

Panorama setorial da Internet

AGOSTO 2010

cetic.br

cgi.br
nic.br

registro.br

cert.br

cetic.br

ceptro.br

W3C WORLD WIDE WEB
Escritório Brasil

Dimensões e características da Web brasileira: uma visão do .gov.br

Ano 2 – Número 3

A Internet e seus impactos

A Internet é uma rede mundial de computadores . A *Web* é uma de suas muitas aplicações, com certeza a maior e a mais conhecida.

A *Web* é, também, a grande responsável pela popularização da Internet, a ponto de hoje ser confundida com esta.

É flagrante o avanço do uso da *Web* pela população brasileira: de 37 milhões de usuários, em 2005, passou para aproximadamente 63 milhões, em 2009.

Dimensões e características da *Web* brasileira: uma visão do .gov.br

A Internet é provavelmente a mais sofisticada tecnologia de informação e comunicação atualmente disponível para a sociedade, principalmente em função da sua forma de organização e de seus impactos nas esferas tecnológicas, social, econômica e política.



Ela é também a infraestrutura necessária para uma de suas maiores e mais conhecida aplicação: a *Web*, grande responsável pela popularização da Internet, a ponto de hoje ser confundida com esta. Internet e *Web* são, portanto, conceitos distintos. A *Web* pode ser definida, grosso modo, como a parte da Internet acessada por meio de navegadores, ou *browsers*.

O impacto do uso da Internet e da *Web* na sociedade, indivíduos e organizações tornou-se objeto de pesquisa, extrapolando o campo especializado da computação aplicada, e atingindo áreas de estudos organizacionais e sociológicos. Por ser essencialmente dinâmica e sem fronteiras, tanto do ponto de vista físico como do virtual, é importante que seja conhecida em detalhes, tanto para assegurar sua livre transformação quanto para permitir sua disponibilidade, confiabilidade e acessibilidade por todos.

Assim, o Comitê Gestor de Internet do Brasil – CGI.br e o Núcleo de Informação e Coordenação do Ponto BR – NIC.br, por meio do W3C Brasil (Consórcio *World Wide Web*) e do Centro de Estudos e Pesquisas em Tecnologias de Redes e Operações – CEPTRON.br, criaram mais uma iniciativa para um melhor conhecimento e entendimento da Internet brasileira: o **Projeto Censo da *Web* .br**. Realizado em parceria com a Secretaria de Logística e Tecnologia da Informação do Ministério do Planejamento, Orçamento e Gestão (SLTI / MPOG), a Associação Brasileira de Entidades Estaduais de Tecnologia da Informação e Comunicação (ABEP) e o Instituto Nacional de Ciência e Tecnologia para a *Web* (INCT *Web*), ainda com o apoio metodológico do Centro de Estudos sobre as Tecnologias de Informação e Comunicação – CETIC.br, esse projeto tem como objetivo realizar estudos sobre a *Web* brasileira, cujo escopo é definido mais adiante.

Desde meados dos anos 90, a *Web* brasileira tem mostrado acentuado crescimento, tanto no número de usuários como no leque de serviços e aplicações oferecidos por meio da rede. É flagrante o avanço de seu uso pela população brasileira: de 37 milhões de usuários, em 2005, passou para aproximadamente 63 milhões, em 2009. Igualmente impressionante é a mudança de comportamento do cidadão, que utiliza cada vez mais serviços transacionais em ambientes virtuais, conforme mostram as pesquisas do CGI.br.

Primeiros passos

Como primeira etapa do projeto, foi realizada uma coleta e análise de dados da *Web* brasileira governamental para estudos preliminares. Os resultados desses estudos ajudarão a responder várias questões inerentes ao desafio de medir o tamanho da *Web*.

Quantos sites há na *Web.br*? Qual o tamanho da *Web.br*, e como se dá seu crescimento? Que tipo de tecnologias são utilizadas? Onde os sítios *Web* estão hospedados? No Brasil ou no exterior?

Dimensões e características da *Web* brasileira: uma visão do .gov.br

Para compreender o fenômeno do desenvolvimento da *Web* brasileira, esta edição do **Panorama Setorial da Internet** apresenta os primeiros resultados do Projeto Censo da *Web* para os sites sob o domínio .GOV.BR. Também são apresentados os aspectos conceituais e metodológicos que fundamentam o projeto.

Como primeira etapa do projeto, foi realizada uma coleta e análise de dados da *Web* brasileira governamental, que servirá de base de validação da coleta, dos mecanismos de análise e da metodologia. Esse estudo ajudará a responder várias questões, complementando e servindo de subsídio para outras ações. Por exemplo: Quantos *sites* há na *Web.br*? Qual o tamanho da *Web.br*, e como se dá seu crescimento? Que tipo de tecnologias são utilizadas? Onde os sítios *Web* estão hospedados? No Brasil ou no exterior? Os sítios são aderentes aos padrões *Web*, como HTML e CSS? Os sítios são acessíveis? Há suporte a IPv6? Quais tecnologias são usadas para os servidores, páginas, imagens, documentos, vídeos etc? Os servidores mantêm seus relógios sincronizados com a Hora Legal Brasileira?

Antes de responder a estas perguntas, é necessário introduzir ao leitor alguns conceitos, algumas definições e também algumas ressalvas.



Definindo o conceito de *Web*

A *World Wide Web*, também conhecida como *Web*, ou simplesmente WWW, é um gigantesco acervo universal de páginas, documentos, dados, aplicações e serviços interligados por meio da rede mundial de computadores, disponibilizado às pessoas de qualquer lugar do globo, a qualquer momento e por diversos dispositivos, desde computadores até aparelhos móveis, como telefones celulares. Esse imenso acervo pode reunir diversos tipos de conteúdos digitais, desde páginas de hipertextos, até arquivos no formato de imagens, figuras, som, vídeos, e códigos de programação, dentre outros. Todo arquivo disponível na *Web* é identificado por um endereço único e exclusivo, chamado de URL, que significa *Uniform Resource Locator*, em português Localizador Padrão de Recursos. Uma URL indica o local onde se localiza o arquivo digital na *Web*. Essa foi a grande invenção de Tim Berners-Lee, que, ao criar todo um sistema de localização na *Web*, possibilitou que os documentos pudessem ser acessíveis em qualquer lugar do globo.

Definindo o WWW

A *World Wide Web*, também conhecida como *Web*, ou simplesmente *WWW*, é um gigantesco acervo universal de páginas, documentos, dados, aplicações e serviços interligados por meio da rede mundial de computadores.

Todo arquivo disponível na Web é identificado por um endereço único e exclusivo, chamado de URL, que significa Uniform Resource Locator, em português Localizador Padrão de Recursos.

Definindo o conceito de Web

Cada um destes acervos é identificado por um nome ou domínio, comumente conhecido por *Website*, site, ou sítio Web. Toda vez que navegamos na Web, digitamos esses nomes para acessarmos os *sites* que desejamos, por exemplo: www.cgi.br, www.google.com.br, www.receita.fazenda.gov.br. É importante notar que a Web, embora seja uma aplicação poderosa e de ampla utilização, é apenas uma parte da rede, uma aplicação.

Os nomes de domínios também têm sua organização própria, não sendo de uso exclusivo da Web, e estão estruturados globalmente em níveis hierárquicos. Os domínios de primeiro nível são chamados de TLDs, acrônimo de *Top Level Domains* e existem diversos tipos, por exemplo o “.NET”, o “.COM”, “.ORG”, entre outros.

Os domínios que identificam um determinado país de origem, como o .br, são chamados de código de país ou *Country Code*. Assim, o .br é um *Country Code Top Level Domain* – ccTLD, o domínio de primeiro nível do Brasil. Para o nosso caso brasileiro, abaixo desse domínio de primeiro nível existem outros níveis, como o “.GOV.BR”, o “.COM.BR”, o “.ORG.BR”. Somente abaixo desses domínios, e seguindo esse esquema hierárquico, os diversos domínios são registrados e criados, por exemplo o domínio “governoeletronico.gov.br”. O correto entendimento dessa estrutura de domínios é importante para também compreender as análises expostas nesse documento.



As dimensões e as características da Web brasileira

Para fins de determinação do escopo, a *Web* brasileira é definida no contexto do **Projeto Censo da Web.br** como a rede formada pelos *sites* de acesso público identificados por um domínio .br, mais os *sites* para os quais há redirecionamentos diretos, via servidor, a partir de um *site.br*, dos quais analisa-se apenas a página principal.

De acordo com dados do *Registro.br*, autoridade de registro para nomes de domínio no Brasil, o mês de julho de 2010 foi encerrado com cerca de 2,17 milhões de nomes de domínios registrados sob esse respectivo ccTLD, os quais contêm *sites* das mais variadas instituições privadas, governamentais, instituições de ensino, organizações do terceiro setor, profissionais liberais, pessoas físicas, etc. Isso exige a realização de estudos sobre universos específicos de nomes de domínios, por exemplo “.COM.BR”, “.ORG.BR”, “.NET.BR”, “.GOV.BR” e outros grupos menores, com o objetivo de medir as características da *Web*.

Coleta de dados

A coleta de dados sobre os domínios do governo foi realizada em outubro de 2009 e fevereiro de 2010. Identificou-se um total de 18.796 *sites* sob o .GOV.BR, a partir de URLs percorridas.

Como ponto de partida para um levantamento mais amplo das dimensões e características do .br, optou-se por uma coleta exaustiva da *Web* governamental brasileira, aquela constante nos *sites* e páginas sob o domínio .GOV.BR.

Os principais resultados e algumas conclusões desse levantamento são apresentados nesta edição do **Panorama Setorial da Internet**. Esse primeiro estudo será de grande utilidade e subsidiará o planejamento de uma coleta mais ampla e detalhada das dimensões de toda a *Web* brasileira sob o ccTLD .br.

.GOV.BR

As dimensões e as características do .gov.br

A coleta de dados sobre os domínios do governo foi realizada em outubro de 2009 e identificou um total de 18.796 *sites* sob o .GOV.BR, a partir de URLs percorridas. A identificação do total de *sites* partiu de dados fornecidos das seguintes fontes:

- Domínios identificados como .GOV.BR (domínios reservados ao Governo Federal), cuja lista foi fornecida pela autoridade de registro para nomes de domínio no Brasil, o Registro.br, com autorização do Ministério do Planejamento, responsável pelo uso dos domínios sob o .GOV.BR.
- Domínios identificados como *sigla-UF.GOV.BR*, registrados pelas empresas estaduais de processamento de dados, vinculadas aos governos estaduais;
- Resultados de consultas e buscas de informações, utilizando ferramentas de busca, com o objetivo de complementar as informações anteriores.

Essas diferentes fontes foram unificadas e serviram como semente para um sistema coletor. Objetivou-se com esse levantamento produzir um cadastro que pudesse contemplar o maior número possível de *sites* governamentais, de tal modo que fosse o mais próximo de um censo da *Web* governamental brasileira. Porém, nem todas as empresas de processamento de dados das unidades das federações e responsáveis pelos registros dos domínios *sigla-UF.GOV.BR* puderam responder em tempo, fato que introduziu mais uma dificuldade para a realização de um censo da *Web* governamental, além daquelas inerentes a própria *Web*.

Ponto de partida

Como ponto de partida para um levantamento mais amplo das dimensões e características do .br, optou-se por uma coleta exaustiva da Web governamental brasileira, aquela constante nos sites e páginas sob o domínio .GOV.BR.

A partir dos resultados da coleta, foi investigada a participação de cada uma das cinco regiões brasileiras e também a do Governo Federal na composição da Web a partir da análise dos subdomínios correspondentes aos estados, por exemplo o subdomínio .SP.GOV.BR foi considerado pertencente a região sudeste, e o .GOV.BR como do governo federal. Avaliaram-se dois aspectos dessa participação:

- ❑ O número total de *sites* correspondentes a cada uma das cinco regiões do país e do governo federal;
- ❑ A quantidade total de *bytes* por região e *Governo Federal*, considerando-se apenas os documentos HTML (isto é, imagens, vídeos e outros tipos de arquivos não foram contabilizados na tabela abaixo. Vide Figura 12 para uma análise de outros tipos de arquivos).

A Tabela 1 apresenta um sumário do número de sites e da quantidade de bytes coletados para cada uma das grandes regiões brasileiras. Embora a coleta tenha chegado a um total de 18.796 sites, foram considerados apenas os que continham pelo menos um documento HTML nessa análise. Os sites satisfazendo essas condições totalizam 11.856.

REGIÃO	Volume em Gigabytes	Número total de sites	participação no tamanho total em bytes	participação no número total de sites do .GOV.BR
SUL	26	3.416	18%	29%
SUDESTE	32	3.358	22%	28%
NORTE	7	816	5%	7%
NORDESTE	27	1.786	18%	15%
GOV.BR	38	1.668	26%	14%
CENTROOESTE	17	812	11%	7%
Total	148	11.856	100%	100%

Tabela 1 – Quantidade e tamanhos de Sites por região geográfica



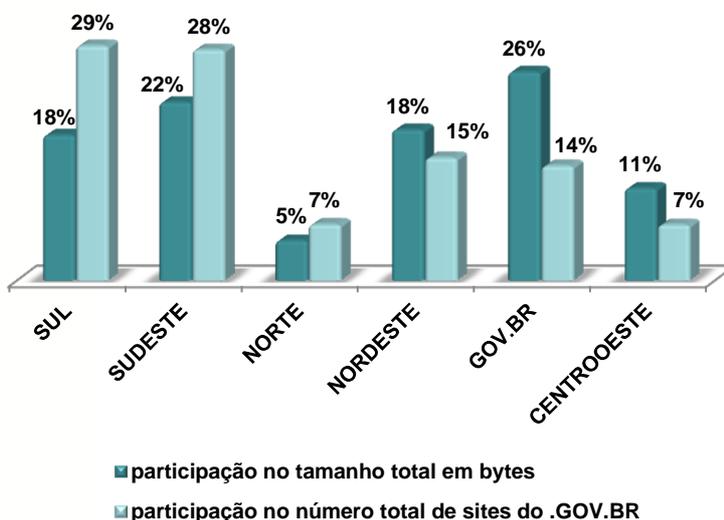
Participação das UFs e regiões na composição da Web brasileira de governo

Dimensões do .GOV.BR

A maior participação em relação ao tamanho em *bytes* na Web de governo foi observada no agrupamento composto pelos *sites* Governo Federal, 26%. Em número absoluto de *sites*, a região Sul apresentou 33% dos 18,7 mil *sites* coletados.

A distribuição percentual dos dados apresentados na Tabela 1 pode ser analisada no gráfico apresentado na Figura 1.

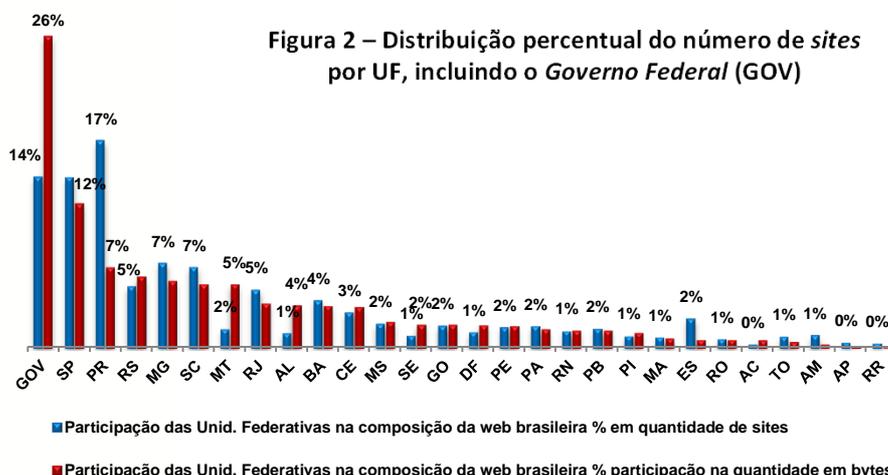
Figura 1 – Participação das regiões e do governo federal na composição da Web de governo



Observou-se uma maior participação em relação ao tamanho em *bytes* na Web de governo no agrupamento composto pelos *sites* do Governo Federal, 26%. Em relação ao número absoluto de *sites*, a região Sul apresentou 33% dos 18,7 mil *sites* coletados.

Em relação à participação das unidades da federação na composição da Web brasileira de governo, o domínio *pr.gov.br*, pertencente ao Estado do Paraná, foi o que apresentou a maior participação em número absoluto de *sites*, cerca de 17%, de todos os *sites* brasileiros de governo coletados, conforme mostrado na Figura 2. O Governo Federal representado pelos *sites* com domínio *.GOV.BR* vem em segundo lugar, empatado com o Estado de São Paulo (*sp.gov.br*). Estes dois últimos participam, cada um, com 14% dos *sites* sob a Web governamental brasileira.

Figura 2 – Distribuição percentual do número de sites por UF, incluindo o Governo Federal (GOV)



Participação das UF's e regiões na composição da Web brasileira de governo

A Web nas UF's

Em relação ao tamanho do conteúdo em *bytes* dos *sites* brasileiros de governo, o Governo Federal é o que apresenta o maior índice, com 26% do total verificado, seguido pelo Estado de São Paulo, com 12%.

Em relação ao tamanho do conteúdo em *bytes* dos *sites* brasileiros de governo, o *Governo Federal* é o que apresenta o maior índice, com 26% do total verificado, seguido pelo Estado de São Paulo, com 12%. Os demais estados apresentam índice inferior a 10% do total de quantidade em *bytes*.

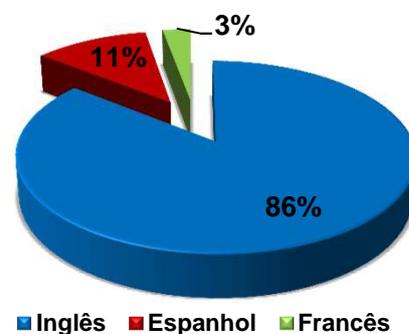
A relação entre o número de *sites* ou eventualmente o número de páginas e a quantidade em *bytes* pode também ser uma abordagem de análise a ser considerada no futuro. Observando-se a Figura 2, verifica-se que os estados, em geral, guardam a mesma participação percentual em relação ao total tanto em número de *bytes* de seus *sites* como na quantidade de *sites* na Web governamental brasileira, indicando que possuem um tamanho médio de *sites* em *bytes* equivalentes. Poucos estados fogem dessa regra. De um lado, os domínios do *Governo Federal* apresentam maior conteúdo em *bytes* em relação aos seus *sites*. De outro, o Estado do Paraná (PR) que apresenta maior quantidade de *sites* em números absolutos e relativos à participação no total da Web governamental brasileira. Investigar as razões pode mais do que demonstrar existir quantidade de domínios governamentais na Web sem conteúdos significativos, ou ainda domínios governamentais com conteúdo excessivo, revelando uma dificuldade no acesso à informação relevante e pública.

Outros idiomas na Web brasileira de governo

Das 3.182.202 páginas que puderam ter seu idioma identificado através de análise automatizada, 97% destas estão em português. O *software* que foi usado para identificar a linguagem faz a comparação do texto contido nas páginas com dicionários com palavras-chave dos idiomas português, inglês, espanhol e francês, contabilizando as palavras que aparecem num determinado documento. Caso o número de palavras-chave de um dos idiomas testados ultrapasse um determinado limite inferior, e não haja ambigüidade (mais de um idioma com palavras-chave suficientes), ele é considerado como identificado. Os resultados para os idiomas estrangeiros estão ilustrados na Figura 3.

É importante lembrar ainda que em 47% das 6,3 milhões de páginas em HTML coletadas, o idioma não pode ser identificado por meio do procedimento utilizado. A amostra é significativa e pode revelar que essencialmente os conteúdos da Web governamental brasileira visam atender ao público interno, tendo pouca quantidade de informação em outros idiomas.

Figura 3 – Outros idiomas encontrados na Web brasileira de governo



Por outro lado, sem uma abordagem semântica não é possível verificar se informações relevantes em outros idiomas são oportunas, sejam, por exemplo, para conteúdo de relações exteriores ou para os fins de turismo. Identificar qual informação pode e convém estar em outro idioma não foi ainda objeto de análise.

Aderência aos padrões HTML do W3C

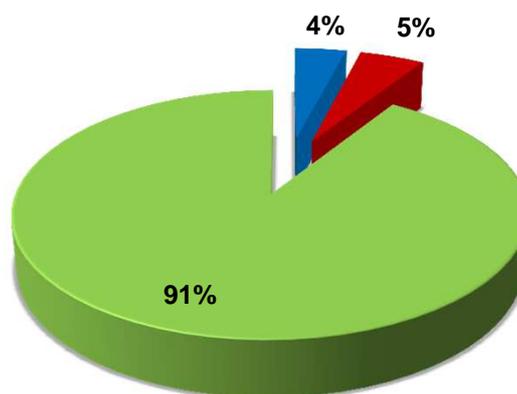
Aderência das páginas

Dos 6,3 milhões de páginas HTML coletadas, cerca de 91% apresentaram não-conformidade aos padrões do W3C, apenas 5% estão completamente de acordo com o padrão e 4% não puderam ser avaliadas.

A avaliação da aderência das páginas HTML aos padrões do W3C foi feita através da aplicação de um *software* validador projetado pelo próprio consórcio. Tal como propõe o W3C e as boas práticas de desenvolvimento *Web*, a aderência aos padrões *Web* é indicador importante da universalidade de acesso por qualquer dispositivo conectado à *Web*, bem como por qualquer ambiente operacional.

Quanto mais aderente aos padrões, melhor a página será acessada por qualquer usuário, independente do dispositivo e de seu ambiente operacional. Por outro lado, páginas não aderentes terão acessos restritos a alguns dispositivos ou sistemas operacionais, donde pressupõe seu caráter de não universalidade. Considera-se que principalmente para conteúdos da *Web* governamental a aderência aos padrões e a universalidade do acesso devem ser constantemente considerados e exigidos.

Figura 4 – Aderência aos padrões HTML do W3C



■ Não puderam ser avaliadas ■ Nenhum erro ■ Mais de 1 erro

Para essa análise, verificou-se a contagem do número de incorreções de acordo com o padrão encontrado pelo *software* validador. Dos 6,3 milhões de páginas HTML coletadas, cerca de 91% apresentaram mais de uma incorreção de aderência, apenas 5% estão completamente de acordo com o padrão, e 4% não puderam ser avaliadas, conforme mostra a Figura 4.

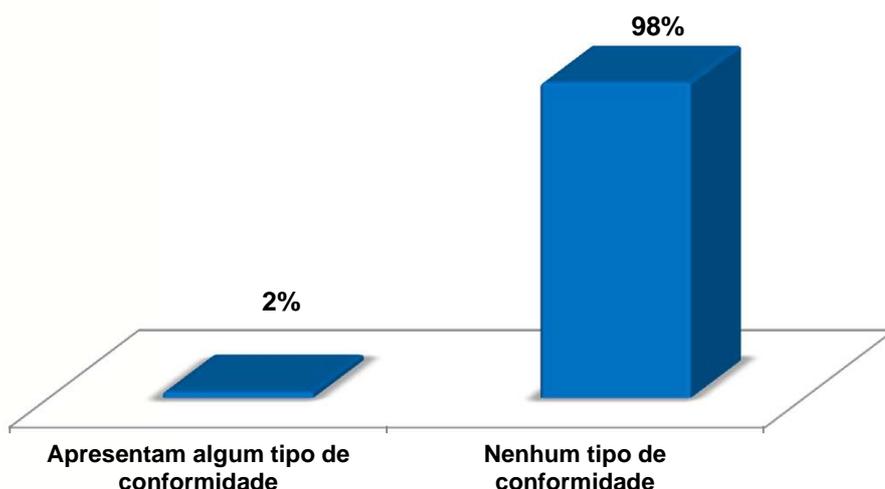
Dos 6,3 milhões de páginas HTML coletadas, 98% não apresentaram nenhuma aderência aos padrões de acessibilidade

Aderência aos padrões de acessibilidade ASES

A aderência a padrões de acessibilidade visa garantir o acesso universal aos *sites Web*, mesmo para portadores de deficiência. Os critérios de acessibilidade são separados em 3 níveis de acessibilidade ou conformidade, definidos pelo padrão WCAG do W3C. O nível de conformidade A é considerado mandatório para que um *site* seja considerado acessível. O nível de conformidade AA consiste em práticas de que deveriam ser seguidas, indo além das mais básicas, e o nível de conformidade AAA, em práticas opcionais, mas que melhorariam ainda mais a acessibilidade do *site*.

O governo brasileiro criou o e-MAG – Modelo de Acessibilidade de Governo Eletrônico, que segue os padrões internacionais e consiste em um conjunto de recomendações a ser considerado para que o processo de acessibilidade dos sítios e portais do governo brasileiro seja conduzido de forma padronizada e de fácil implementação. Criou ainda o ASES, um *software* que auxilia o desenvolvedor *Web* na construção de *sites* acessíveis, em conformidade com o e-MAG.

Figura 5 – Proporção de páginas aderentes aos padrões de acessibilidade ASES



Fez-se a avaliação da aderência das páginas HTML coletadas aos padrões de acessibilidade através dos mesmos testes utilizados no ASES. O processo de avaliação consiste da contagem de conformidades das páginas.

Dos 6,3 milhões de páginas HTML coletadas, 98% não apresentaram nenhuma aderência aos padrões de acessibilidade conforme mostrado na Figura 5.

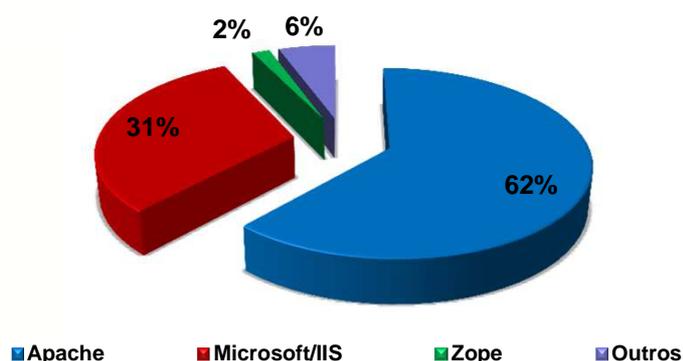
Tecnologias utilizadas pelos servidores Web governamental

Uso das tecnologias baseadas em código aberto

As tecnologias baseadas em *software* de código aberto foram encontradas em mais de 60% das páginas coletadas. Plataformas proprietárias ocupam pouco mais de 30% da fatia de servidores Web de governo.

Um servidor Web pode ser definido como todo o computador dedicado, totalmente ou em parte, a disponibilizar e gerenciar o acesso a conteúdos digitais. Em geral, fazem uso de *software* específico, projetado exclusivamente para desempenhar de forma otimizada as suas funções. Dentre os resultados obtidos da coleta de dados do .GOV.BR, pode-se destacar o mapeamento das tecnologias de disponibilização e armazenagem de informações. A seguir, apresenta-se o gráfico relativo à participação das principais tecnologias usadas em servidores Web do .GOV.BR(Figura 6).

Figura 6 – Tecnologias utilizadas pelos servidores Web

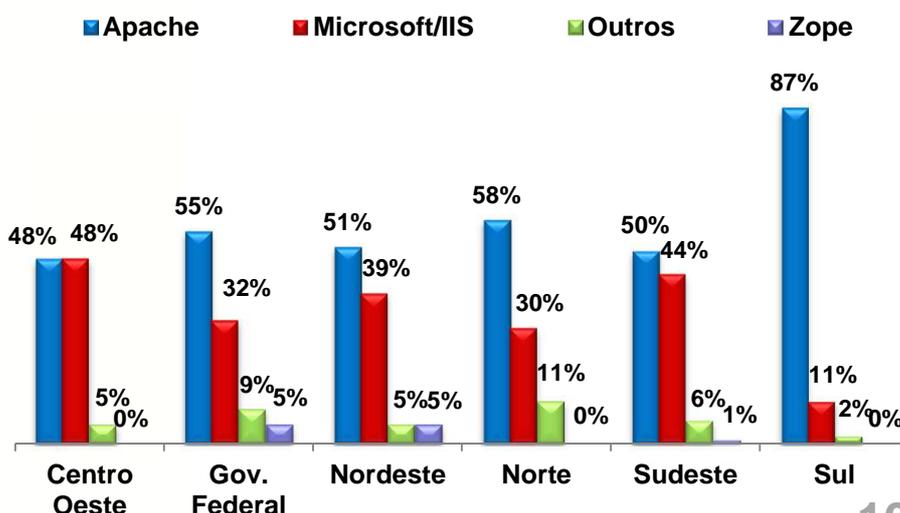


As tecnologias baseadas em *software* de código aberto foram encontradas em mais de 60% dos sites coletados. Plataformas proprietárias ocupam pouco mais de 30% da fatia de sistemas servidores de conteúdo na Web brasileira de governo.

Tecnologias utilizadas pelos servidores Web nas cinco regiões brasileiras

Conforme ilustrado no gráfico a seguir, as páginas coletadas sob subdomínios relativos a unidades federativas da região Sul apresentam a maior incidência de servidores Web baseados em sistemas de código aberto, e também o menor percentual de uso de sistemas proprietários, considerando inclusive as páginas do Governo Federal.

Figura 7 – Tecnologias utilizadas pelos servidores Web por região



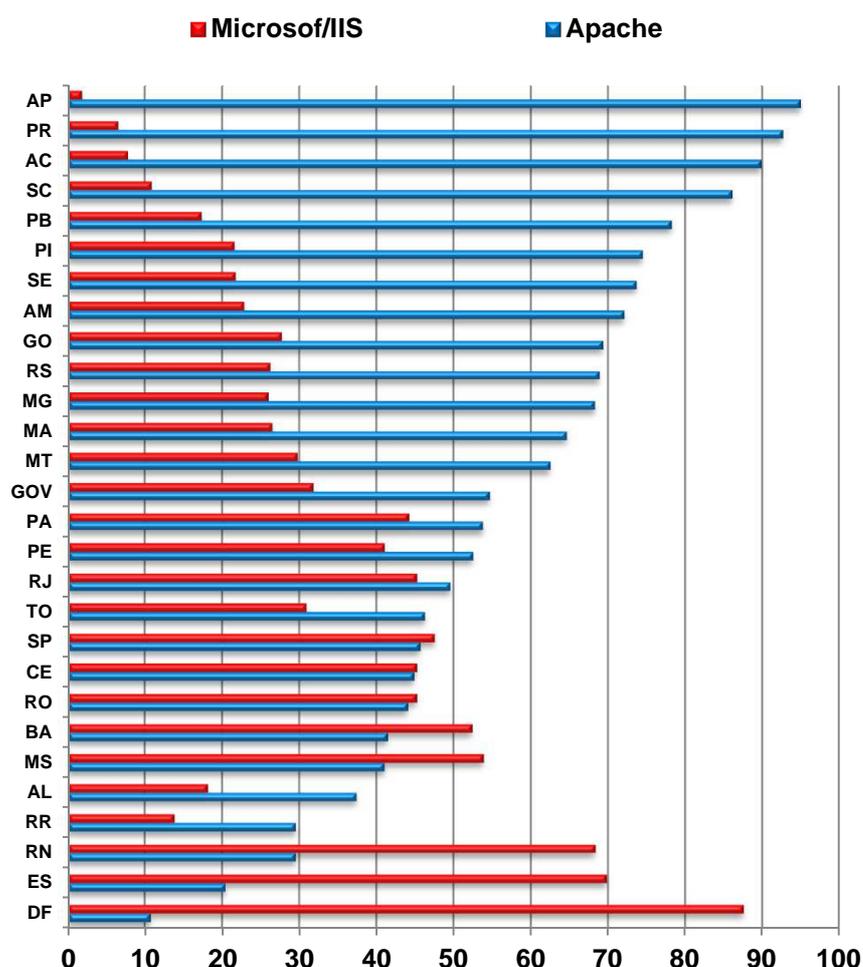
As tecnologias utilizadas pelos servidores Web nas UFs

Tecnologias nas UFs

O Amapá e o Paraná são os primeiros colocados em uso relativo de sistema de código aberto para servir conteúdo na Web.

Também foi verificada a utilização do tipo de plataforma servidora por unidade da federação. O gráfico da Figura 8 apresenta o uso do tipo de plataforma servidora de Web, relativo ao total de sites daquela UF. O Amapá e o Paraná são os primeiros colocados em uso relativo de sistema de código aberto para servir conteúdo na Web. Em relação ao uso de software proprietários, verifica-se que o DF é o estado que mais utiliza esse tipo de sistema para servir conteúdo dentre os demais.

Figura 8 – Tecnologias utilizadas pelos servidores Web



Dentre todos os sites estudados, aqueles pertencentes ao governo e aos municípios dos estados do Rio de Janeiro, São Paulo, Ceará e Rondônia, são o que mais apresentam quantidades semelhantes de sites utilizando tecnologias proprietárias ou não.

Objetos mais frequentes

Dos 192,2 milhões de *links* encontrados nas páginas da *Web .GOV.BR*, cerca de 89% correspondem a algum tipo de arquivo gráfico, 8,3% correspondiam a algum tipo de arquivo hipertexto e 2,5% a algum tipo de arquivo de texto como .DOC, .PDF, .XML, .ODT.

Somente os domínios com algum conteúdo verificável podem ser considerados como *sites* estruturados; para tanto, levantou-se quantos domínios apontavam pelo menos um documento HTML, comumente chamado de página HTML, ou seja, um arquivo que pode ser interpretado por um navegador *Web*, conhecido também como *browser*.

O termo *site*, equivalente a *Website* ou sítio eletrônico, refere-se a um conjunto de páginas HTML referenciadas por um mesmo nome (considerado aqui como o nome de domínio completo) na Internet. Por exemplo, <http://www.prefeitura.sp.gov.br> (considerou-se como URL tudo o que está depois do <http://> e antes da primeira *"/*). As URLs <http://site.prefeitura.sp.gov.br/pagina1.html> e <http://site.prefeitura.sp.gov.br/calendario/evento.html> fazem parte do mesmo *site*, para efeito dessa pesquisa, enquanto <http://www.prefeitura.sp.gov.br> refere-se a um *site* diferente.

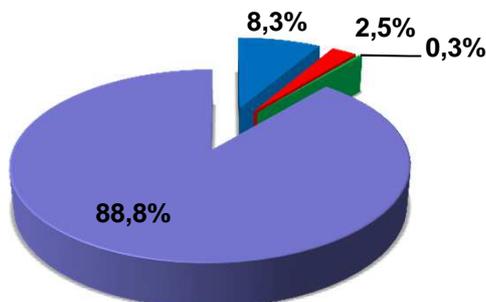
Do total inicial de 18.796 *sites*, apenas 11.586 apresentaram essas características. Efetuou-se também o levantamento do total de arquivos digitais para cada grupo.

A coleta de dados realizada identificou 7.947.607 arquivos digitais; destes, 6.331.256 são documentos em formato HTML, criados ou não por sistemas automatizados de geração de conteúdo. Os demais 1.616.351 arquivos digitais não HTML são arquivos em outros formatos, como: TXT, SWF, EXE, ZIP, RAR. Ainda segundo a coleta, o número médio de documentos HTML por *site* é de 534 documentos. Todos esses números descrevem de forma sucinta algumas características dos *sites* de governo presentes na *Web* brasileira.

Objetos mais usados nas páginas da Web brasileira de governo

O levantamento indicou que entre todos os 192,2 milhões de *links* encontrados nas páginas da *Web .GOV.BR*, cerca de 89% correspondem a algum tipo de arquivo gráfico, 8,3% correspondiam a algum tipo de arquivo hipertexto e 2,5% algum tipo de arquivo de texto como .DOC, .PDF, .XML, .ODT, conforme apresenta a Figura 9.

Figura 9 – Objetos mais frequentes na *Web* brasileira de governo



- Hipertexto
- Documento
- Planilha + Apresentação + Banco de Dados + Áudio + Vídeo
- Gráficos

Objetos no .GOV.BR

Dos links para objetos gráficos identificados, cerca de 99% apontavam para imagens em formato .GIF, .JPG, .PNG ou .BMP. Os arquivos em formato .PDF representam 80% dentre todos os tipos de documentos coletados....

Figura 11 – Tipos de hipertexto mais utilizados

Base utilizada: 15.957.331 objetos coletados

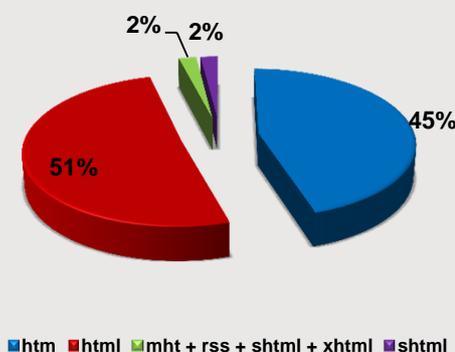
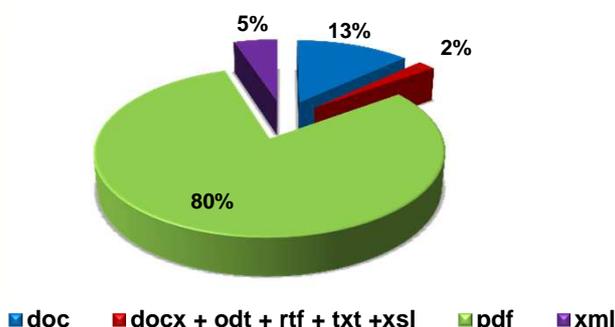


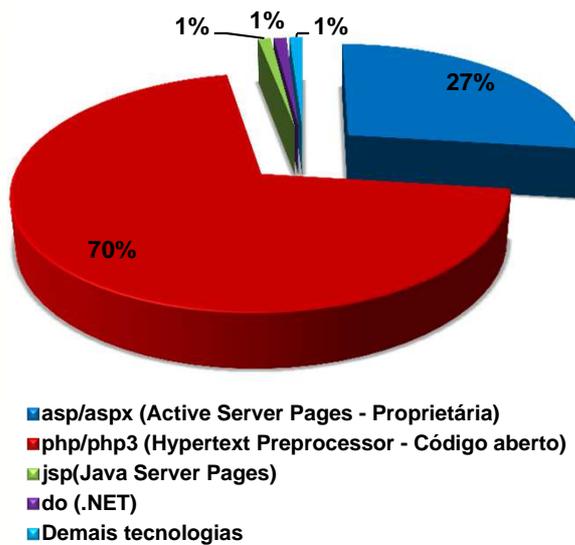
Figura 12 - Tipos de documentos mais utilizados

Base utilizada: 4.821.244 objetos coletados



As tecnologias empregadas na distribuição de informação de maneira automatizada dos sites do governo brasileiro são basicamente de dois tipos: PHP e ASP. As tecnologias baseadas em sistemas de código aberto, como o PHP, foram aquelas que predominaram no conjunto total das páginas de governo coletadas. 70% das páginas HTML coletadas tinham a extensão .PHP.

Figura 10 – Tipos de tecnologia utilizada para gerar documentos



Dos links para objetos gráficos identificados, cerca de 99% apontavam para imagens em formato .GIF, .JPG, .PNG ou .BMP. Os arquivos em formato .PDF representam 80% dentre todos os tipos de documentos coletados; já os arquivos em formato .DOC representam 13%. Esses resultados estão ilustrados nos gráficos das Figura 11 e 12.

O Horário dos servidores

Apenas pouco mais da metade dos servidores está corretamente sincronizada, e o restante apresenta diferenças em relação a Hora Legal Brasileira entre 1 segundo e até mais de que duas horas, denotando a necessidade de revisão nas configurações.

A sincronização dos relógios dos servidores, estações de trabalho e outros dispositivos conectados à Internet é importante para o correto funcionamento de muitas aplicações, bem como em situações em que a análise dos registros (*logs*) feitos pelas aplicações é necessária para tratar incidentes de segurança e eventos correlatos. O NIC.br provê um serviço público e gratuito que fornece a Hora Legal Brasileira via Internet, em conjunto com o Observatório Nacional, instituição responsável pela sua definição. Oferece ainda um sítio *Web* com informações e instruções sobre como utilizar esse serviço: o <http://ntp.br>. O CGI.br recomenda formalmente a sincronização de todos os dispositivos ligados à rede. Para mais detalhes veja a resolução **CGI.br/RES/2008/009/P** disponível em:

<http://www.cgi.br/regulamentacao/resolucoes.htm>

Para aferir a sincronização dos servidores que hospedam os *sites* .GOV.BR, foi obtido o horário de seus relógios, via protocolo http, e este comparado com a hora correta. O resultado, ver Figura 13, mostra que apenas pouco mais da metade dos servidores está corretamente sincronizada, e o restante apresenta diferenças em relação a Hora Legal Brasileira entre 1 segundo e até mais de que duas horas, denotando a necessidade de revisão nas configurações.

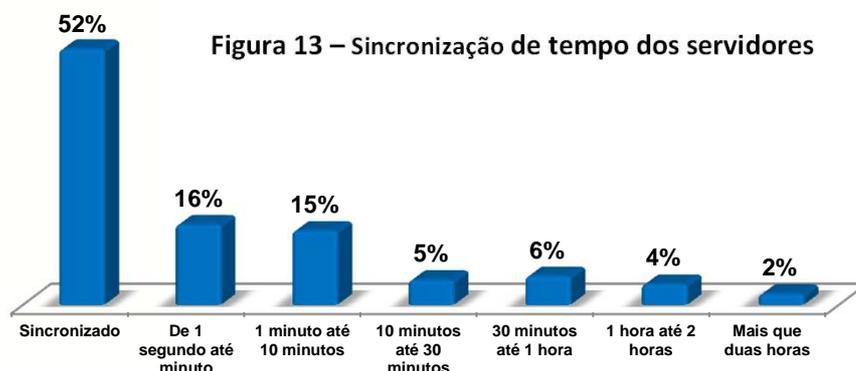


Figura 14 – Geolocalização dos Servidores Web



Geolocalização dos IPs

Estimou-se a geolocalização dos servidores com o uso da base de dados GeoLite, da empresa MaxMind¹, consultada a partir de seus endereços IP.

Servidores *Web* hospedados em locais distantes dos usuários implicam uma velocidade menor de acesso, por conta do tempo de tráfego dos pacotes. A hospedagem dos servidores no exterior, além disso, colabora para o aumento dos custos de acesso à Internet no Brasil, já que implica maior utilização dos enlaces internacionais, com custo alto, pelas operadoras de telecomunicações.

Cerca de 6% dos *sites* .GOV.BR estão hospedados fora do país.

1 - "This product includes GeoLite data created by MaxMind, available from <http://maxmind.com/>"

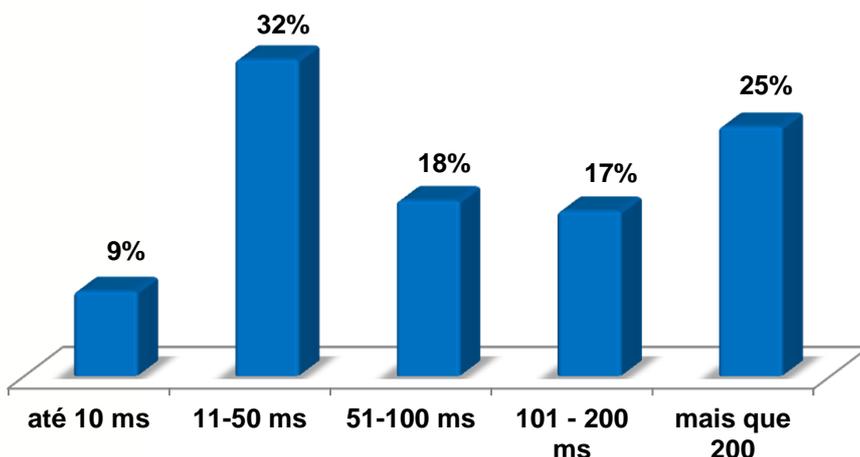
Tempo médio de respostas dos servidores brasileiros

O tempo médio de resposta dos servidores, nesse estudo, consiste no tempo que levaram para responder uma consulta http simples (HEAD), incluindo o tempo de ida e volta dos pacotes de dados, mais o tempo de processamento do servidor. O teste é influenciado, portanto, pela localização do medidor, na rede do NIC.br, em São Paulo.

O tempo de resposta dos servidores

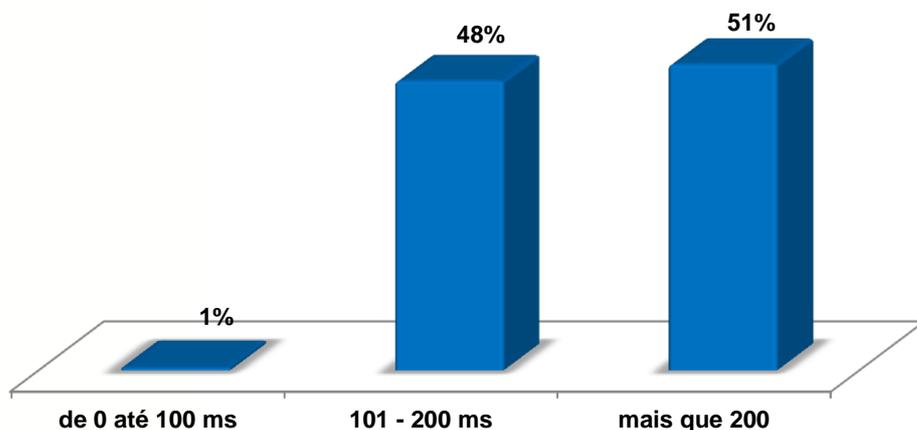
O tempo de resposta dos sites hospedados fora do Brasil têm resultados piores. Dos hospedados no país, aproximadamente 59% apresentaram tempo de resposta de até 100ms.

Figura 15 – Distribuição do tempo médio de resposta para sites hospedados no Brasil



De forma simplificada, o indicador pode ser considerado uma medida de desempenho do site, do ponto de vista de um usuário localizado em São Paulo. Pode se notar como os sites hospedados fora do Brasil (Figura 16) têm resultados piores do que os hospedados no país (Figura 15). Dos hospedados no país, aproximadamente 59% apresentaram tempos até 100ms, o que é um bom resultado, mas apontam para a possibilidade de melhoria na infraestrutura dos demais servidores e na própria infraestrutura da Internet brasileira.

Figura 16 – Distribuição do tempo médio de resposta para sites hospedados no exterior



IPV6 no .GOV.BR

Nenhum *site* estava disponível via protocolo IPV6 no censo da *Web para o .GOV.BR*.

O protocolo IP é a base tecnológica que sustenta a Internet; é quem torna possível a utilização inteligente da infraestrutura de telecomunicações que faz da Internet uma rede muito mais ubíqua, versátil e acessível, em comparação aos serviços convencionais de telecomunicações. Todas as aplicações Internet, inclusive a *Web*, amparam-se, num nível mais básico, neste protocolo. A informação na Internet é dividida em pacotes que trafegam de forma independente pela rede, e o IP é quem permite que eles encontrem seu caminho, identificando cada dispositivo na Internet com um número, o “endereço IP”.

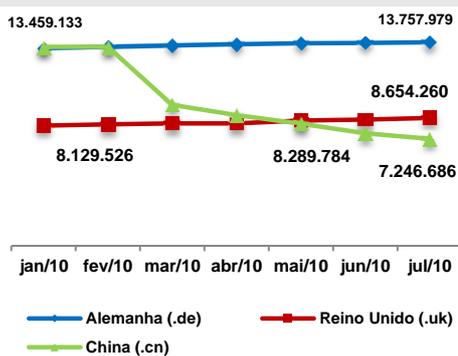
A versão mais usada hoje do protocolo, o IPv4, tem cerca de 4 bilhões de endereços possíveis, mas cerca de 94% desse total já é utilizado. Com o IPv4, a Internet só consegue manter o atual ritmo de expansão por mais 1 ou 2 anos. Por isso um novo protocolo, o IPv6, será introduzido na rede; ele deverá funcionar lado a lado com o IPv4 durante um período de transição e o substituirá a longo prazo, possibilitando a continuidade da expansão da Internet. Todos devem fazer a implantação de forma célere, pois quando o IPv4 esgotar-se, paulatinamente surgirão serviços e usuários que conseguirão comunicar-se utilizando apenas o IPv6.

O *Governo Federal* mostra compreensão sobre a gravidade da situação ao estabelecer no documento de referência da e-PING, que: “Os órgãos da Administração Pública Federal deverão se interconectar utilizando IPv4 e planejar sua futura migração para IPv6. Novas contratações e atualizações de redes devem prever suporte à coexistência dos protocolos IPv4 e IPv6 e a produtos que suportem ambos os protocolos².”

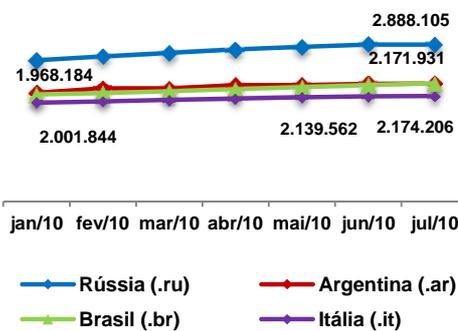
Nenhum *site* estava disponível via protocolo IPV6 no censo da *Web para o .GOV.BR*.

NOMES DE DOMÍNIO NO MUNDO

A alteração da regulamentação para o registro de nomes de domínios na China tira o ccTLD **.cn** da liderança absoluta que agora é ocupada pela Alemanha com cerca de 13,7 milhões de registros no mês de julho.



Desempenho do registro de nomes de domínios da Alemanha, China e Reino Unido desde janeiro de 2010.



Desempenho do registro de nomes de domínios da Rússia, Argentina, Brasil e Itália desde janeiro de 2010.

cgi.br nic.br

A dinâmica do registro de nomes de domínio ao redor do mundo

O registro de nomes de domínio no mundo

A Alemanha e o Reino Unido como os líderes no registro de nomes de domínio entre todos os demais ccTLDs. A China continua em queda pelo sexto mês consecutivo desde a alteração da regulamentação para o registro de nomes de pessoas físicas. O fim do preços promocionais também podem explicar essa queda, visto que até fevereiro de 2010 o registro de nomes de domínios para pessoas físicas era cerca de um dólar.

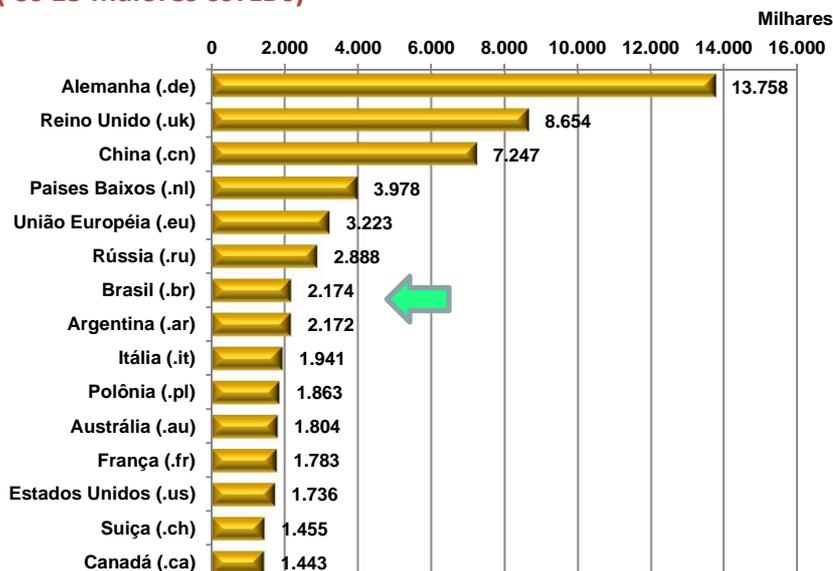
Rank	ccTLD	Domínios	Ref.	Fonte
1	Alemanha (.de)	13.757.979	Julho/10	http://www.denic.de/
2	Reino Unido (.uk)	8.654.260	Julho/10	http://www.nominet.org.uk/intelligence/statistics/
3	China (.cn)	7.246.686	Julho/10	http://www.cnnic.cn/html/Dir/2003/12/13/2020.htm
4	Países Baixos (.nl)	3.978.085	Julho/10	http://www.domain-registry.nl/
5	União Européia (.eu)	3.223.130	Julho/10	http://www.eurid.eu/
6	Rússia (.ru)	2.888.105	Julho/10	http://cctld.ru/
7	Brasil (.br)	2.174.206	Julho/10	http://registro.br/
8	Argentina (.ar)	2.171.931	Julho/10	http://www.latinamericann.org/
9	Itália (.it)	1.940.529	Julho/10	http://www.nic.it/
10	Polônia (.pl)	1.862.596	Julho/10	http://www.dns.pl/
11	Austrália (.au)	1.803.927	Julho/10	http://www.auda.org.au/
12	França (.fr)	1.783.115	Julho/10	http://www.afnic.fr/actu/stats
13	Estados Unidos (.us)	1.735.912	Julho/10	http://www.domaintools.com/internet-statistics/
14	Suíça (.ch)	1.454.660	Junho/10	https://www.nic.ch/
15	Canadá (.ca)	1.442.809	Julho/10	http://www.cira.ca/

O registro de nomes sob ccTLD **.br** permanece em crescimento, atingindo em julho de 2010 pouco mais de 2,17 milhões de nomes. O país continua ocupando a nona posição entre todos os ccTLDs

O registro de nomes entre os principais gTLDs

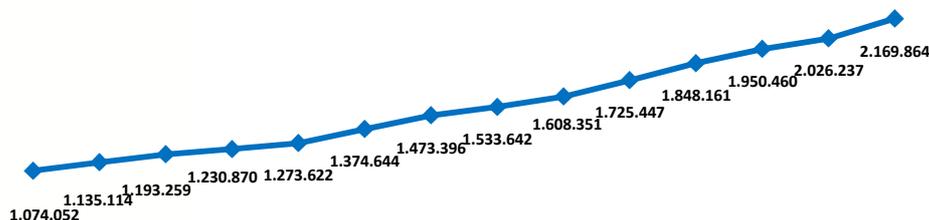
Rank	gTLD	Domínios	Ref.	Fonte
1	.com	88.839.287	Julho/10	http://www.domaintools.com/
2	.net	13.239.274	Julho/10	http://www.domaintools.com/
3	.org	8.614.406	Julho/10	http://www.domaintools.com/
4	.info	6.757.195	Julho/10	http://www.domaintools.com/
5	.biz	2.142.388	Julho/10	http://www.domaintools.com/

Informações sobre nomes de domínio no mundo (os 15 maiores ccTLDs)



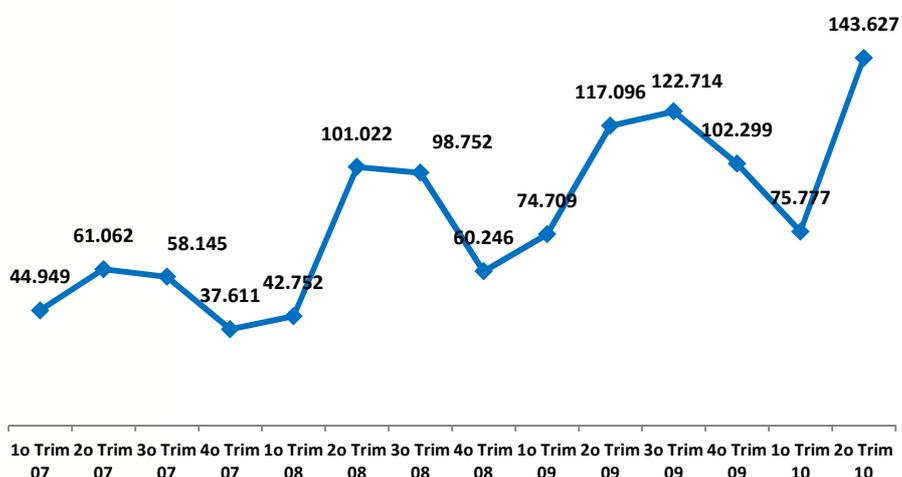
Dinâmica de crescimento do registro de nomes de domínios no Brasil – Julho/2010

O segundo trimestre de 2010 encerrou com cerca 2,17 milhões de nomes de domínios registrados sob o ccTLD .br. Este número representa um crescimento de 7,1% em relação ao trimestre anterior e 25,7% em relação ao mesmo período no ano anterior.



1o Trim 07 2o Trim 07 3o Trim 07 4o Trim 07 1o Trim 08 2o Trim 08 3o Trim 08 4o Trim 08 1o Trim 09 2o Trim 09 3o Trim 09 4o Trim 09 1o Trim 10 2o Trim 10

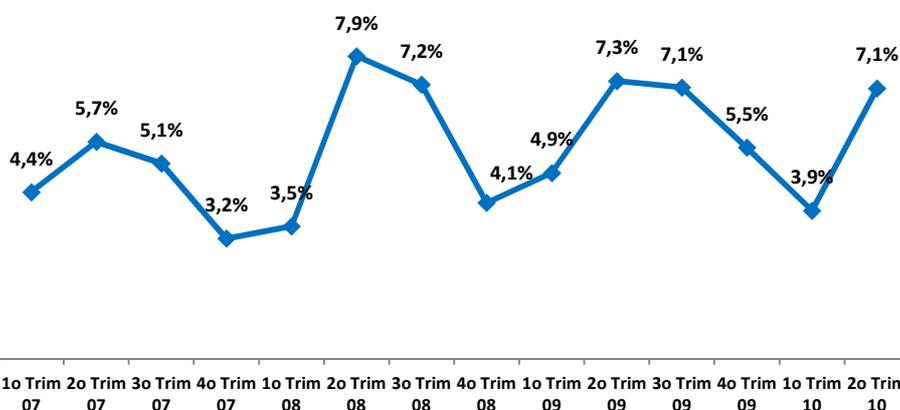
Saldo total de registros por trimestre entre Jan/2007 e Jul/2010



1o Trim 07 2o Trim 07 3o Trim 07 4o Trim 07 1o Trim 08 2o Trim 08 3o Trim 08 4o Trim 08 1o Trim 09 2o Trim 09 3o Trim 09 4o Trim 09 1o Trim 10 2o Trim 10

O saldo total entre novos registros e cancelamentos é o maior já observado em toda a história do registro de nomes sob o ccTLD .br, foram pouco mais de 143 mil novos registros neste segundo trimestre.

O saldo total entre novos registros e cancelamentos é de pouco mais de 143 mil nomes de domínio para o final do segundo trimestre. Este é o maior saldo já observado na história do registro de nomes sob o ccTLD .br.



1o Trim 07 2o Trim 07 3o Trim 07 4o Trim 07 1o Trim 08 2o Trim 08 3o Trim 08 4o Trim 08 1o Trim 09 2o Trim 09 3o Trim 09 4o Trim 09 1o Trim 10 2o Trim 10

Mesmo sendo este o maior crescimento da história em número absoluto do saldo de registros sob ccTLD .br, a taxa de crescimento relativa observada desse segundo trimestre é ligeiramente menor do que aquela verificada no segundo trimestre de 2009 que foi de 7,3%.

O Projeto SIMET



O Sistema de Medição de Tráfego de Última Milha, SIMET, tem como principal objetivo mensurar a qualidade da conexão à Internet. Para isso, o SIMET usa um *applet* java.

Para a utilização do SIMET é necessário autorizar sua execução e acesso à Internet, caso seu sistema operacional ou firewall solicitem. O *applet* instalado na máquina do usuário utiliza os protocolos TCP e UDP para realizar uma série de testes, enviando e recebendo dados dos servidores do sistema.

Os servidores do SIMET estão localizados no NIC.br e nos Pontos de Troca de Tráfego do PTT Metro. O SIMET identifica de forma automática quais servidores o usuário final está apto a utilizar, e lhe dá a opção de escolher entre aqueles que estão disponíveis. Muitas vezes apenas um servidor é mostrado, o do NIC.br, em São Paulo. Isso significa que o provedor do usuário de Internet não está diretamente conectado a nenhum dos PTTs participantes do projeto e que o teste será feito pela Internet (os dados podem passar por provedores intermediários). Isso não é um problema, mas havendo opção, sempre é melhor escolher os servidores conectados aos PTTs, porque os resultados dos testes serão mais precisos.

É importante salientar que embora a equipe que projetou e construiu o SIMET tenha tomado um cuidado muito grande com a exatidão dos resultados, alguns fatores, como o desempenho do hardware do computador do usuário, a execução simultânea de outros programas, a utilização compartilhada da rede local e o meio físico utilizado para o acesso a Internet podem afetar o desempenho dos testes. O SIMET pode, devido a estes fatores, apresentar em alguns casos resultados que não correspondem às exatas características de sua conexão.

Todos podem ajudar fazendo o teste identificado e, assim, também ajudar a divulgar o SIMET. Vale ressaltar ainda que além de ajudar a verificar se a sua conexão à Internet está OK, ele ajuda o NIC.br e o CGI.br a obter informações sobre as conexões à Internet em todo o Brasil. Isso permitirá, o futuro, que outras ações sejam tomadas em favor da melhoria da qualidade da infra-estrutura de Internet no país.

O SIMET ainda está em sua versão beta. Mais informações podem ser obtidas em <http://simet.nic.br/>.

Fonte: CEPTR0 (julho de 2010)

Site oficial do projeto SIMET: <http://simet.nic.br>

O *applet* instalado na máquina do usuário utiliza os protocolos TCP e UDP para realizar uma série de testes, enviando e recebendo dados dos servidores do sistema.