTRADA NO USUÁRIO

Conforme Fridlein (2003), a medição centrada no site é muito boa para dizer o que está acontecendo no site, mais não diz o porquê isso está acontecendo, e também não diz quem são os usuários que estão acessando o *site*. Com o tempo, tem-se a necessidade de criar relacionamentos fortes e lucrativos com os clientes para poder obter sucesso da empresa. Com isso, a medição centrada no *site* não é suficiente, pois não informa sobre os gostos, desgostos, motivações e necessidade dos clientes.

As medidas utilizadas para medir o site da Web centrada no usuário são mais instáveis do que a medição centrada no *site*. Isso ocorre porque, em geral, elas são medidas mais flexíveis, como a satisfação ou a lealdade do cliente. A medida centrada no usuário não possui nada de muito novo, o modo com que cada uma das informações são implementadas é que são novos. Isso é utilizado, pois esses elementos estão se tornando mais importantes na busca da vantagem competitiva.

O atendimento ao cliente também é extremamente importante. Na Web não existe a parte física, por isso, tudo se refere à marca e ao serviço que se é oferecido, de modo que há grandes oportunidades para criar ou destruir esse valor. A natureza remota e interativa da Web se torna muito importante tentar ganhar confiança dos seus clientes, pois você não pode vê-los, não pode falar facilmente com eles de forma direta e não

pode evitar que expressem rapidamente seus sentimentos a milhares de outros clientes. Uma vez perdidos, têm-se poucas chances de tê-los de volta.

Conforme Fridlein (2003), a medição centrada no usuário é um elemento vital de qualquer estratégia de medição de *site* e, ajuda a avaliar se você está atingindo os clientes certos e se está criando valores com sucesso tanto para eles como para a empresa. As medições do *site* Web centradas no usuário são as mesmas utilizadas pelo marketing tradicional. De modo geral, as técnicas de chegar às medidas centradas no usuário geralmente são um pouco diferentes dos métodos utilizados tradicionalmente. Uma das únicas maneiras de saber quem são os usuários que estão acessando o site ao certo é quando os mesmo informam ou quando se utiliza uma prática que não é aprovada: a compra das informações dos usuários. Mesmo quando tiverem recebido informações pessoais de quem são os usuários no site, ou pelo menos quem eles dizem que são ainda é necessário identificá-los em visitas repetidas.

Para isso, existem os *cookies* e a seção de *login*²⁴ na página. Os *cookies* que são pequenos arquivos de texto que ficam nas máquinas dos usuários e armazenam informações sobre ele e que, quando é retornado ao site que o implantou, reenvia às informações ao servidor. Os *cookies* ajudam a reconhecer o usuário que está retornando ao *site*. Esses são específicos para cada máquina, sendo assim, se o computador for compartilhado por mais de uma pessoa, os resultados não serão exatos. Podem também ser excluídos ou bloqueados pelo próprio usuário. Alguns deles só ficam por determinado tempo, como por exemplo: os *cookies* de sessão que só armazenam informações enquanto o usuário está no *site*.

²⁴ Ao efetuar o *logon*, o usuário estabelece uma conexão com um sistema on-line.

Já quando os usuários efetuam o *login*, ou seja, informam o seu nome de usuário e a sua senha pessoal no *site*, às informações pessoais sobre os usuários são mais precisas.

4.3 TECNICAS DE MEDIÇÃO

Conforme Krishnamurthy e Rexford (2001), as medições da Web podem ser coletadas de várias maneiras, dependendo exclusivamente do tipo de informação que está sendo requerida.

4.3.1 *Logging* no Servidor

O servidor Web gera um *log* como parte do processamento das solicitações. Cada entrada no *log* referência um pedido HTTP tratado pelo servidor, incluindo algumas informações sobre quem está solicitando. As empresas que desenvolvem esses servidores seguem normas informais para os formatos dos *logs*. Com isso, facilitou o desenvolvimento de diversas ferramentas para a análise dos mesmos. Essas ferramentas fazem com que os dados retirados se tornem números estatísticos, fazendo com que o servidor ofereça algumas informações valiosas para o administrador do servidor Web.

Os *logs* do servidor Web, têm formado base para a maioria dos estudos de pesquisa que tem caracterizado o tráfego HTTP ou avaliadas novas técnicas para melhorar o desempenho da Web (KRISHNAMURTHY; REXFORD, 2001).

As maiorias dos servidores Web fazem o *logging* por padrão, sem as devidas configurações, esses arquivos não oferecem informações muito detalhadas, como por

exemplo: registram apenas o pedido, a URI do pedido e o código da resposta, não informando, por exemplo, de onde a requisição veio.

Em outras palavras, um *log* de um servidor WEB, poderia ser usado para analisar os padrões de acesso dos usuários e a popularidade relativa dos recursos de um *site* da Web.

4.3.2 Logging no Proxy

Conforme Krishnamurthy e Rexford, os *proxys* da Web, assim como os servidores, geram arquivos de *logs* como parte de sua operação normal. Ao contrário dos *logs* de um servidor Web, os de um servidor *proxy* registram o acesso aos diversos *sites* que uma organização ou um provedor visitou.

Esses arquivos oferecem informações úteis sobre o padrão dos acessos dos usuários da empresa.

O *log* de um *proxy* inclui os pedidos que são executados pelo *cache*²⁵ do *proxy*. Assim, eles podem ser usados para determinar a popularidade relativa de diferentes páginas da Web.

Esses arquivos possuem várias das mesmas desvantagens dos produzidos pelo servidor Web. Ele não consegue ver os pedidos que são efetuados diretamente pelo *cache*, não conseguindo ver todos os pedidos para todos os servidores. Isso dificulta a caracterização de certas propriedades da Web, como a taxa de pedidos para os sites e os recursos mais populares. Além disso, os clientes que utilizam o *proxy*, não representam as idéias do restante dos usuários na Web (REXFORD; KRISHNAMURTHY, 2001).

²⁵ Refere-se a uma porção de memória, entretanto os *browsers* e *proxys* possuem um *cache* próprio onde arquivam as últimas imagens e páginas HTML visitadas.

Finalmente, assim como os *logs* dos servidores Web, muitas instituições comerciais não colocam os *logs* do seu *proxy* à disposição.

4.3.3 Logging no Cliente

A coleta de *log* no computador do usuário possui a facilidade de oferecer uma visão detalhada de quais *sites* o usuário navegou. O *browser* do computador poderia até registrar as horas em que foi utilizado a Internet. Poderia também registrar os pedidos que recebem resposta pelo *cache* do *browser*, além de operações do mouse e do teclado no próprio navegador.

Comparado com os *proxys* e servidores Web, o usuário trata com muito menos pedido de cada vez, e mesmo armazenando tudo que o usuário faz isso não faz com que o *browser* se torne lento.

Os *logs* dos *browsers* do cliente, não obedecem a um padrão para o formato do *log*. Os *browsers* populares não geram *logs* por padrão, com isso, para coletar os esses *logs* do cliente, é necessário que se modifique do software de navegação, como, por exemplo, adicionando algum *plugin*²⁶ ou instalando programas a parte. Outro meio para fazer com que sejam registrados os sites visitados, é instalando um *proxy* rodando na própria máquina do usuário (REXFORD; KRISHNAMURTHY, 2001).

²⁶ Programas como se fossem acessórios para os *browsers* que desempenham uma tarefa específica, como som, vídeo, entre outros.

4.3.4 Monitoração de Pacotes

Conforme Rexford e Krishnamurthy (2001), capturar os *logs* do cliente, do *proxy* ou do servidor Web, podem gerar um *overhead*²⁷ no computador, fazendo com que o mesmo fique mais lento. Além disso, os *logs* que são coletados no nível de aplicação possuem pouca ou nenhuma informação sobre a atividade da rede nos níveis do TCP e do IP. A monitoração do tráfego dentro da rede oferece um modo de coletar os dados detalhados sem afetar o desempenho dos componentes do servidor. No entanto, a coleta dos dados da rede exige que se utilize um modo de capturar os pacotes IP individuais enquanto trafegam por um *link* qualquer da rede ou por um roteador.

Um monitor de pacotes pode produzir informações detalhadas na atividade Web nos níveis HTTP, TCP e IP. Esse monitor é capaz de registrar informações referentes à hora detalhada dos pacotes. Isso é muito importante, pois é possível, analisar as propriedades de atraso, *throughput*²⁸ e a perda de transferências para a Web.

As informações, também incluem detalhes em nível de transporte, por exemplo, quando e como as conexões TCP terminam.

Assim como em um *proxy*, um monitor de pacotes pode não capturar os pedidos que são respondidos pelo *cache* do *browser* ou as mensagens HTTP que foram criptografadas usando *Secure Socket Layer* (SSL). Além disso, quanto maior a velocidade do *link* de Internet, maior se torna a dificuldade de monitorar os pacotes. À medida que os pacotes chegam, é necessário capturar os pacotes, realizar os processamentos necessários e armazenar os dados. Para gerar os dados em nível de

²⁷ Um custo adicional em processamento ou armazenamento, que como consequência, piora o desempenho de um programa ou dispositivo de processamento. Usado normalmente para se referir os custos adicionais indesejáveis ou que poderiam ter evitado.

²⁸ Em computadores é uma medida de quantidade de trabalho que pode ser processado em um período de tempo definido, em redes é uma medida da quantidade de dados que podem serem transferidos com êxito em um período de tempo definido.

HTTP, o software precisa ser capaz de associar os pacotes recebidos com a transferência apropriada.

Em comparação com os outros, a monitoração de pacotes introduz problemas adicionais sobre as violações da privacidade do usuário. O monitor de pacotes possui acesso irrestrito ao *link*, obtendo todas as informações que passam por meio dele.

4.3.5 Medição Ativa

Os *logs* dos usuários, do *proxy*, do servidor Web e da monitoração de pacotes, geralmente não incluem informações suficientes para a avaliação do desempenho dos usuários da Web, tendo que caracterizar um problema de lentidão no *download* de uma página da Web, incluindo algumas imagens embutidas. Se uma ou mais dessas imagens embutidas estivessem localizadas em um servidor diferente, o *log* do servidor da página não irá oferecer as informações sobre todos os pedidos. Um *log* de um *proxy* ou de um servidor, não captura o tempo de resposta com a consulta ao DNS²⁹ para traduzir o endereço da página. Um *log* do *browser* do usuário se estivesse disponível, capturaria o tempo de um único usuário conectando com um determinado conjunto de páginas da Web, e o tempo que esse cliente irá levar, pode não representar o de outros usuários (REXFORD; KRISHNAMURTHY, 2001).

O uso de *log* dos *browsers* dos clientes, *logs* de *proxy*, *logs* do servidor ou monitoração de pacotes, usado para estudar o desempenho do usuário, possui suas limitações. Por esse motivo, uma técnica alternativa é gerar pedidos de uma forma controlada e observar o seu desempenho. Isso é conhecido como medição ativa. A

²⁹ Serviço onde estão armazenadas as ligações entre endereços os endereços IPs e os domínios da Web.

medição ativa da Web envia um agente do usuário para enviar os pedidos e registrar as informações sobre as respostas, como hora e cabeçalhos HTTP.

4.4 ANTES DA ANÁLISE PROPRIAMENTE DITA

Conforme Rexford (2001), antes de uma ferramenta iniciar a analisar dos *logs* do servidor Web, essa efetua um pré-processamento, que envolve três etapas:

- a) analisar os dados para verificar se não há registros errados: esse é o primeiro passo. Os dados da medição consistem em vários registros com vários campos. Essa análise envolve a identificação dos limites entre os registros do arquivo de *log*, e os dados que não estiver com a sintaxe esperada, descartar;
- b) remover os campos desnecessários: nos campos em que não contiver nenhuma informação que é útil para a análise, como por exemplo, os itens destacados na Figura 9, também descartar;
- c) transformar os dados para um formato que esteja hábil para analisar: nesse último estágio de pré-processamento, é transformado o dado para uma sintaxe mais simples, mais adequado ao processamento, como por exemplo, os campos que não diferenciam maiúsculas de minúsculas, poderiam ser convertidos para um único formato.

```
200.215.36.17 [24/May/2006:23:06:32 +0000]  
"GET /secom/Imagens/fotolegenda/santosdumont.jpg HTTP/1.0"  
304 "Mozilla/3.01 (compatible;)"
```

Figura 9. Arquivo AccessLog
Fonte: THE APACHE SOFTWARE FOUNDATION (2006)

Assim, com a abordagem das métricas eletrônicas finalizou-se o estudo dos aspectos teóricos envolvidos nesta pesquisa. A seguir, descreve-se a parte prática da pesquisa realizada que consiste na análise dos *logs* no servidor Web da Prefeitura Municipal de Criciúma.

5 ANÁLISE NOS ARQUIVOS DE LOG DA PMC

A PMC possui toda a rede de computadores estruturada, sendo dividida em quatro *racks*³⁰ espalhados pela PMC, possuindo mais de 150 computadores e 8 servidores utilizando Windows e Linux. Esses servidores estão disponibilizando diversos serviços para outros computadores, entre eles estão os seguintes servidores: servidor Web, servidor *proxy*, servidor de DNS, servidor dos sistemas betha, servidor de arquivos, entre outros. A Prefeitura possui 254 IP's reais, onde alguns estão sendo utilizados em seus servidores para serem acessados pelo meio da Internet e um link de 2 MB para a Internet.

Falando no servidor Web da PMC, ele estava sendo executado em um Servidor HP NetServer Ip1000r de 1,2 Ghz com 256 MB de memória SDRAM ECC de 133 Mhz, contando ainda com um HD de 15 GB SCSI, utilizando o sistema operacional Slackware Linux 10.1 e o servidor Web Apache 1.3.

Esse servidor foi configurado para realizar o armazenamento de todas as informações possíveis solicitadas pelos *browsers* dos usuários. Neste servidor são encontrados os arquivos correspondentes ao *site* para que possa por meio dele, divulgar algumas informações destinadas aos contribuintes, tais como:

- a) andamentos das obras;
- b) *links* para outros sites de utilidade pública, por exemplo, horário dos ônibus;
- c) oferecer à população serviços on-line a fim de facilitar o dia-a-dia da população, como encontrar a impressão do boleto bancário do IPTU.

³⁰ Elemento fixo que é colocado verticalmente sobre algo onde é destinado ao empilhamento.

A Prefeitura não dispunha de informações de quantos acessos estavam sendo realizados ao site, quantos desses acessos eram realizados internamente e externamente, quais serviços e informações eram mais utilizados pelos visitantes. Tendo essas informações, pode-se justificar o investimento feito. Pensando em obter estas informações, realizou-se a análise dos arquivos de *log* utilizando ferramentas computacionais centradas no servidor Web, onde foi utilizada a técnica de *logging* no servidor e definiram-se as métricas utilizadas. Analisaram-se os resultados obtidos de forma a conseguir a informação dos acessos aos serviços mais utilizados no *site* da Prefeitura Municipal de Criciúma.

5.1 SOFTWARES UTILIZADOS NO DESENVOLVIMENTO

A partir de setembro de 2005, foram instalados no servidor, quatro softwares que foram responsáveis em efetuar a análise dos *logs* para que posteriormente, fosse possível disseminar aos setores responsáveis os resultados obtidos.

Os programas que foram instalados para essa função estão na Tabela 5.

Tabela 5. Informações sobre os softwares instalados

Nome do Software	Versão	Licença	Endereço na Internet
Webalizer	2.01	<i>Open-Source</i>	http://www.mrunix.net/webalizer/
Awstats	6.5	<i>Free</i>	http://awstats.sourceforge.net/
http-analyze	2.4	Educacional	http://http-analyze.org/
Analog	6.0	<i>Free</i>	http://www.analog.cx/

Fonte: CARRADORE. (2006)

Foram instalados mais de um software de análise de *logs* para verificar se algum software oferecia mais informações que os outros.

5.1.1 Webalizer

O Webalizer exibe por meio de uma página HTML estática, os gráficos e as estatísticas, mostrando informações como, por exemplo, a Figura 10.

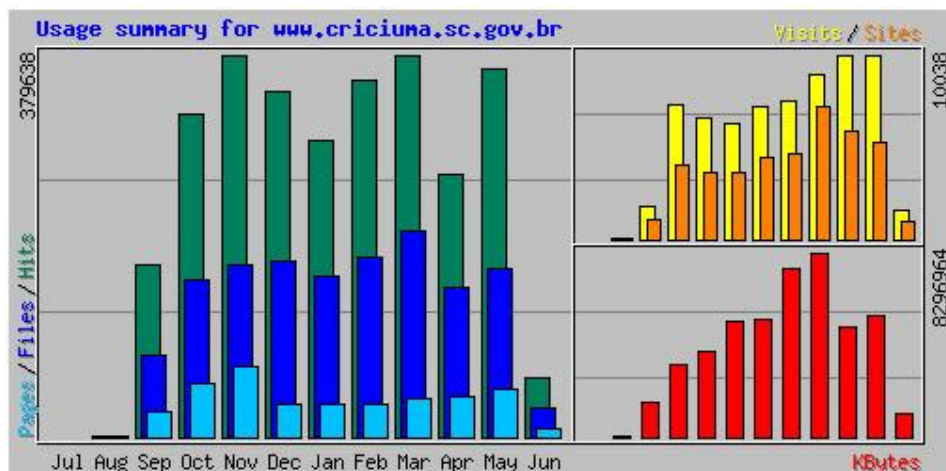


Figura 10. Informação Mensal dos Acessos
Fonte: WEBALIZER (2006)

Este software mostra diversas informações. A página inicial mostra a Figura 10, onde se têm a informação da quantidade de páginas, arquivos, *hits*, visitas e *sites* visualizados e o tráfego em Kbytes em que o servidor trafegou. Abaixo dessa figura, possui uma tabela com esses dados divididos em total por mês e média diária, como se pode ver na Figura 11.

Summary by Month										
Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Jun 2006	9962	4753	1259	259	988	1068942	1559	7559	28521	59776
May 2006	11776	5371	1537	323	5273	5442673	10038	47652	166520	365078
Apr 2006	8673	4959	1329	334	5872	4943861	10020	39894	148784	260214
Mar 2006	12224	6593	1235	289	7173	8296964	8968	38286	204405	378973
Feb 2006	12649	6404	1196	269	4668	7617966	7546	33496	179314	354179
Jan 2006	9506	5136	1077	233	4485	5317132	7247	33402	159226	294690
Dec 2005	11050	5651	1071	203	3595	5185201	6315	33224	175182	342568
Nov 2005	12654	5730	2339	221	3662	3846474	6639	70172	171921	379638
Oct 2005	10367	5060	1692	237	4009	3230791	7365	52473	156863	321404
Sep 2005	5684	2725	834	60	1028	1565341	1817	25042	81771	170540
Aug 2005	953	591	326	1	1	6445	1	326	591	953
Totals						46521790	67515	381526	1473098	2928013

Figura 11. Informações dos Acessos no Apache
Fonte: WEBALIZER (2006)

Caso se queira ver as informações mais detalhadamente, deve-se escolher um mês. Após clicar sobre o mês desejado, é aberta uma nova página Web onde se mostram as informações referentes ao mês solicitado. Nessa página, o Webalizer mostra a quantidade de *hits*, de arquivos visualizados, de páginas vistas, de visitas feitas e o total de Kbytes trafegados que teve o servidor, como se pode ver na Figura 12.

Além dessas informações, e ainda tendo em vista a Figura 12, o Webalizer mostra também a quantidade de visitas únicas, de arquivos diferentes que foram requisitados pelo servidor Web, de *links* externos diferentes que redirecionaram um visitante ao *site* do servidor e de *browsers* diferentes. Também exibe uma quantidade média e máxima de *hits* por hora e arquivos, páginas, *hits*, visitas e Kbytes trafegados por dia.

Monthly Statistics for May 2006		
Total Hits	365078	
Total Files	166520	
Total Pages	47652	
Total Visits	10038	
Total KBytes	5442673	
Total Unique Sites	5273	
Total Unique URLs	913	
Total Unique Referrers	368	
Total Unique User Agents	339	
	Avg	Max
Hits per Hour	490	4226
Hits per Day	11776	28580
Files per Day	5371	9076
Pages per Day	1537	3443
Visits per Day	323	536
KBytes per Day	175570	298944

Figura 12. Informações Mostradas pelo Webalizer
Fonte: WEBALIZER (2006)

De acordo com a Figura 13, o Webalizer também mostra qual a quantidade de códigos de resposta que o Apache retornou para o *browser* do visitante no mês.

Hits by Response Code	
Code 200 - OK	166520
Code 206 - Partial Content	1594
Code 301 - Moved Permanently	7585
Code 304 - Not Modified	185813
Code 400 - Bad Request	56
Code 403 - Forbidden	15
Code 404 - Not Found	3342
Code 405 - Method Not Allowed	137
Code 408 - Request Timeout	16

Figura 13. Códigos de Resposta ao Browsers
Fonte: WEBALIZER (2006)

O Webalizer também mostra algumas das informações referidas na Figura 12 divididas por dia, conforme a Figura 14. Essas informações são muito úteis ao administrador do servidor, pois é possível saber quais são os dias da semana em que se pode fazer manutenção no servidor, pois caso for preciso deixá-lo fora do ar, não terá muitas visitas perdidas. Como em gráfico se não consegue obter o valor exato das informações, o Webalizer mostra também os dados obtidos em texto, conforme a Figura 15.

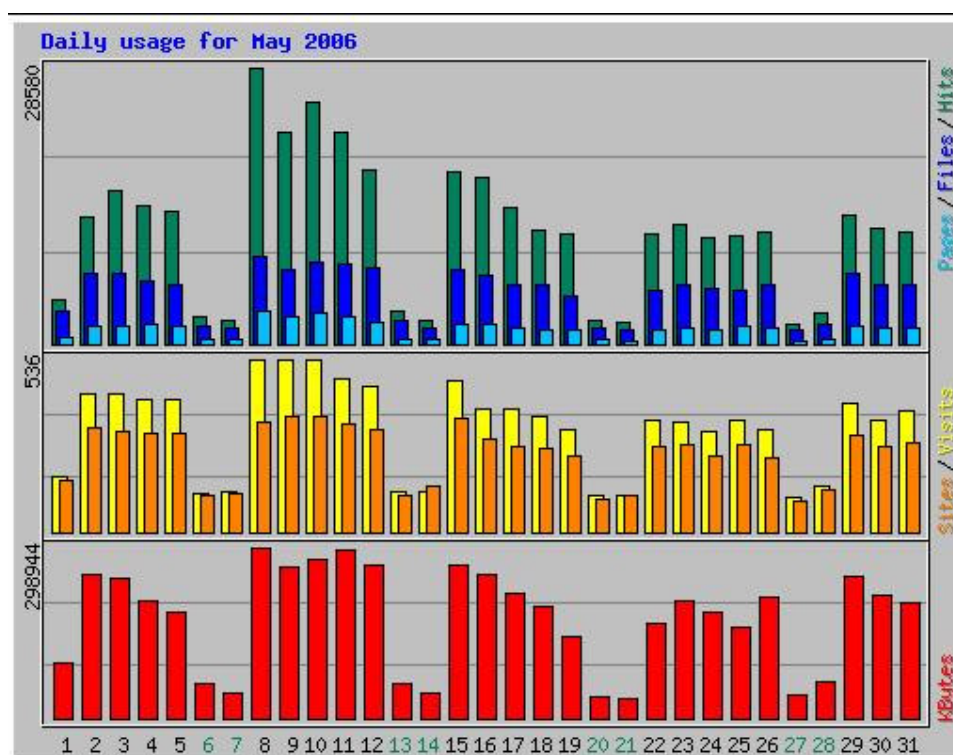


Figura 14. Informações do Webalizer em Gráfico Separadas por Dia
Fonte: WEBALIZER (2006)

Daily Statistics for May 2006												
Day	Hits		Files		Pages		Visits		Sites		KBytes	
1	4586	1.26%	3369	2.02%	674	1.41%	172	1.71%	162	3.07%	98020	1.80%
2	13190	3.61%	7362	4.42%	1790	3.76%	426	4.24%	325	6.16%	252779	4.64%
3	15860	4.34%	7199	4.32%	1901	3.99%	428	4.26%	311	5.90%	244066	4.48%
4	14359	3.93%	6421	3.86%	2059	4.32%	413	4.11%	306	5.80%	205802	3.78%
5	13691	3.75%	6135	3.68%	1879	3.94%	408	4.06%	303	5.75%	185185	3.40%
6	2825	0.77%	1810	1.09%	435	0.91%	122	1.22%	116	2.20%	61676	1.13%
7	2356	0.65%	1624	0.98%	406	0.85%	125	1.25%	120	2.28%	43169	0.79%
8	28580	7.83%	9076	5.45%	3443	7.23%	533	5.31%	341	6.47%	298944	5.49%

Figura 15. Informações do Webalizer em Texto Separadas por Dia
Fonte: WEBALIZER (2006)

E demonstra também, por meio de um gráfico, a utilização, de hora em hora, a quantidade de páginas visualizadas, de arquivos vistos e o total de *hits*, como pode ser visto na Figura 16. Essa informação também é muito útil ao administrador do servidor, pois, como foi falado nas estatísticas de dia da semana, é possível saber quais são os horários em que se pode fazer manutenção no servidor. Tendo o mesmo problema comentado na Figura 15, o software também mostra as informações em texto, de acordo com a Figura 17.

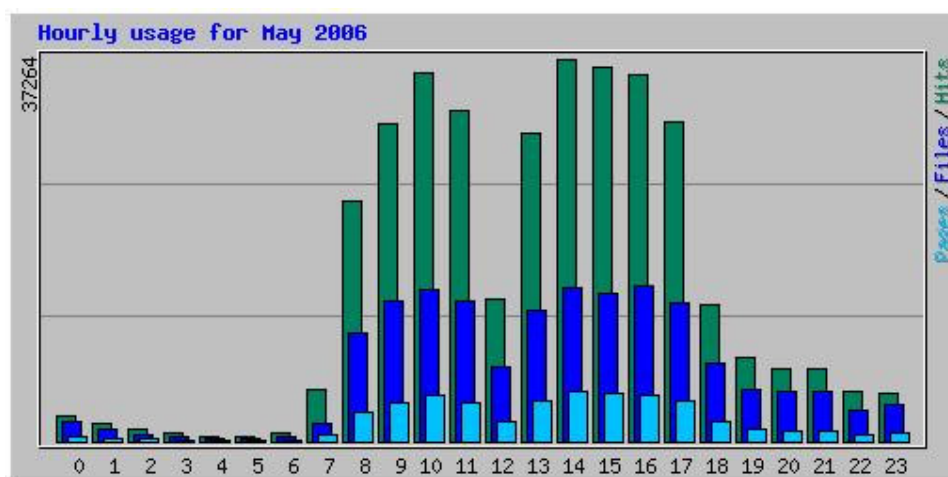


Figura 16. Informações do Webalizer em Gráfico Separadas por Hora
Fonte: WEBALIZER (2006)

Hourly Statistics for May 2006												
Hour	Hits			Files			Pages			KBytes		
	Avg	Total		Avg	Total		Avg	Total		Avg	Total	
0	77	2394	0.66%	59	1856	1.11%	13	421	0.88%	1513	46898	0.86%
1	55	1734	0.47%	39	1237	0.74%	10	335	0.70%	1041	32282	0.59%
2	36	1120	0.31%	20	633	0.38%	6	211	0.44%	692	21446	0.39%
3	28	894	0.24%	15	490	0.29%	5	160	0.34%	468	14513	0.27%
4	15	484	0.13%	9	297	0.18%	3	109	0.23%	482	14930	0.27%
5	15	484	0.13%	9	304	0.18%	3	99	0.21%	221	6843	0.13%
6	28	877	0.24%	12	384	0.23%	5	164	0.34%	428	13283	0.24%

Figura 17. Informações do Webalizer em Texto Separadas por Hora
Fonte: WEBALIZER (2006)

Com esse software, é possível também verificar quais são os *links* do *site* que tiveram mais *hits*. É possível também definir que o *link* que possuir mais *hits* pode ser também o que possui mais visibilidade. Geralmente, o link que possui maior visibilidade é a página inicial, nesse caso: www.criciuma.sc.gov.br. No entanto, em um *site* com diversos *links*, pode acontecer que muitos visitantes vão ao *site* unicamente para encontrar alguma informação e esta não está na página inicial, fazendo com que os visitantes coloquem esse *link* em seus favoritos³¹. Com essa estatística, problemas podem ser detectados, tornando o *site* mais dinâmico. Essa informação pode ser vista por meio da Figura 18.

³¹ Recurso dos programas de navegação que permitem o armazenamento dos endereços da Internet preferidos.

Top 30 of 913 Total URLs					
#	Hits		KBytes		URL
1	12109	3.32%	18995	0.35%	/
2	9205	2.52%	515686	9.47%	/secom/
3	7814	2.14%	16511	0.30%	/secom/fotolegenda.html
4	2171	0.59%	2973	0.05%	/avisos.html
5	551	0.15%	4352	0.08%	/secom/secretarias.html
6	432	0.12%	844	0.02%	/IPTU lei decretos.html
7	432	0.12%	1743	0.03%	/classificacao.html
8	389	0.11%	1971	0.04%	/secom/links.html
9	385	0.11%	4143	0.08%	/secom/fotos.html
10	310	0.08%	27116	0.50%	/Estagio Probatorio/caaep0002.05.html

Figura 18. *Links* mais Visualizados
 Fonte: WEBALIZER (2006)

O Webalizer também informa quais foram os *links* em que trafegaram mais dados, conforme a Figura 19. Essa informação é importante, pois caso o *site* estiver lento, e o problema não for detectado por falta de recursos de hardware e, for descoberto que o *link* de *upload*³² com a Internet está sempre com alta utilização, por meio desse item é possível descobrir quais são as páginas em que estão trafegando mais dados. Depois de definir quais são, é possível verificar se essas não possuem algo de errado com elas, como por exemplo, as imagens com alta definição. Quanto maior a definição, maior o tamanho da figura, e em muitas páginas Web, as imagens não precisam possuir alta definição e, com isso, podem-se diminuir as resoluções das figuras, consumindo menos recurso de banda.

³² É a transferência de arquivo de um computador para outro usando qualquer protocolo de comunicação.

Top 30 of 913 Total URLs By KBytes					
#	Hits		KBytes		URL
1	9205	2.52%	515686	9.47%	/secom/
2	49	0.01%	35522	0.65%	/ACT/orientador.pdf
3	310	0.08%	27116	0.50%	/Estagio Probatório/caaep0002.05.html
4	12109	3.32%	18995	0.35%	/
5	7814	2.14%	16511	0.30%	/secom/fotolegenda.html
6	83	0.02%	10790	0.20%	/ACT/educacaoinfantilalterada.PDF
7	118	0.03%	8544	0.16%	/ACT/05a08.PDF
8	57	0.02%	7056	0.13%	/ACT/educacaoinfantil.PDF
9	551	0.15%	4352	0.08%	/secom/secretarias.html
10	385	0.11%	4143	0.08%	/secom/fotos.html

Figura 19. Links que mais Trafegam
Fonte: WEBALIZER (2006)

Outra informação que útil que o Webalizer informa é qual a página de entrada e saída do *site*. A principal página de entrada, como comentada anteriormente, pode mostrar se existe alguma página interna do site que os usuários estão visualizando sem passar pela página inicial. A página de saída é importante saber, pois pode haver algum *link* em uma das páginas do *site* que faz com que o usuário acabe saindo do mesmo. Essa informação se pode ver na Figura 20 e 21.

Top 10 of 119 Total Entry Pages					
#	Hits		Visits		URL
1	12109	3.32%	7238	82.14%	/
2	9205	2.52%	539	6.12%	/secom/
3	2171	0.59%	156	1.77%	/avisos.html
4	385	0.11%	134	1.52%	/secom/fotos.html
5	389	0.11%	54	0.61%	/secom/links.html
6	81	0.02%	47	0.53%	/secom/WebAlbum/Cidade Mineira/
7	7814	2.14%	45	0.51%	/secom/fotolegenda.html
8	432	0.12%	44	0.50%	/PTU lei decretos.html
9	310	0.08%	36	0.41%	/Estagio Probatório/caaep0002.05.html
10	269	0.07%	33	0.37%	/secom/secom.html

Figura 20. Páginas de Entrada do mês de Maio
Fonte: WEBALIZER (2006)

Top 10 of 148 Total Exit Pages					
#	Hits		Visits		URL
1	7814	2.14%	3668	37.97%	/secom/fotolegenda.html
2	12109	3.32%	3171	32.83%	/
3	9205	2.52%	620	6.42%	/secom/
4	2171	0.59%	487	5.04%	/avisos.html
5	551	0.15%	186	1.93%	/secom/secretarias.html
6	432	0.12%	162	1.68%	/classificacao.html
7	310	0.08%	128	1.33%	/Estagio Probatório/caaep0002.05.html
8	389	0.11%	128	1.33%	/secom/links.html
9	432	0.12%	105	1.09%	/IPTU lei decretos.html
10	385	0.11%	88	0.91%	/secom/fotos.html

Figura 21. Páginas de Saída do mês de Maio
 Fonte: WEBALIZER (2006)

Este programa também informa quem acessou o servidor, sendo identificado pelo seu IP utilizado. Quando o visitante, que está visualizando o *site* da PMC, estiver dentro das dependências da mesma, ele está utilizando um IP que é o do servidor *proxy*. Com essa informação, é possível saber se quem está visualizando o site da Prefeitura está utilizando um computador dentro das dependências da PMC, ou seja, funcionários, ou se são visitantes externos. Quando uma empresa coloca o seu *site* na Internet, ela deseja saber quem o está visualizando, se está valendo o investimento, e isso não é diferente com a PMC. Com essa informação é possível saber se são visitantes externos ou pessoas que trabalham na prefeitura que estão acessando o *site*. Essa informação é classificada conforme a quantidade de *hits*, como mostra a figura 22, e em total de dados trafegados.

Top 30 of 5273 Total Sites									
#	Hits		Files		KBytes		Visits		Hostname
1	17870	4.89%	1292	0.78%	50379	0.93%	254	2.53%	200.215.36.17
2	16758	4.59%	6239	3.75%	317043	5.83%	105	1.05%	200.193.33.1
3	9889	2.71%	3553	2.13%	129425	2.38%	135	1.34%	200.193.20.174
4	5354	1.47%	1072	0.64%	47724	0.88%	124	1.24%	200.102.53.196
5	5147	1.41%	248	0.15%	13948	0.26%	47	0.47%	200.180.33.154
6	4975	1.36%	1349	0.81%	47705	0.88%	135	1.34%	200.192.77.254
7	4666	1.28%	303	0.18%	13660	0.25%	80	0.80%	200.193.64.110
8	3107	0.85%	383	0.23%	16884	0.31%	127	1.27%	200.138.215.43

Figura 22. IP dos Visitantes ao Site
Fonte: WEBALIZER (2006)

Uma outra informação que é exibida pelo Webalizer, é quais as páginas que referenciam a página da PMC, ou seja, quais são as páginas na Internet que possuem um *link* para o site da Prefeitura Municipal de Criciúma. O Webalizer conta também como referência as páginas internas do próprio *site*, como pode ser visto na Figura 23.

Top 30 of 368 Total Referrers			
#	Hits		Referrer
1	123939	33.95%	http://www.criciuma.sc.gov.br/secom/
2	97063	26.59%	http://www.criciuma.sc.gov.br/
3	46938	12.86%	http://www.criciuma.sc.gov.br/secom/fotolegenda.html
4	46767	12.81%	- (Direct Request)
5	17699	4.85%	http://www.criciuma.sc.gov.br
6	5516	1.51%	http://www.criciuma.sc.gov.br/avisos.html

Figura 23. Sites da Internet que referenciam a página da PMC
Fonte: WEBALIZER (2006)

O Webalizer também mostra quais são as palavras ou frases que os visitantes utilizam em *sites* de buscas para encontrar a página da PMC, conforme a Figura 24. Com essa informação, é possível saber as principais palavras que os usuários utilizam

para encontrar a página da PMC e com isso o *webmaster*³³ pode colocar essas no cabeçalho do *site*, para fazer com que os buscadores encontrem mais facilmente o site da PMC. O Webalizer mostra também quais são os *browsers* que os visitantes utilizam. Com essa informação, o *webmaster*, pode antes de colocar na Internet uma nova ferramenta, testar utilizando os *browsers* mais usados, para verificar se vai ocorrer algum problema.

Top 20 of 474 Total Search Strings			
#	Hits		Search String
1	816	25.06%	criciuma
2	386	11.86%	prefeitura de criciuma
3	366	11.24%	prefeitura municipal de criciuma
4	236	7.25%	prefeitura criciuma
5	179	5.50%	criciãma
6	109	3.35%	prefeitura municipal de criciãma
7	71	2.18%	prefeitura de criciãma
8	60	1.84%	prefeitura criciãma
9	50	1.54%	criciumatrans
10	49	1.50%	criciuma sc

Figura 24. Palavras que foram utilizadas para encontrar o *site* em Maio
Fonte: WEBALIZER (2006)

5.1.2 http-analyze

Como o Webalizer, o http-analyze também exibe as informações retiradas do *log* por meio de uma página HTML estática, utilizando gráficos e as estatísticas. Mostra essas informações como, por exemplo, a Figura 25.

³³ Alguém que tem a tarefa de criar e realizar a gestão de um determinado site.

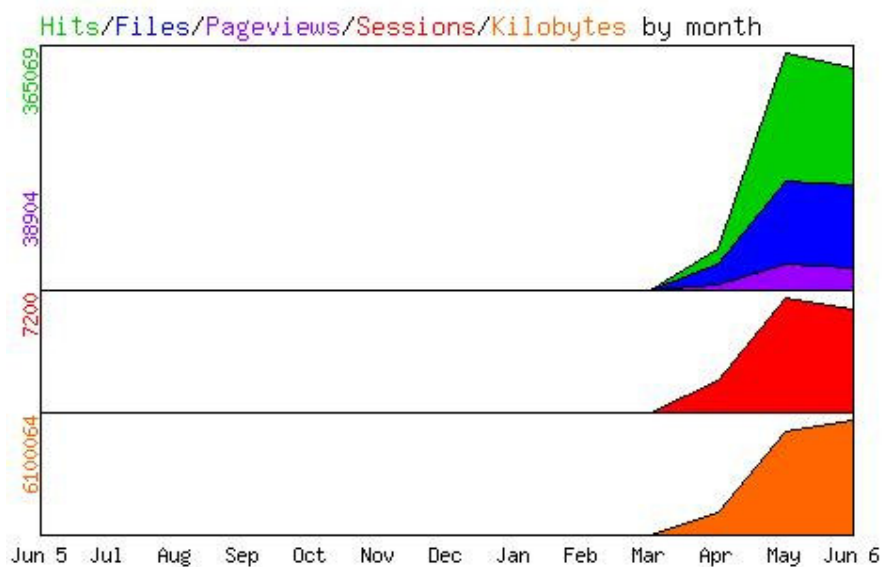


Figura 25. Informações Exibidas pelo http-análise em Gráfico

Fonte: HTTP-ANALYZE (2006)

Esse programa não se aprofunda muito nas estatísticas, oferecendo apenas algumas informações básicas. Na Figura 26, o programa está exibindo informações referentes aos últimos 12 meses. Informações sobre *hits* que o servidor enviou, a quantidade de arquivos visualizados, de páginas que foram vistas, o total de Kbytes em que o servidor navegou e a quantidade de sessões feitas. Sessões são os acessos de um único IP dentro de 24 horas. Como na Figura 25 é difícil de definir o número exato das estatísticas, o software mostra uma caixa de texto com as mesmas informações, como se pode ver na Figura 26.

Short statistics for June 2006 (updated more frequently)						
Month	Hits	Files	Cached	Pageviews	Sessions	KB sent
June 2006	67812	31875	33734	6851	1275	1220013
May 2006	365069	166441	185813	38904	7200	5443560
April 2006	63029	37682	23094	7996	1967	1178139
March 2006	0	0	0	0	0	0
February 2006	0	0	0	0	0	0
January 2006	0	0	0	0	0	0
December 2005	0	0	0	0	0	0
November 2005	0	0	0	0	0	0
October 2005	0	0	0	0	0	0
September 2005	0	0	0	0	0	0
August 2005	0	0	0	0	0	0
July 2005	0	0	0	0	0	0
Total	495910	235998	242641	53751	10442	7841712
Average	41325	19666	20220	4479	870	653476

Figura 26. Informações Exibidas pelo http-análise em Texto
 Fonte: HTTP-ANALYZE (2006)

Para ver as informações mais detalhadas, é preciso escolher um mês. Com isso, o programa mostra as informações referentes ao mês escolhido. Essas são as quantidades totais de *hits*, total de arquivos enviados (código 200), total de arquivos que não foram enviados por que já se encontravam no *cache* do *browser* do visitante (código 304) e também mostra a quantidade de outros códigos a que o servidor respondeu. Informa também sobre o total de páginas vistas, o total em Kbytes que foi pedido ao servidor, tendo essa informação separada por dados que foram enviados e dados que não foram por quê os arquivos já estavam no *cache* do *browser*, como se pode ver na Figura 27.

Monthly Summary			
Total hits (any request)		100.00%	365069
Total files sent (Code 200)		45.59%	166441
Total files saved by cache (Code 304)		50.90%	185813
Other response codes (see below)		3.51%	12815
Total pageviews			38904
Total KB requested		100.00%	11832150
Total KB transferred (Code 200)		46.01%	5443560
Total KB saved by cache (Code 304)		53.99%	6388590

Figura 27. Informações Exibidas pelo http-analyze
Fonte: HTTP-ANALYZE (2006)

Outras informações são o total de *links* visitados dentro do *site*, de visitantes únicos, de sessões, e o total de *browsers* diferentes que os visitantes utilizaram. Também mostra a quantidade máxima e a média por dia dos *hits*, arquivos enviados, arquivos que estavam no *cache* do usuário, páginas visualizadas, sessões e dados trafegados, como mostra a Figura 28.

Total unique URLs		915
Total unique sites		5272
Total user sessions		7200
Total unique agents		459
Total unique referrer URLs		2313
	Maximum	Average
Hits per day	28580	11776
Files sent per day	9075	5369
Files cached per day	18621	5993
Pageviews per day	2763	1254
Sessions per day	345	232
KB sent per day	298946	175571

Figura 28. Outras Informações Exibidas pelo http-analyze
Fonte: HTTP-ANALYZE (2006)

O http-analyze mostra também informações referentes aos *logs*, aos códigos que o apache enviou aos *browsers* e quais foram os métodos que os *browsers* utilizaram para requisitar os dados do servidor Web, como mostra a Figura 29.

Logfile statistics	
Total logfile entries read	365078
Total logfile entries processed	365069
Empty Request Method	9
Other Response Codes	
Partial Content (Code 206)	1594
Moved Permanently (Code 301)	7585
Bad Request (Code 400)	56
Forbidden (Code 403)	15
Not Found (Code 404)	3342
Method Not Allowed (Code 405)	31
Request Timeout (Code 408)	7
Request Methods other than GET/POST	
HEAD	3
PUT	2
OPTIONS	92
CONNECT	88

Figura 29. Mais Algumas Informações Exibidas pelo http-analyze
Fonte: HTTP-ANALYZE (2006)

O software também mostra relatórios sobre *hits*, arquivos visualizados, páginas visualizadas, sessões e a quantidade de dados que foram trafegados e classificados por hora, dia e dias da semana. Mostra também, em forma de lista, quais foram e quantas foram as vezes que os *links* foram acessados, mostrando também o que os visitantes tentaram acessar e não conseguiram, pois não existia. Essa informação é muito importante para o *webmaster*, pois é possível saber quais são os *links* que não estão funcionando e com isso arrumá-los. Lista também quais foram os *browsers* utilizados pelos visitantes, e as páginas que fazem referência a página da PMC.

5.1.3 analog

O analog mostra todas as estatísticas em apenas uma página HTML. Dos programas testados, o analog é o que mostra as informações de um jeito mais simples.

No início da página das estatísticas retiradas do site, mostra informações gerais, como a quantidade de requisições completadas pelo servidor Web, a média de requisições completadas por dia, o total de requisições por página, a quantidade de requisições por página a cada dia, a quantidade de requisições que não foram completadas, o total de arquivos diferentes que foram requisitados, a quantidade de visitantes únicos no *site*, o total de dados trafegados e a média diária do tráfego, como se pode ver na Figura 30.

```

Figures in parentheses refer to the 7-day period ending 07-Jun-2006 04:40.
Successful requests: 491,775 (87,811)
Average successful requests per day: 11,330 (12,544)
Successful requests for pages: 55,481 (9,239)
Average successful requests for pages per day: 1,278 (1,319)
Failed requests: 4,910 (616)
Redirected requests: 10,354 (1,956)
Distinct files requested: 1,049 (769)
Distinct hosts served: 7,245 (1,366)
Data transferred: 7.71 gigabytes (1.59 gigabytes)
Average data transferred per day: 181.92 megabytes (232.77 megabytes)

```

Figura 30. Informações Gerais do Analog
Fonte: ANALOG (2006)

O analog também mostra a estatística de quantas requisições e quantas páginas foram vistas, mostrando essa informação separada por hora, dias da semana e mês, conforme a Figura 31.

day	reqs	pages	
Sun	15738	2336	+++++++
Mon	99354	10392	+++++++
Tue	109011	12002	+++++++
Wed	88533	9824	+++++++
Thu	84120	9810	+++++++
Fri	79132	8832	+++++++
Sat	15887	2285	+++++++

Figura 31. Informações Classificada por Dias da Semana
Fonte: ANALOG (2006)

Conforme a Figura 32, o software também mostra quais foram e quantas vezes foram usadas as palavras para encontrar o site do servidor Web.

reqs	search term
3536	criciuma
2341	prefeitura
2154	de
989	municipal
828	crici�ma
300	sc
199	da
138	fotos

Figura 32. Palavras Usadas para Encontrar o Site no Servidor Web
Fonte: ANALOG (2006)

O analog mostra em um gr fico no formato de pizza e em uma tabela de texto, quantas vezes e quais foram os sistemas operacionais utilizados pelos visitantes, como pode ser visto na Figura 33.

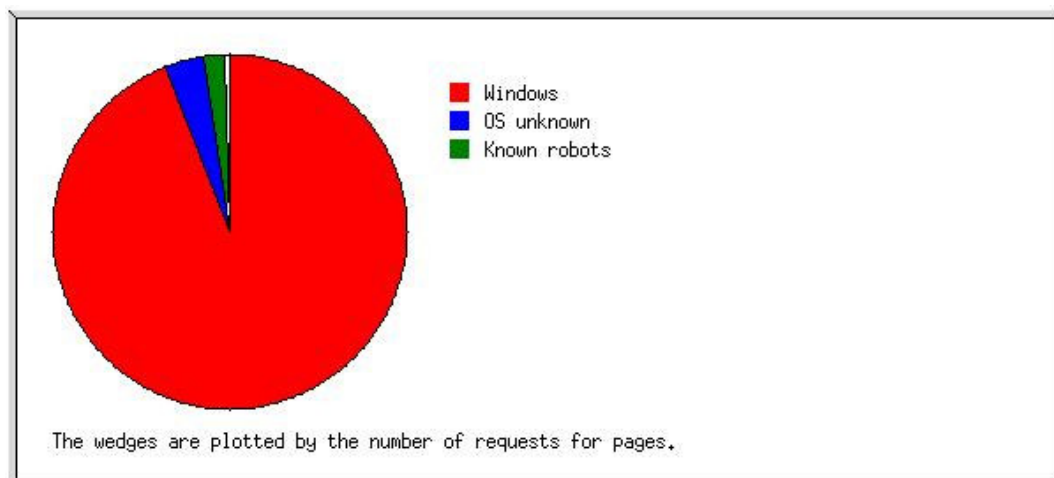


Figura 33. *Browsers* Utilizados pelos Visitantes
Fonte: ANALOG (2006)

O analog também mostra quantas vezes e quais foram os códigos enviados ao browser do visitante, a quantidade de arquivos enviados dividindo essa informação por tamanho de arquivo (caso o site estiver com muita utilização de *upload*, com essa estatística pode-se saber se o site está utilizando arquivos muitos grandes, como por exemplo, fotos de alta resolução, e com isso conseguir diminuir esses arquivos), a quantidade em que foram solicitados os arquivos, dividido por tipo de arquivo, por exemplo: jpg, html, gif, os diretórios do site mais utilizados e os arquivos mais vezes visualizados.

5.1.4 Awstats

O Awstats é o software, que informa a maior quantidade de estatísticas, além de ser também o mais organizado.

O Awstats mostra as informações recolhidas do *log* divididas por mês. Para verificar as estatísticas de algum mês, basta escolher qual o mês desejado e o ano no início da página HTML.

Ele mostra a quantidade de visitantes únicos, número de visitas, páginas visualizadas, *hits* e a quantidade de dados trafegados pelo servidor, dividindo essas informações por mês, conforme a Figura 34.

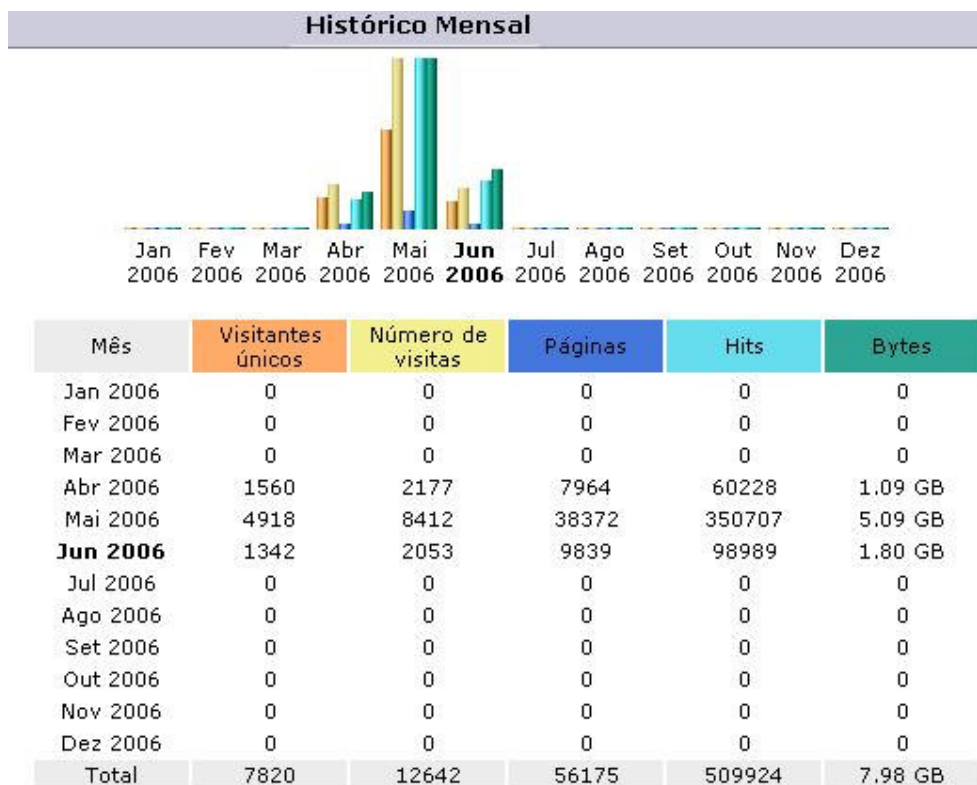


Figura 34. Informações Mostradas pelo Awstats
Fonte: AWSTATS (2006)

Também mostra as informações referentes ao número de visitas, páginas visualizadas, *hits* e a quantidade que o servidor trafegou sendo divididas essas informações por dias do mês e a quantidade de páginas visitadas, *hits* e bytes trafegados, divididos por hora e dia da semana como mostra a Figura 35.



Figura 35. Informações Separadas por Dia da Semana do Mês de Maio
Fonte: AWSTATS (2006)

O Awstats mostra o IP dos visitantes que estiveram no site, mostrando também à quantidade de páginas visitadas, os *hits*, a quantidade de bytes que esses IP trafegaram e quando foi a sua última visita.

Mostra quais foram os *spiders* que vasculharam o site do servidor, quantos *hits* esses softwares requisitaram e quanto trafegaram.

Uma informação que é muito importante e que só o Awstats identificou foi a duração das visitas. Ele identifica a quantidade de visitas feitas, dividindo por tempo, como pode se ver na Figura 36.

Duração das visitas			
	Número de visitas: 8412 - Med: 475 s	Número de visitas	Por cento
0s-30s		3836	45.6 %
30s-2mn		1993	23.6 %
2mn-5mn		688	8.1 %
5mn-15mn		648	7.7 %
15mn-30mn		361	4.2 %
30mn-1h		463	5.5 %
1h+		423	5 %

Figura 36. Informações Separadas por Dia da Semana
Fonte: AWSTATS (2006)

Este software também mostra os tipos de arquivos que possuíram mais *hits*, quais os *links* que tiveram mais acesso, quais foram as páginas de entrada ao *site* e as páginas que o visitante acaba visualizando por último e quais os sistemas operacionais e *browsers* com suas versões utilizadas pelos visitantes.

O Awstats é também o único que mostra de onde o visitante acessou o *site*, se foi por meio de sites de busca, um link de um *site* externo ou se foi digitando diretamente a URL da página. Documenta a partir de qual o site que os visitantes estavam visualizando antes de chegar à página da PMC, mostra também quais foram as palavras e frases utilizadas nos sites de buscas para também chegar à página, quantas pessoas adicionam o site da PMC nos favoritos (outra informação que apenas o Awstats informa) e a quantidade de erros HTTP que o servidor relatou. O Awstats também é o único dos quatro softwares que remove os tráfegos e *hits* não vistos como, por exemplo, os dados trafegados e os *hits* solicitados pelos *spiders* de sites de busca, como o do *site* de busca google.com, das estatísticas gerais do *site*. Pelo motivo de que o Awstats não registrar os tráfegos do que são visualizados por visitantes, algumas estatísticas, como por exemplo, o número de visitantes totais no *site*, é menor, porém é mais exato que nos outros programas.

5.2 MÉTRICAS UTILIZADAS E RESULTADOS OBTIDOS

Para definir o perfil dos acessos ao *site* da Prefeitura Municipal de Criciúma, algumas métricas foram definidas. O software Webalizer foi instalado e configurado no servidor Web em setembro de 2005, já o Awstats foi instalado em meados de abril de 2006. Como não foi possível analisar mais de um mês inteiro com o Awstats, para definir o resultado das métricas, será utilizado o software Webalizer, mas como há

informações que apenas o Awstats informa, como por exemplo, o tempo de acesso dos visitantes, na definição do resultado dessas métricas, estará sendo utilizado o Awstats.

De acordo com Friedlein (2003), para se ter um conhecimento instantâneo, devem-se utilizar as estatísticas referentes a uma semana, mas é necessário por volta de no mínimo três meses para se obterem as reais tendências. Com isso, será utilizado, sempre que possível o Webalizer e os meses de março, abril e maio para realizar as análises e obter a resposta às métricas. Quando isso não for possível, será utilizado o Awstats com os dados do mês de maio. Abaixo estão as métricas que foram utilizadas e as estatísticas que se obtiveram:

- a) a média mensal de visitas totais: para definir a quantidade de visitas mensais, foram verificados nas estatísticas, que o mês de março teve um total de 8.968 visitas, o mês de abril teve 10.020 visitas e o mês de maio teve um total de 10.038 acessos ao site, de acordo com a Figura 11. Com essas informações, chega-se a obter a média de acessos total ao site da PMC que é de 9.675 visitas por mês. Sabendo que a população de Criciúma tem por volta de 170 mil habitantes, é importante saber a quantidade de visitas que o site da PMC possui, pois através dessa métrica pode-se ter uma idéia da quantidade de pessoas que visitam o *site* e definir se será preciso oferecer mais serviços na página para aumentar o número de visitas.
- b) a média mensal de visitas únicas: de acordo com a Figura 37, o mês de março possuiu o total de 7.173 visitantes únicos, o mês de abril teve 5.872 visitantes únicos, conforme a Figura 38 e o mês de maio teve 5.273 visitantes únicos, conforme a Figura 12. Tendo esses dados em

mãos, chega-se ao resultado de que a quantidade média de visitantes únicos é de 6.106.

Monthly Statistics for March 2006	
Total Hits	378973
Total Files	204405
Total Pages	38286
Total Visits	8968
Total KBytes	8296964
Total Unique Sites	7173
Total Unique URLs	804
Total Unique Referrers	64
Total Unique User Agents	50

Figura 37. Visitas Únicas em Março
Fonte: WEBALIZER (2006)

Monthly Statistics for April 2006	
Total Hits	260214
Total Files	148784
Total Pages	39894
Total Visits	10020
Total KBytes	4943861
Total Unique Sites	5872
Total Unique URLs	764
Total Unique Referrers	327
Total Unique User Agents	360

Figura 38. Visitas Únicas em Abril
Fonte: WEBALIZER (2006)

- c) a média mensal da quantidade de visitas que não estão sendo originadas de computadores do pátio da PMC: Para descobrir qual é a quantidade de visitas ao *site* da PMC que não originaram do pátio da PMC foi obtido a quantidade total de visitas a página da PMC e subtraído com o número de acessos ao *site* feito pelo servidor *proxy*. De acordo com o item a), desse capítulo, o mês de março teve 8.968 visitas, o mês de abril teve

10.020 visitas e o mês de maio teve um total de 10.038 acessos. Visualizadas as estatísticas obtidas com o Webalizer, pode-se dizer que o número de acessos ao *site* da PMC pelo servidor *proxy* foi o total de 108 visitas no mês de março, 79 visitas em abril e 105 visitas em maio. Com esses dados, determina-se que a quantidade média por mês de usuários externos que visualizam o site da PMC é de 9.578 visitas.

- d) a média de visitas totais por dia: a Figura 11 nos mostra a média diária dos meses de março, abril e maio. O mês de março teve uma média de 289 visitas por dia, abril teve uma média de 334 visitas e no mês de maio, a média de visita foi de 323 visitantes. Tendo essas informações, a média de visita diária no *site* da PMC é de 315 visitantes.
- e) qual é o principal caminho em que um usuário utiliza para visitar o *site*: quando um usuário visita o site da PMC, ou ele encontrou o link em algum outro site, ou encontrou por um site de busca ou então ele digitou diretamente o endereço do *site* no *browser*. De acordo com a Figura 39, a maioria dos visitantes acessa o site da PMC digitando o endereço do site no *browser*. Essa métrica é importante saber, pois pode, por exemplo, saber se alguma propaganda de marketing feita, está obtendo o retorno esperado.

Conectado a partir de				
Origem	Páginas	Por cento	Hits	Por cento
Endereço direto / Favoritos	18093	81.4 %	35924	86.8 %
Links de um NewsGroup				
Link de um Buscador - Lista completa	3615	16.2 %	3712	8.9 %
- Google	2648 2649			
- MSN	441 481			
- Yahoo	420 433			
- Google (Images)	87 128			
- AltaVista	19 19			
- Unknown search engines	0 2			
Link de uma página externa (outros sites que não buscadores) - Lista completa	519	2.3 %	1715	4.1 %
- http://buscador.terra.com.br/Default.aspx	89 89			
- http://busca.uol.com.br/www/index.html	37 37			
- http://www.interlegis.gov.br/comunidade/20040817103027/santa-cat...	35 35			
- http://www.amrec.com.br/municipios/index.php	22 22			
- http://www.guiacatarinense.com.br/criciuma/menu.htm	21 21			
- http://buscador.terra.com.br/default.aspx	19 19			
- http://www.guiademidia.com.br/cidades/santacatarina.htm	19 19			
- http://www.virtualmidiacri.com.br	17 17			
- http://www.actuonibus.com.br/consulta.php	15 15			
- http://pt.wikipedia.org/wiki/Crici%C3%BAma_(Santa_Catarina)	13 13			
- Outros visitantes	232 1428			
Origem Desconhecida				

Figura 39. Conectados a Partir De
Fonte: AWSTATS (2006)

f) a principal página de entrada do site: conforme as Figura 40, 41 e 20, a primeira página que os visitantes visualizaram quando se conectaram ao servidor Web da PMC foi a página inicial do *site*. Através dessa métrica é possível saber se os visitantes estão visualizando algum link do *site* que pode estar nos favoritos, fazendo com que os visitantes não visualizem a página inicial.

Top 10 of 114 Total Entry Pages					
#	Hits	Visits	URL		
1	8184	5229	2.16%	70.49%	/
2	8079	1026	2.13%	13.83%	/secom/
3	502	188	0.13%	2.53%	/secom/fotos.html
4	6358	139	1.68%	1.87%	/secom/fotolegenda.html
5	1529	105	0.40%	1.42%	/avisos.html
6	311	61	0.08%	0.82%	/secom/emigrante.html
7	699	56	0.18%	0.75%	/secom/secretarias.html
8	115	54	0.03%	0.73%	/secom/WebAlbum/Ensaio Dine/
9	317	51	0.08%	0.69%	/secom/secom.html
10	393	49	0.10%	0.66%	/Estagio Probatório/caaep0002.05.html

Figura 40. Páginas de Entrada ao Site no Mês de Março
Fonte: WEBALIZER (2006)

Top 10 of 128 Total Entry Pages					
#	Hits		Visits		URL
1	10811	4.15%	7172	81.32%	/
2	6927	2.66%	583	6.61%	/secom/
3	458	0.18%	150	1.70%	/secom/fotos.html
4	1713	0.66%	128	1.45%	/avisos.html
5	495	0.19%	65	0.74%	/secom/links.html
6	5742	2.21%	64	0.73%	/secom/fotolegenda.html
7	408	0.16%	48	0.54%	/Estagio Probatorio/caaep0002.05.html
8	659	0.25%	42	0.48%	/IPTU lei decretos.html
9	96	0.04%	42	0.48%	/secom/WebAlbum/Cidade Mineira/
10	257	0.10%	40	0.45%	/secom/agenda.html

Figura 41. Páginas de Entrada ao Site no Mês de Abril
Fonte: WEBALIZER (2006)

g) a principal página de saída do site: conforme as Figuras 42, a 43 e 21, a página que os visitantes visualizam por último, antes de sair da página da PMC é o endereço www.criciuma.sc.gov.br/secom/fotolegenda.html. Essa métrica pode ajudar a descobrir se existe algum *link* do *site* que faz com que os usuários acabem saindo da página.

Top 10 of 122 Total Exit Pages					
#	Hits		Visits		URL
1	6358	1.68%	3152	36.90%	/secom/fotolegenda.html
2	8184	2.16%	2241	26.24%	/
3	8079	2.13%	939	10.99%	/secom/
4	1529	0.40%	443	5.19%	/avisos.html
5	699	0.18%	258	3.02%	/secom/secretarias.html
6	551	0.15%	215	2.52%	/classificacao.html
7	393	0.10%	166	1.94%	/Estagio Probatorio/caaep0002.05.html
8	464	0.12%	160	1.87%	/secom/links.html
9	317	0.08%	94	1.10%	/secom/secom.html
10	502	0.13%	89	1.04%	/secom/fotos.html

Figura 42. Páginas de Saída ao Site no Mês de Março
Fonte: WEBALIZER (2006)

Top 10 of 137 Total Exit Pages					
#	Hits		Visits		URL
1	10811	4.15%	4139	43.01%	/
2	5742	2.21%	2726	28.33%	/secom/fotolegenda.html
3	6927	2.66%	549	5.71%	/secom/
4	1713	0.66%	362	3.76%	/avisos.html
5	492	0.19%	197	2.05%	/classificacao.html
6	408	0.16%	179	1.86%	/Estagio Probatonio/caaep0002.05.html
7	584	0.22%	172	1.79%	/secom/secretarias.html
8	659	0.25%	147	1.53%	/IPTU lei decretos.html
9	495	0.19%	140	1.45%	/secom/links.html
10	458	0.18%	84	0.87%	/secom/fotos.html

Figura 43. Páginas de Saída ao Site no Mês de Abril
Fonte: WEBALIZER (2006)

- h) a palavra principal que os usuários utilizam para encontrar o site da PMC em sites de buscas: já foi definido que o principal modo com que o visitante acessa o site da PMC é digitando o endereço na barra de URL do *browser*. Caso o visitante não saiba qual é o endereço do *site* da PMC, de acordo com as Figuras 44, 45 e 24, o visitante utiliza a palavra “criciuma” nos sites de busca para encontrar a página da PMC.

Top 20 of 38 Total Search Strings				
#	Hits		Search String	
1	19	17.76%	criciuma	
2	13	12.15%	prefeitura de criciuma	
3	8	7.48%	prefeitura de criciu ^{ma}	
4	8	7.48%	prefeitura municipal de criciuma	
5	7	6.54%	crici ^{ma}	
6	6	5.61%	prefeitura municipal de crici ^{ma}	
7	4	3.74%	prefeitura criciuma	
8	3	2.80%	giss sc	
9	2	1.87%	criciuma sc	
10	2	1.87%	criciuma/sc	

Figura 44. Palavras que foram utilizadas para encontrar o site em Março
Fonte: WEBALIZER (2006)

Top 20 of 648 Total Search Strings			
#	Hits		Search String
1	758	21.41%	criciuma
2	437	12.34%	prefeitura municipal de criciuma
3	381	10.76%	prefeitura de criciuma
4	274	7.74%	prefeitura criciuma
5	214	6.04%	criciãma
6	122	3.45%	prefeitura municipal de criciãma
7	82	2.32%	prefeitura de criciãma
8	61	1.72%	prefeitura criciãma
9	54	1.52%	criciuma sc
10	47	1.33%	prefeitura municipal criciuma

Figura 45. Palavras que foram utilizadas para encontrar o *site* em Abril
Fonte: WEBALIZER (2006)

- i) o principal programa de navegação que os usuários utilizam para visualizar o *site* da PMC: De acordo com as Figuras 46, 47 e 48, o *browser* Internet Explorer é o software mais utilizado para visualizar o *site* da PMC. Quando algum novo serviço ser disponibilizado no *site*, pode-se verificar se funcionam nos *browsers* utilizados pelos visitantes.

Top 11 of 50 Total User Agents			
#	Hits		User Agent
1	8014	2.11%	Microsoft Internet Explorer
2	774	0.20%	Netscape/Mozilla
3	18	0.00%	msnbot/1.0 (+http://search.msn.com/msnbot.htm)
4	5	0.00%	mebusca
5	5	0.00%	psbot/0.1 (+http://www.picsearch.com/bot.html)
6	2	0.00%	MSFrontPage/4.0
7	2	0.00%	Microsoft Data Access Internet Publishing Provider Cache Mana
8	2	0.00%	Microsoft Data Access Internet Publishing Provider DAV
9	2	0.00%	Microsoft Data Access Internet Publishing Provider Protocol D
10	1	0.00%	Microsoft Office Protocol Discovery
11	1	0.00%	Yahoo-MMCrawler/3.x(mms dash mncrawler dash support at yahoo

Figura 46. *Browser* utilizado pelos visitantes ao *site* no mês de Março
Fonte: WEBALIZER (2006)

Top 15 of 360 Total User Agents		
#	Hits	User Agent
1	241976	92.99% Microsoft Internet Explorer
2	16173	6.22% Netscape/Mozilla
3	649	0.25% msnbot/1.0 (+http://search.msn.com/msnbot.htm)
4	206	0.08% Googlebot-Image/1.0
5	110	0.04% GbPlugin
6	88	0.03% Yahoo-MMCrawler/3.x (mms dash mmcrawler dash support at yahoo)
7	81	0.03% msnbot/0.9 (+http://search.msn.com/msnbot.htm)
8	64	0.02% oBot
9	50	0.02% Microsoft-WebDAV-MiniRedir/5.1.2600
10	37	0.01% Opera/9.00 (Windows NT 5.1; U; pt-BR)
11	35	0.01% Gigabot/2.0/gigablast.com/spider.html
12	33	0.01% ConveraMultiMediaCrawler/0.1 (+http://www.authoritativeweb.co)
13	31	0.01% MSFrontPage/4.0
14	26	0.01% larbin-2.6.3 cumbuca@codo.softaplic.com.br
15	20	0.01% Microsoft Data Access Internet Publishing Provider Protocol D

Figura 47. *Browser* utilizado pelos visitantes ao *site* no mês de Abril
Fonte: WEBALIZER (2006)

Top 15 of 339 Total User Agents		
#	Hits	User Agent
1	336659	92.22% Microsoft Internet Explorer
2	26648	7.30% Netscape/Mozilla
3	556	0.15% msnbot/1.0 (+http://search.msn.com/msnbot.htm)
4	130	0.04% msnbot/0.9 (+http://search.msn.com/msnbot.htm)
5	108	0.03% GbPlugin
6	100	0.03% LinkWalker
7	91	0.02% Yahoo-MMCrawler/3.x (mms dash mmcrawler dash support at yahoo)
8	67	0.02% Java/1.4.1_04
9	60	0.02% Microsoft-WebDAV-MiniRedir/5.1.2600
10	57	0.02% Googlebot-Image/1.0
11	40	0.01% TurnitinBot/2.0 http://www.turnitin.com/robot/crawlerinfo.htm
12	37	0.01% Opera/9.00 (Windows NT 5.1; U; en)
13	26	0.01% Microsoft Data Access Internet Publishing Provider Protocol D
14	23	0.01% Custom crawler cumbuca@trivialcomsabor.com
15	21	0.01% MSFrontPage/4.0

Figura 48. *Browser* utilizado pelos visitantes ao *site* no mês de Maio
Fonte: WEBALIZER (2006)

- j) o principal sistema operacional e versão que os usuários utilizam para visualizar o *site* da PMC: conforme a Figura 49, o sistema operacional que os visitantes utilizam geralmente é o Windows e de acordo com a Figura 50, a versão mais utilizada pelos visitantes é o Windows XP.






Sistema Operacional (Primeiros 10) - Lista completa/Versões - Desconhecido			
Sistema Operacional		Hits	Por cento
	Windows	332449	94.7 %
	Desconhecido	16731	4.7 %
	Linux	1363	0.3 %
	Macintosh	147	0 %
	OS/2	17	0 %

Figura 49. Sistema Operacional utilizado pelos visitantes em Maio
Fonte: AWSTATS (2006)
















Sistema Operacional			
Versões		Hits	Por cento
Windows		332449	94.7 %
	Windows XP	176864	50.4 %
	Windows NT	1824	0.5 %
	Windows Me	2687	0.7 %
	Windows 98	125033	35.6 %
	Windows 95	1036	0.2 %
	Windows 2003	2784	0.7 %
	Windows 2000	22221	6.3 %
Linux		1363	0.3 %
	Ubuntu	25	0 %
	Fedora	14	0 %
	Debian	310	0 %
	GNU Linux (Unknown or unspecified distribution)	1014	0.2 %
Macintosh		147	0 %
	Mac OS X	109	0 %
	Mac OS	38	0 %
Outros visitantes		16748	4.7 %
	Desconhecido	16731	4.7 %
	OS/2	17	0 %

Figura 50. Versão dos Sistemas Operacionais utilizados em Maio
Fonte: AWSTATS (2006)

- k) o tempo médio das visitas ao site: de acordo com a Figura 51, o tempo médio que um visitante fica no site da PMC é de 7 minutos e 55 segundos.

Duração das visitas		
Número de visitas: 8412 - Med: 475 s	Número de visitas	Por cento
0s-30s	3836	45.6 %
30s-2mn	1993	23.6 %
2mn-5mn	688	8.1 %
5mn-15mn	648	7.7 %
15mn-30mn	361	4.2 %
30mn-1h	463	5.5 %
1h+	423	5 %

Figura 51. Tempo em que um visitante fica no *site* da PMC no mês de Maio
Fonte: AWSTATS (2006)

- l) a média aproximada da quantidade que o servidor Web trafega por mês: conforme a Figura 11, o site da PMC trafegou um total de 8 Gbytes e 100 Mbytes aproximadamente no mês de março, 4 Gbytes e 827 Mbytes aproximadamente em abril e 5 Gbytes e 315 Mbytes aproximadamente no mês de maio. Com esses dados, se tem que o servidor Web trafega uma média aproximada de 6 Gbytes e 64 Mbytes por mês. Caso o servidor Web estiver lento e não for problema de recursos de hardware, pode-se através desta métrica, descobrir se a utilização da banda de Internet não está com sua utilização quase que completa.
- m) a média aproximada mensal da quantidade em que o servidor Web trafega para cada visitante: Visualizando a Figura 11, tem-se a informação que o site da PMC teve 8.968, 10.020 e 10.038 visitantes totais nos meses de março, abril e maio respectivamente. Nesta mesma figura obtém-se a informação que o servidor Web trafegou 8 Gbytes e

100 Mbytes aproximadamente no mês de março, 4 Gbytes e 827 Mbytes aproximadamente em abril e 5 Gbytes e 315 Mbytes no mês de maio. Com esses dados, têm-se a informação de que cada visitante trafega uma média aproximadamente de 660 Kbytes.

- n) os dias da semana em que o site da PMC é mais visitado: Como pode ser visto através da Figura 35, o site da PMC é acessado pelos visitantes em dias úteis da semana. Nos fins de semana o acesso ao site diminui consideravelmente. Através dessa métrica é possível saber quando, por exemplo, que o administrador de rede pode efetuar manutenção no servidor sem fazer com que perca muitas visitas.
- o) quais são os horários em que o site da PMC é mais visitado: conforme mostra a figura 52, o site da PMC é mais visitado no período entre as 8 às 11 horas e das 13 às 17 horas.

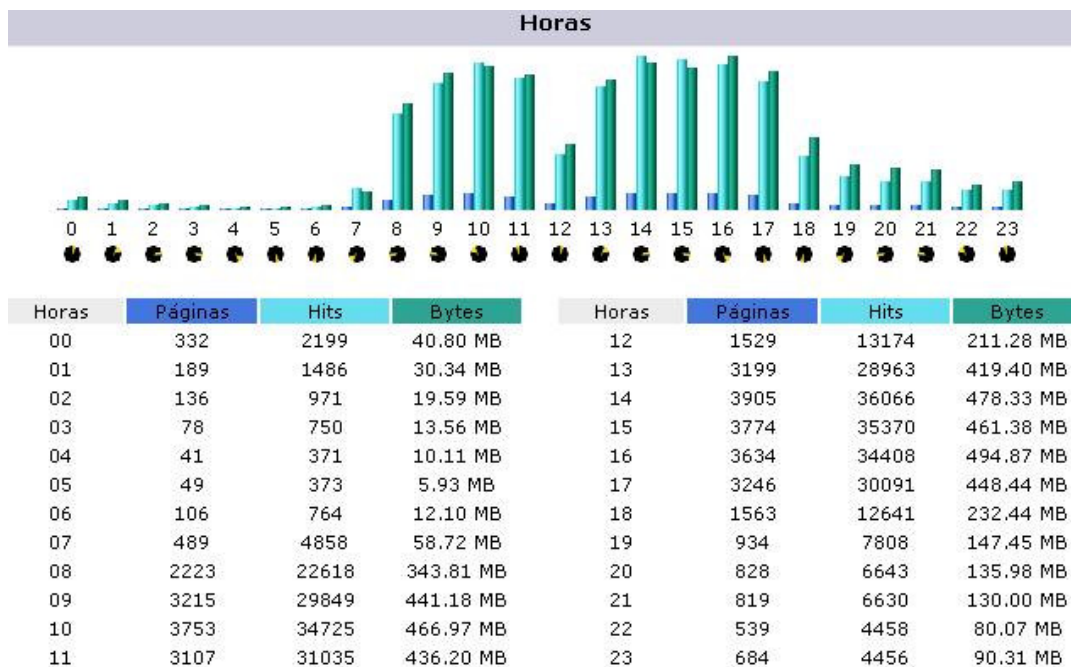


Figura 52. Informações Separadas por Hora do Mês de Maio
Fonte: AWSTATS (2006)

Tendo o resultado das métricas acima, pode-se definir o perfil dos acessos ao *site* da Prefeitura Municipal de Criciúma: o *site* da PMC possui uma média de 9.675 visitas totais por mês. Caso forem apenas visitantes únicos, essa média é de 6.106 visitantes diferentes. Obteve-se também que a quantidade média total de visitas externas que a página da PMC possui é de 9.578. Com essas informações, se tem uma média de 315 visitas por dia e que a maioria dos visitantes digitam diretamente o endereço da página no seu navegador, tendo a primeira página que esses visualizam a página principal do site e a última página que esses acabam visualizando sendo a www.criciuma.sc.gov.br/secom/fotolegenda.html.

Quando um usuário não sabe o endereço do site da PMC e utiliza um *site* de busca para encontrá-lo, é utilizada a palavra “criciuma”. Geralmente esses visitantes utilizam o *browser* da Microsoft, o Internet Explorer e o sistema operacional também da Microsoft: o Windows XP e ficam por volta de 7 minutos e 55 segundos navegando no site.

Esse servidor Web trafega uma média de 6 Gbytes e 64 Mbytes por mês e cada visitante trafega cerca de 660 Kbytes. Sabe-se também que o servidor da prefeitura recebe mais acessos nos dias de semana entre as 8 às 11 horas e 13 às 17 horas.

A Tabela 6 mostra a quais setores ou pessoas responsáveis da PMC serão/ foram encaminhadas às informações obtidas através desse trabalho.

Tabela 6. Para onde serão disseminadas as informações

Métrica	Setor/ Responsável
A	Setor de Informática/ Governo
B	Setor de Informática/ Governo
C	Setor de Informática/ Governo/ Prefeito
D	Setor de Informática/ Governo/ Prefeito
E	Setor de Informática/ <i>Webmaster</i>
F	Setor de Informática/ <i>Webmaster</i>
G	Setor de Informática/ <i>Webmaster</i>
H	Setor de Informática/ <i>Webmaster</i>
I	Setor de Informática/ <i>Webmaster</i>
J	Setor de Informática/ <i>Webmaster</i>
K	Setor de Informática/ Governo
L	Setor de Informática/ Administrador do Servidor
M	Setor de Informática/ Administrador do Servidor
N	Setor de Informática/ Administrador do Servidor
O	Setor de Informática/ Administrador do Servidor

Fonte: CARRADORE. (2006)

As informações obtidas com esse trabalho já estão sendo encaminhadas para os setores ou para as pessoas responsáveis. Com isso, estão se tomando algumas atitudes, como a troca do servidor Web para outro com mais recurso de hardware, a contratação de um funcionário para a criação de um novo *site*, entre outros.

O servidor Web foi trocado devido ao fato que foi descoberto que o *site* possui mais visitas do que era imaginado. Com a quantidade de serviços que o servidor Web estava disponibilizando, a utilização da memória física estava em 100% e esta por sua vez começou a aumentar a frequência de acessos à memória swap³⁴. Como o uso dessa memória é mais lento e espera-se que a nova página da Prefeitura esteja em funcionamento em breve e, ela irá disponibilizar mais serviços que a atual, é esperado que o *site* tenha mais visitas. Com isso, para não ocorrer problemas de desempenho foi trocado o servidor Web para um servidor Xeon de 2.0 GHz, com 2 GB de memória e HD de 60 GB SCSI, utilizando o sistema operacional Suse Linux 10.1 com o apache 2.2.0.

³⁴ É um arquivo que fica no disco rígido. É utilizado pelo sistema operacional quando a memória física do computador se esgota. É mais lenta que a memória física.

CONCLUSÃO

A experiência adquirida ao longo desse trabalho identifica que a análise dos arquivos de *logs* de um servidor Web é uma tarefa imprescindível, pois é possível através dele, obter diversas informações referentes aos acessos dos visitantes aos *sites* hospedados por um servidor deste tipo.

Uma instituição que possui a análise dos acessos ao seu *site* poderá, por exemplo, saber se será necessário realizar algum tipo de investimento futuro no *site*, ou se será necessário desenvolver novos serviços para conseqüentemente atrair mais visitante.

Quando se deseja medir algo, são utilizadas métricas para definir o que se deseja obter para conseqüentemente obter os resultados desejados com essas medições.

O objetivo proposto nesse trabalho, que foi a análise dos arquivos de *logs* do servidor Web da PMC para conseqüentemente definir o perfil dos acessos no mesmo, foi alcançado, pois hoje a Prefeitura está obtendo elementos que informam do perfil de utilização de seu *site*.

Com a conclusão desse trabalho observa-se que ainda é possível no futuro, estudar se o Apache possui perda de desempenho, conforme a quantidade de informações em que ele irá gravar no arquivo de *log*. Poder-se-á também, realizar *data mining* para fazer o cruzamento de informações e obter informações mais detalhadas.

REFERÊNCIAS

A.I.S.A. Aprenda Internet Sozinho Agora. Disponível em:
<<http://www.aisa.com.br/index1.html>>. Acesso em 4 jun. 2006.

ANALOG: WWW logfile analysis. Disponível em: <<http://www.analog.cx/>>. Acesso em 30 mai. 2006.

Assessoria de infra-estrutura e novos investimentos. In: INSTITUTO EUVALDO LODI. **Desburocratização eletrônica nos estados brasileiros**. 2002. Disponível em: <http://www.firjan.org.br/notas/media/Desburocrat_estados.PDF>. Acesso em: 19 de maio 2006.

AWStats – Free log file analyser for advanced statistics (GNU GPL). Disponível em <<http://awstats.sourceforge.net/>>. Acesso em 15 abr. de 2006.

DEMÉTRIO, Rinaldo. **Internet**. São Paulo: Érica, 2001. 276 p.
EBAVE/FGV – *E-Government*. Disponível em:
<http://www.ebape.fgv.br/e_government/asp/dsp_oquee.asp>. Acesso em 13 nov. 2005.

Felipini, Dailton. Segurança na Internet. **Net**. São Paulo, [200-?]. Disponível em:
<http://www.e-commerce.org.br/Artigos/seguranca_internet.htm>. Acesso em: 11 mar. 2006.

FERRER, Florência; SANTOS, Paula. ***E-government*: o governo eletrônico no Brasil**. São Paulo: Saraiva, 2004. 258 p.

FREITAS, Rogério Afonso (...[et al.]). **Portais Corporativos**. Rio de Janeiro: Brasport, 2004. 128 p.

FRIEDLEIN, Ashley. **Como gerenciar sites Web de sucesso**. Rio de Janeiro: Campus, 2003. 412 p.

Guia do Hardware. Disponível em <<http://www.guiadohardware.net/tutoriais/112/>>. Acesso em 27 mai. 2006.

Governo Eletrônico. Disponível em: <<http://www.governoeletronico.gov.br/>> Acesso em 10 nov. 2005.

HALLAM-BAKER, Phillip; BEHLENDORF, Brian. Extended Log File Format. **Net**, Cambridge - USA, [199-?]. Disponível em: <<http://www.w3.org/TR/WD-logfile.html>>. Acesso em: 12 jun. 2006.

HEIDE, Ann; STILBORNE, Linda. **Guia do professor para a Internet: completo e fácil**. 2. ed. Porto Alegre: ArtMed, 2000. 337 p.

Home of The Webalizer. Disponível em <<http://www.mrunix.net/webalizer/>>. Acesso em 1 jun. 2006.

HTTP-ANALYZE – A Logfile analyzer for web servers. Disponível em <<http://http-analyze.org/>>. Acesso em 1 jun. 2006.

HUNT, Craig. **Servidores de redes com linux: o recurso essencial para administradores de sistemas**. São Paulo: Market books, 2000. 715 p.

JARVIS, Mark. Comércio Eletrônico é mais seguro que comprar na rua. **Net**. São Paulo, [200-?]. Disponível em: <http://www.e-commerce.org.br/Artigos/comercio_eletronico_seguran%E7a.htm>. Acesso em: 11 mar. 2006.

KNIGHT, Peter T. PPPs e Infra-estrutura de e-Governo. **Net**, Rio de Janeiro, fev. 2005. Disponível em: <<http://www.tedbr.com/publicacoes/bancohoje02-05.htm>>. Acesso em: 8 nov. 2005.

Log Files – Apache HTTP Server. Disponível em <<http://httpd.apache.org/docs/1.3/logs.html>>. Acesso em 22 nov. 2005.

MELLO, Leonor Queiroz e. A importância da Internet no marketing dos produtos tradicionais de qualidade. **Net**. Lisboa, jul. 2004. Disponível em: <<http://www.agroportal.pt/a/2004/lmello.htm>>. Acesso em: 15 mai. 2006.

NAKAMURA, Rodolfo Reijiro. **E-commerce na internet: fácil de entender**. São Paulo: Érica, 2001. 240 p.

NEMETH, Evi (...[et al.]). **Manual do administrador do sistema Unix**. 3.ed Porto Alegre: Bookman, 2002. 895 p.

NORTHCUTT, Stephen (...[et al.]). **Desvendando: segurança em redes**. Rio de Janeiro: Campus, 2002. 650 p.

Núcleo de Informação e Coordenação. Disponível em: <<http://www.nic.br>>. Acesso em 15 mai. 2006.

Prefeitura Municipal de Criciúma. Disponível em <<http://www.criciuma.sc.gov.br>>. Acesso em 13 abr. 2006.

REXFORD, Jennifer; KRISHNAMURTHY, Balachander. **Redes para a web**. Rio de Janeiro: Ed. Campus, 2001. 651 p.

SIQUEIRA, Ethevaldo. Um retrato do *e-government* no Brasil. **Net**, São Paulo, ago. 2004. Convergência Digital. Disponível em: <<http://www.estadao.com.br/tecnologia/coluna/ethevaldo/2004/ago/09/119.htm>>. Acesso em: 10 nov. 2005.

TEIXEIRA JUNIOR, José Helvécio. **Redes de computadores: serviços, administração e segurança**. São Paulo: Makron Books, 1999. 493 p.

The Apache Software Foundation. Disponível em <<http://apache.org>>. Acesso em 4 jun. 2006.

WADLOW, Thomas A. **Segurança de redes: projeto e gerenciamento de redes seguras**. Rio de Janeiro: Campus, 2000. 269 p.

ZELENOVSKY, Ricardo; MENDONÇA, Alexandre. Funcionamento de Modems. **Net**, Rio de Janeiro, jul. 2001. Disponível em: <<http://www.clubedohardware.com.br/artigos/373>>. Acesso em: 11 jun. 2006.

APÊNDICE A – INSTALAÇÃO E CONFIGURAÇÃO DO APACHE

O servidor Web Apache geralmente está incluído nas distribuições Linux. Quando é realizada a instalação do sistema operacional, o administrador tem a possibilidade de escolher se deseja instalar esse pacote junto com o Linux. Muitas pessoas que instalam o Linux pela primeira vez podem já o ter instalado sem saber, pois ele é instalado no perfil padrão. O nome que o Apache é reconhecido no Linux é o httpd. Para descobrir se o Apache já está instalado e funcionando, basta ir à linha de comando e digitar:

```
# ps ax | grep httpd
```

Se o Apache estiver instalado e funcionando, uma lista de processos aparecerá na tela. Com isso, basta você carregar a interface gráfica, abrir um navegador Web e colocar o seguinte endereço:

```
http://localhost/
```

Você verá a seguinte página da Figura 53.

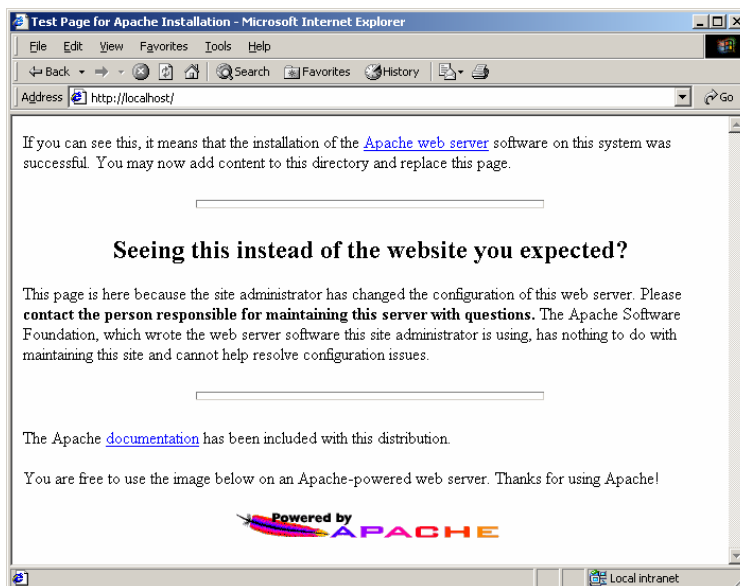


Figura 53. Página inicial do servidor Web Apache
Fonte: THE APACHE SOFTWARE FOUNDATION (2006)

Isso comprova que o servidor está em funcionamento.

Caso nenhum processo aparecer, então você terá que instalar o Apache. Para isso, visite o *site* <http://httpd.apache.org> e faça o *download* da última versão do servidor. É necessário fazer o *download* da versão para compilar, ou seja, a versão com a extensão *.tar.gz*.

Após fazer o *download*, é necessário voltar para a linha de comando. Utilizando um exemplo de que a instalação do Apache foi salva no diretório */root/* e que o arquivo de instalação for, por exemplo, *httpd-2.2.2.tar.gz*, caso o usuário não estiver utilizando o *root*, digitar:

```
# su
```

Após esse comando, o usuário deverá entrar com a senha de administrador. Depois de realizar o acesso, digitar os seguintes comandos:

```
# cd /root
```

```
# tar -xvzf httpd-2.2.2.tar.gz
```

```
# cd httpd-2.2.2
```

```
# ./configure
```

```
# make
```

```
# make install
```

Depois desses comandos, o apache estará instalado em */usr/local/apache2/*.

Para iniciar o serviço, digite os seguintes comandos:

```
# cd /usr/local/apache2/bin
```

```
# ./httpd
```

Com isso, basta voltar ao *browser* e digitar o endereço <http://localhost> que irá aparecer a página padrão do Apache.

Para um servidor Web padrão, o apache já esta funcionando corretamente, mais é necessário editar algumas opções do arquivo *httpd.conf* que fica no diretório

/usr/local/apache2/conf/. Lembrando que antes de editar, é recomendável fazer um *backup* desse arquivo:

ServerAdmin: Nessa opção, coloca-se o e-mail do administrador do servidor Web.

ServerName: Aqui, é necessário colocar o nome do *host*, como por exemplo: `www.criciuma.sc.gov.br`.

Port: Define qual porta o servidor Web irá utilizar. Por padrão ele utiliza a porta 80.

MaxClients: Define o número máximo de clientes que podem conectar ao mesmo tempo no servidor.

HostnameLookup: Aqui é definido se o servidor deve ou não traduzir os endereços IP para o nome de *hosts* (*On/Off*).

DocumentRoot: Define onde ficarão os arquivos de páginas.

ErrorLog: Define onde ficará o arquivo de *log* que registrará os erros.

CustomLog: Define onde ficará o arquivo de *log* dos acessos ao servidor Web.

LogLevel: É recomendado definir essa opção como *debug*, pois com essa opção, será informado os erros e todos os tipos de *logging* possíveis.

LogFormat: Definem o formato das entradas do arquivo de *log*. Essa opção deve ser definida da seguinte maneira para informar o máximo de *logging* possível:

```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\""  
combined
```

```
LogFormat "%h %l %u %t \"%r\" %>s %b" common
```

```
LogFormat "%{Referer}i -> %U" referer
```

```
LogFormat "%{User-agent}i" agent
```

BindAddress: Define em qual IP e porta que o servidor Web irá monitorar para responder as requisições.

Depois de editar essas opções, digite os seguintes comandos:

```
# cd /usr/local/apache2/bin
```

```
# ./http -k restart
```

O servidor Web Apache irá reiniciar com as opções definidas no arquivo de configuração.

É possível editar outras configurações, mais a ênfase desse apêndice é apenas fazer com que o servidor Web esteja em funcionamento e que envie para o arquivo de *log*, as maiores informações possíveis.