

dimensões e características da

Web

brasileira: um estudo do .gov.br

2011

egi.br

Brasília Internet Steering
Committee

nic.br

Brasília Network
Information Center

Análise das características técnicas da web brasileira: um estudo do “.gov.br”

Temos a satisfação de apresentar a segunda edição da *Análise das características técnicas da web brasileira: um estudo do “.gov.br”*, desenvolvido por W3C Brasil Ceptro e Cetic, áreas do Comitê Gestor da Internet no Brasil (CGI.br) e Núcleo da Informação e Comunicação do Ponto br (NIC.br).

Este estudo tem como objetivo acompanhar ao longo do tempo a evolução técnica das páginas web brasileiras, em seus aspectos de aderência aos padrões internacionalmente aceitos e o quanto elas referenciam tecnologias abertas ou fechadas.

O estudo também objetiva acompanhar ao longo do tempo alguns aspectos técnicos dos servidores web que indiquem sua prontidão para novo protocolo IPv6, para a hora legal brasileira e para tempo de respostas aceitáveis.

Até o presente momento esse estudo está desenhado para os sítios da Web governamental brasileira, isto é, aqueles sítios que estão debaixo do domínio “.gov.br”. Todas as páginas abertas da Web governamental foram coletadas para prover os dados e as análises deste estudo (vide o capítulo Metodologia para a explicação de como foram feitas as coletas).

Com o estudo buscamos oferecer indicadores aos gestores públicos que apontem o status atual de suas organizações, revelem melhores práticas e estimulem a melhoria da web brasileira de tal forma que ela seja para todos, aberta, interoperável e preparada para novas tecnologias e padrões.

Os resultados que ora apresentamos referem-se aos dados coletados no final de 2011. Eles não apresentam significativas diferenças com relação ao ano anterior. Mas, alguns avanços e recuos merecem destaques nesse apresentação.

Ainda que seja modesta a adoção do IPv6 na Web governamental brasileira, houve um crescimento no número de sítios que de alguma maneira foi identificado movimento na direção de adoção do protocolo. Destacam-se Espírito Santo, Minas Gerais e Rio de Janeiro com ações acima da média.

O indicador de aderência aos padrões Web HTML também sofreu variação positiva. Páginas Web codificadas em HTML sem erros, aderentes aos padrões HTML do W3C, apresentam melhor desempenho em sua apresentação, possibilitando uma experiência de navegação na Web mais rica para os usuários. O indicador de páginas Web sem erros cresceu de 5,0 % para 6,38%. Ainda que pequeno o resultado verificado, ele aponta que está havendo uma preocupação com a aderência a padrões. Destacaram-se os Estados de Minas Gerais, Maranhão e Tocantins com aqueles que apresentaram um nível de páginas sem erros acima de 10%.

Outro indicador que apresentou uma variação significativa foi o de Acessibilidade na Web. Páginas Web aderentes às diretrizes de Acessibilidade do W3C significam que elas podem ser acessadas por pessoas com algum tipo de deficiência, seja leve ou severa. Tecnologias estão disponíveis para garantir a essas pessoas uma navegação o mais rica possível, porém para funcionar elas necessitam que as páginas sigam os padrões de acessibilidade.

O indicador de acessibilidade cresceu de 2,0 % em 2010 para 4,82 % em 2011. Sem dúvida que proporcionalmente é um índice baixo de acessibilidade, porém é significativo que tenha mais que dobrado de tamanho, o que significa uma bem maior preocupação dos governos com a inclusão digital de pessoas com deficiência.

O indicador de distribuição de tipos de objetos na Web mostra dados sobre a quantidade de objetos usados nas páginas Web governamental brasileira nas categorias apresentação, áudio e vídeo, documento, gráficos, hipertexto e planilha. O destaque é o significativo aumento do uso do PDF como extensão para documento na Web. A variação de 80,2% para 90,5% indica a preferência governamental por uma plataforma fechada para publicar seus documentos.

Por outro lado, houve um crescimento importante no uso do formato gráfico PNG na preferência para exibição de imagens gráficas, seguindo uma tendência para formatos abertos e mais leves em ambiente Web.

Outro indicador que se destaca é o tempo de resposta dos servidores web que mede se o tempo de resposta verificado está dentro dos padrões que

considera excelente tempo inferior a 50 milissegundos e razoável inferior a 200 milissegundos. Nesse indicador, houve uma degradação no tempo de resposta, pois a média caiu de 190 para 360 milissegundos, o que indica uma baixa qualidade dos servidores e/ou serviços providos.

Convidamos a todos para ler atentamente todos os indicadores que apresentamos neste documento. Estamos certos que eles contém informações preciosas para ajudar na melhoria da qualidade da web governamental brasileira. Essa segunda edição já apresenta com números a evolução da Web em diversos indicadores, ainda que tímidos na maioria. Mas, permite concluir que temos gestores e desenvolvedores mais atentos à necessidade de termos uma web brasileira de tal forma que ela seja para todos, aberta, interoperável e preparada para novas tecnologias e padrões.

Os desafios técnicos para o estudo da Web brasileira

Nesta segunda edição do projeto, nossos objetivos não se desviaram muito dos originais, no entanto foi dado um foco principal às análises evolutivas e na validação dos dados. Assim, algumas das principais perguntas que queríamos e ainda queremos ajudar a responder englobam: Que tipos de sítios são desenvolvidos na Web brasileira? Quais são os mais ou menos importantes? São sítios grandes ou pequenos? E, são voltados ao público nacional ou internacional?

Conduzimos também um projeto para a disseminação do IPv6 no país, e o acompanhamento de sua adoção na Web brasileira é um ótimo indicador da efetividade de nossas ações, bem como dos acontecimentos internacionais a cerca da implantação desse protocolo. De forma semelhante, gostaríamos de manter dados sobre a proporção de servidores Web sincronizados com a hora correta, o que é recomendado pelo CGI.br e possibilitado através do serviço NTP.br oferecido em conjunto com o Observatório Nacional.

Outro dos mais importantes itens que motivou o desenvolvimento e evolução deste projeto veio do escritório brasileiro do W3C e está ligada à necessidade do governo de conhecer melhor a aderência aos padrões de acessibilidade de seus sítios. Inicialmente, percebeu-se a relevância de se conhecer melhor a Web brasileira e decidimos nos dedicar ao projeto. Até então, não tínhamos ideia, contudo, do tamanho do desafio ao qual nos proporíamos, principalmente em seus aspectos técnicos.

A forma como é constituída a Web, por si só, traz diversas dificuldades. Por exemplo, não há realmente uma “Web brasileira”; a World Wide Web, como o próprio nome diz, é uma rede de alcance mundial. Limitar o escopo do estudo foi o primeiro dos desafios. O que faríamos? Consideraríamos a linguagem das páginas? Se o fizéssemos, haveria a dificuldade em si, de identificar o idioma, e o risco de considerarmos sítios dos demais países lusófonos. Consideraríamos, então, a geolocalização dos servidores utilizados para hospedar a Web? Sabíamos de antemão que uma grande percentagem dos domínios “.br” estavam hospedados fora do país. Decidimos considerar apenas a Web formada pelos domínios “.br”, conscientes de que há sítios nacionais hospedados em

outros domínios que ficariam fora do estudo. Para minimizar o problema, consideramos no estudo os sítios sob outros domínios encontrados por meio de um redirecionamento a partir de uma URL que apontasse para um “.br”.

A Web é uma rede cujos conteúdos estão interligados através de documentos de hipertexto. Seu estudo é possível por um processo de análise e coleta sucessiva das páginas, a partir de um conjunto de sítios previamente conhecidos. Essa busca é feita de forma automática por um programa de computador normalmente chamado de crawler, coletor, ou batedor. Nem toda a Web está interligada, embora a maior parte dela esteja, existem “ilhas” de tamanhos variados sem ligação com o restante da rede. Isso significa que o conjunto inicial de sítios a partir dos quais a pesquisa é feita influencia o resultado e que encontrar o conjunto adequado, geralmente o mais completo possível, é um passo importante. Na coleta do “.gov.br”, por exemplo, a situação ideal seria conhecermos os domínios registrados diretamente sob o “.gov.br”, mais os domínios registrados sob as siglas das unidades federativas, como “.sp.gov.br”. Os primeiros estão sob responsabilidade do Governo Federal, e obtivemos a base; os demais são responsabilidade dos Governos Estaduais e contamos com o apoio da ABEP (Associação Brasileira de Entidades Estaduais de TICs) em sua obtenção. Ainda assim, na primeira rodada de coletas, apenas 8 unidades federativas nos enviaram os dados e, nesta última, não houveram envios. Na primeira, complementamos os dados utilizando sítios encontrados em buscadores Web e, na de 2011 utilizamos como base a lista de sítios encontrados na coleta anterior.

Há também armadilhas para o processo de coleta: sítios com um número infinito de páginas, geradas dinamicamente. Elementos simples, como um calendário gerado automaticamente no sítio, podem criar situações desse gênero. Limites de tamanho e profundidade têm de ser estabelecidos, com o risco de impedirem a coleta de partes de sítios maiores que estes. Nesta evolução do projeto, fizemos algumas modificações no crawler com o intuito de solucionar alguns desses problemas, como a normalização das URLs, porém eles continuaram existindo.

Outra das dificuldades técnicas a ser lembrada é aquela que apelidamos de “Web profunda”: a parte da rede em que é requerida a autenticação do usuário para a navegação, por exemplo a maior parte dos sítios de relacionamento ou comunidades. Essa parte da Web é inacessível através do método utilizado, tendo ficado fora do estudo. Existe ainda a possibilidade de serem consultados servidores temporariamente indisponíveis, ou de serem encontrados sítios sem o arquivo robots.txt, responsável por especificar quais de suas páginas podem ou não ser visitadas por batedores automatizados.

Consideramos, ainda, os recursos de tempo, processamento, conectividade e disco, para coletar, armazenar e processar os dados: mesmo agora, com duas análises do estudo concluídas, temos dificuldade em estimar o que seria necessário para um estudo no formato censitário de toda a Web “.br”. Com isso, optamos por realizar análises amostrais desse domínio a serem divulgadas posteriormente.

Ao aventarmos a possibilidade de fazer o estudo, um dos primeiros passos foi procurar por pesquisas similares realizadas anteriormente, e por ferramentas. Encontramos algumas pesquisas de cunho acadêmico, inclusive realizadas sobre a Web brasileira, que nos auxiliaram no processo. Encontramos também algumas ferramentas que poderiam ser aproveitadas para a coleta dos dados. Em particular, estudamos três sistemas para essa finalidade: o Nutch, um coletor utilizado para a criação de buscadores; o Heritrix, usado no Web Archive, um projeto que mantém um arquivo histórico de parte relevante da Web; e o WIRE, utilizado em um dos estudos acadêmicos que encontramos, escrito justamente com a finalidade de realizar estudos sobre a Web, tendo já embutidas algumas ferramentas de análise que consideramos de interesse: análise do tamanho das páginas, tipos de documentos, idiomas, cálculo de rankings, etc. A conclusão foi: começar o estudo utilizando qualquer uma delas traria vantagens em relação ao desenvolvimento de uma ferramenta inteiramente nova. Optamos pelo WIRE, principalmente pela existência das funcionalidades de análise, já incorporadas ao programa. No entanto, dado alguns problemas de desempenho observados com sua utilização em domínios maiores, começamos a considerar a utilização das outras opções.

Além das modificações realizadas para a primeira coleta do “.gov.br”, que incluíam, por exemplo, mudanças em sua forma de armazenamento e no tratamento de redirects nas páginas, nesta segunda coleta, foram corrigidos alguns problemas quanto à normalização das URLs e tratamento de redirects dentro dos domínios em estudo, que ajudaram a tornar o WIRE mais estável. Também, trabalhamos em conjunto com o INWEB na criação de um Data Warehouse com o intuito de padronizar a geração dos dados e das análises que são feitas a cada novo domínio avaliado.

Gostaríamos, com o TIC Web, de responder a várias questões que não estavam contempladas nos resultados das análises feitas pelo WIRE. Por exemplo: a geolocalização dos servidores, a aderência ao IPv6 e ao NTP, e a aderência aos padrões HTML e de acessibilidade (eMAG / WCAG). Essas análises poderiam ser incorporadas ao WIRE ou implementadas numa ferramenta separada. Optamos pela segunda alternativa, de forma a evitar a inserção acidental de novos bugs no código do WIRE, com o qual ainda não estávamos completamente familiarizados. Foi criada a ferramenta cujo nome provisório é ConNeCTOR, que realiza os testes citados, e tem a função adicional de armazenar tanto os dados do WIRE, quanto os de suas próprias análises, num banco de dados único. Reutilizamos, quando possível, ferramentas já prontas. Por exemplo, para verificar a aderência ao padrão HTML usamos o validador criado pelo W3C, rodando localmente, o qual é consultado pelo ConNeCTOR. Para os testes de acessibilidade, incorporamos ao programa rotinas do ASES, programa desenvolvido pelo Governo Brasileiro. Nesta segunda etapa, houve um grande esforço acerca da melhora do desempenho e confiabilidade dessas ferramentas.

Ao terminar a análise dos dados dessa segunda coleta da Web “.gov.br”, concluímos que temos um conjunto de ferramentas confiáveis que nos possibilitaram uma análise evolutiva de certos domínios Web e que mais análises devem ser realizadas para melhor entendermos a Web brasileira. Sabemos, no entanto, de limitações que ainda precisam ser vencidas, por isso modificações continuam a ser feitas, seguidas de testes extensivos. Dentre as modificações em curso, podemos destacar: a contagem do tamanho dos objetos não HTML presentes nas páginas, como imagens e vídeos, sem baixá-los,

através de consultas HTTP HEAD; a melhora no tratamento às “armadilhas” citadas anteriormente e a melhora na velocidade das coletas e análises. Além disso, há a necessidade de polir a utilização das ferramentas de análises estatísticas automatizadas baseados em soluções de Data Warehouse e Data Mining.

Neste intervalo de tempo, também fizemos a publicação dos códigos com licenças livres. Porém, ainda não se trabalhou a ideia da geração de uma comunidade acerca desses sistemas.

Trabalhou-se, ainda, em dois sistemas baseados nessas duas ferramentas. O primeiro é chamado de TopSites¹, ele possui a tarefa de analisar semanalmente os sítios listados pela plataforma Alaxa.com², que classifica e disponibiliza gratuitamente o um milhão de sítios mais populares no mundo. Como a frequência e quantidade de sítios são significativamente superiores às das coletas usuais, apenas as informações relacionadas diretamente ao sítio e sua página inicial são extraídas.

O Segundo, chamado Analisador de Sites Automático (ASA)³, é uma plataforma desenvolvida com o intuito de auxiliar desenvolvedores na tarefa de manter sítios em conformidade com os principais padrões Internet. Na prática, um sítio Web que esteja em conformidade com tais padrões consegue prover melhor qualidade de serviços a seus usuários e, o ASA ajuda desenvolvedores ao realizar análises periódicas nos sítios cadastrados e sinalizar as inconsistências presentes.

1 <http://labs.ceptro.br/topsites>

2 <http://www.alexacom.com/>

3 <http://asa.nic.br/>

Metodologia

A palavra censo origina-se no latim census e significa a “contagem ou enumeração completa” de uma população de indivíduos ou objetos determinados. Portanto, um censo é o resultado final de uma contagem específica que define o conjunto de dados estatísticos sobre as diversas variáveis de uma população investigada.

Para o estudo do tamanho e das características da Web governamental brasileira foi realizado um censo com as mesmas definições adotadas na edição anterior desta pesquisa. O censo foi resultado de uma contagem específica que define o conjunto de dados estatísticos sobre as diversas variáveis de uma população investigada. Sua realização tem como requisito o conhecimento dos objetos de análise e procedimentos metodológicos, isto é, a definição prévia dos limites populacionais. Assim as unidades de análise e os procedimentos metodológicos para coleta e tabulação dos resultados seguiram as mesmas premissas.

Assim sendo, assume-se que um conteúdo pertence à Web governamental brasileira se o domínio de topo do nome do seu sítio Web respeita uma das seguintes condições:

1. Domínios identificados como .gov.br (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;
2. Domínios identificados como sigla-uf.gov.br, registrados pelas empresas estaduais de processamento de dados, vinculadas aos governos estaduais;
3. Resultados de consultas e buscas de informações, utilizando ferramentas de busca, com o objetivo de complementar as informações anteriores.

Consideraram-se, ainda, em algumas das análises os links para

documentos presentes nas páginas de sítios .gov.br, mesmo que estejam hospedados fora desta hierarquia de domínios.

Vale ressaltar também que situações de redirecionamento a partir de uma requisição para página Web do domínio .gov.br também são consideradas. Assim, se existe uma página Web encontrada que não está sob a hierarquia .gov.br, mas existe um redirecionamento a partir de um domínio sob o .gov.br, então essa página é considerada.

O estudo da Web é possível através de um processo de análise e coleta sucessivas das páginas que satisfaçam as condições necessárias. Essa busca é realizada de forma automatizada através de um programa de computador normalmente chamado de coletor (crawler Web). Foram estabelecidos limites de tamanho e profundidade uma vez que nem toda a Web possui conteúdos interligados através de documentos de hipertexto, uma vez que há “ilhas” de tamanhos variados sem ligação com o restante da rede e há ainda a Web profunda, a qual para ser acessada demanda a autenticação do usuário.

A1: Tamanho da Web – Sítios

1.1 Definição do indicador

Quantidade e volume (em bytes) dos sítios da Web governamental brasileira (.gov.br).

1.2 Propósito

Identificar o tamanho da Web em relação à quantidade e volume de sítios da Web governamental brasileira (.gov.br).

1.3 Metodologia

Obtém-se o indicador a partir dos resultados apresentados automaticamente pelo coletor de páginas chamado WIRE. O coletor diferencia os sítios a partir da formação das URLs encontradas, analisando cada string. Na obtenção do volume do sítio somente são considerados os recursos escritos na linguagem HTML e suas derivadas. Não são consideradas o tamanho de outros objetos referenciados nas páginas Web.

1.4 Apresentação dos resultados

Um resumo descritivo dos dados sobre o tamanho da Web governamental brasileira (.gov.br) em relação a sítios está apresentado em três tabelas (A1-1, A1-2, A1-3).

A Tabela A1-1 exibe as informações relacionadas ao volume de sítios e algumas medidas descritivas dos dados. Ao todo, foram identificados 18.229 sítios, que ocupam aproximadamente 172,35 GigaBytes. A média de tamanho é de 9,68 MB, sendo que o maior sítio ocupa 1224,9 MB.

Tabela A1-1: Tamanho da Web governamental brasileira (.gov.br) e medidas resumo

Número de sítios	Tamanho					
	Total (MB)	Mínimo (MB)	Mediana (MB)	Média (MB)	Máximo (MB)	CV
18.229	176.486,47	0,00	12,88	9,68	1.224,92	4,95

A Tabela A1-2 apresenta essas informações sobre a quantidade de sítios e tamanho segundo as unidades federativas (UF) . Pode-se destacar que Roraima (RR) possui menos sítios se comparada as outras UF, além disso esses sítios são menores. Ao todo foram coletados 61 sítios que totalizaram 41,68 MB dele. Já o estado com maior número de sítios e dados foi São Paulo, com 2.267 e 20.447 MB. Alagoas foi o estado com maior tamanho médio 47,15 MB.

Tabela A1-2: Tamanho da Web governamental brasileira (.gov.br) e medidas resumo por unidades federativas

UF	Número de sítios	Páginas / Sítio	Tamanho					CV
			Total (MB)	Mínimo (MB)	Mediana (MB)	Média (MB)	Máximo (MB)	
.GOV.BR	3.104	493,44	37997,42	0	1	12,24	870,84	4,85
.outros	2.260	0,91	44,37	0	1	0,02	3,75	4,74
AC	68	657,34	1057,59	0	10	15,55	367,18	3,38
AL	199	1347,99	9382,24	0	19	47,15	543,28	2,45
AM	225	159,56	885,93	0	60	3,94	220,35	4,5
AP	77	171,39	187,92	0	34	2,44	109,78	5,2
BA	648	493,24	8767,82	0	29	13,53	655,24	4,24
CE	527	554,75	9671,66	0	10	18,35	519,69	3,65
DF	232	512,07	3220,05	0	100	13,88	397,12	3,29
ES	384	150,24	1454,41	0	43	3,79	201,36	4,86
GO	285	296,48	2471,47	0	8	8,67	492,03	4,82
MA	176	249,21	1016,44	0	359	5,78	150,85	3,28
MG	1.181	500,16	13600,7	0	48	11,52	462,65	3,94
MS	275	259,13	1830,08	0	55	6,65	265,43	4,65
MT	233	1130,91	10545,1	0	52	45,26	1036,33	3
PA	262	290,42	1863,69	0	13	7,11	157,53	3,15
PB	271	241,77	1893,9	0	13	6,99	353,72	5,28
PE	289	297,45	2415,33	0	10	8,36	426,96	4,74
PI	152	318,26	1119,55	0	1	7,37	251,22	4,91
PR	1.648	355,08	13336,5	0	8	8,09	399,49	4,78
RJ	782	353,92	6159,09	0	24	7,88	535,17	4,89
RN	260	206,86	2377,03	0	54	9,14	1224,92	8,46
RO	131	417,19	2431,93	0	42	18,56	942,85	5,22
RR	61	39,56	41,68	0	14	0,68	12,66	2,89
RS	803	539,89	9053,55	0	40	11,27	391,86	3,72
SC	992	409,88	8507,9	0	186	8,58	306,7	3,76
SE	242	653,62	3715,47	0	15	15,35	315,63	3,08
SP	2.267	372,64	20447,22	0	45	9,02	619,75	4,57
TO	195	231,12	990,46	0	16	5,08	203,04	4,87

A Tabela A1-3 apresenta os sítios que possuem os maiores volumes. Os 3 maiores sítios em volume são do município de São José de Mipibu (RN), da instituição SECOM (MT) e o Portal do Servidor do município de Porto Velho (RO).

Tabela A1-3: Lista dos 10 maiores sítios em MB da Web governamental brasileira

URL	Tamanho (MB)
www.saojosedemipibu.rn.gov.br	1.224,92
www.secom.mt.gov.br	1.036,33
portalservidor.portovelho.ro.gov.br	942,85
site.portalcofen.gov.br	870,84
www.colider.mt.gov.br	843,97
www.crbio-sc.gov.br	828,68
www.crbio3.gov.br	828,30
www.crbio-rs.gov.br	828,25
www.crbio03.gov.br	817,18
www.plenarinho.gov.br	726,18

1.5 Análise

Os resultados mostram que tamanho médio dos sítios dos estados (11,85 MB para sítios das Ufs, exceto .GOV.BR e outros) é próximo ao tamanho médio dos sítios da Web governamental brasileira (sítios .gov.br, excluindo sítios classificados como “outros” – Tabela A1-2), que é igual a 12,24 MB. Em relação as UF, os estados com maior tamanho médio são Alagoas (47,15 MB) e Mato Grosso (45,26 MB). Provavelmente isso tenha relação com o volume de páginas por sítio, já que esses estados possuem valores de 1.348 e 1.131, respectivamente. Ao avaliar os estados com menor tamanho médio de seus sítios (Roraima com 0,68 MB, Amapá com 2,44 MB, e Espírito Santo com 3,79 MB), pode-se notar a mesma relação, ou seja, são estados que possuem reduzido volume de páginas por sítio (Roraima tem apenas 39,56, Amapá tem 171,39, e Espírito Santo tem 150,24), sendo a média 415,19 páginas por sítio, enquanto a média em sítios do Governo Federal é de 493,44.

Ao comparar o tamanho da Web, relativa ao numero de sítios e o tamanho da população (dados do Censo 2010, IBGE), observa-se proporções ligeiramente diferente, Sendo as Regiões Norte e Sudeste com proporção menor de paginas e a região Sul com proporção superior de paginas frente ao tamanho das respectivas populações. Segue dados sumarizados:

- Centro-Oeste (CO): 1.220 (9,48% dos sítios) e 7% da população;
- Norte (N): 824 (6,40% dos sítios) e 8% da população;
- Nordeste (NE): 2.764 (21,48% dos sítios) e 28% da população;
- Sul (S): 3.443 (26,76% dos sítios) e 15% da população;
- Sudeste (SE): 4.614 (35,86% dos sítios) e 42% da população.

Em relação à quantidade de sítios, pode-se observar que os 10 estados brasileiros com maior volume de sítios representam 74,01% do total (9.521 sítios). São eles: SP, PR, MG, SC, RS, RJ, BA, CE, ES, e PE. Esses 10 estados brasileiros com maior volume de sítios representam 72,79% da população do país.

A partir desses resultados, é possível notar que, de forma geral, a quantidade de sítios nos estados segue aproximadamente outros indicadores demográficos, físicos, geopolíticos e sociais do país. Isso é importante para que se possa desenvolver políticas no sentido de melhorar o atendimento de usuários e geradores de conteúdo e, ao mesmo tempo, promover a produção de conteúdo digital relevante em regiões onde existe maior carência em termos de conteúdos da Web governamental brasileira.

1.6 Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se a ocorrência de um aumento do volume de dados de 148,4 GB para 172,4 GB (ou 176.486 MB), o que equivale a um aumento de 16,2%. Já o número de páginas Web coletadas aumentou de 6.334.054 para 6.874.926, o que equivale a um aumento de 8,5%. O volume de sítios aumentou de 12.891 para 18.229 sítios, equivalente a um aumento de 41,4%. Por fim, o número médio de páginas por sítio passou de 491,36 para 377,14.

A2: Tamanho da Web – Páginas Web

- **Definição do indicador**

Quantidade e volume (em bytes) das páginas da Web governamental brasileira (.gov.br).

- **Propósito**

Identificar o tamanho da Web em relação à quantidade e volume de páginas da Web governamental brasileira (.gov.br).

- **Metodologia**

Obtém-se o indicador a partir dos resultados apresentados automaticamente pelo coletor de páginas chamado WIRE. Considera-se apenas as páginas Web que responderam à requisição de suas URLs contendo código HTTP 2XX. Para filtrar eventuais discrepâncias e erros na geração das páginas durante a coleta, foi estabelecido um limite de 500 KB como quantidade máxima a ser baixado.

- **Apresentação dos resultados**

Um resumo descritivo dos dados sobre o tamanho da Web governamental brasileira (.gov.br) em relação a páginas Web está apresentado em duas tabelas (A2-1, A2-2).

A Tabela A2-1 exibe as informações gerais relacionadas ao volume de páginas Web e algumas medidas descritivas dos dados. No total, foram identificadas 6.874.926 páginas que ocupam aproximadamente 172,35 GigaBytes. A média de tamanho foi de 26,29 KB, sendo que a maior página Web ocupa 488,3 KB.

Tabela A2-1: Tamanho da Web governamental brasileira (.gov.br) e medidas resumo

Número de páginas	Tamanho (KB)					
	Total	Mínimo	Mediana	Média	Máximo	CV
6.874.926	180.722.145,28	0,00	21,28	26,29	488,28	1,04

A Tabela A2-2 apresenta as informações sobre o tamanho da Web segundo as unidades federativas (UF). Pode-se destacar que Roraima (RR) apresentou os menores valores de tamanho, sendo identificadas 2413 páginas que totalizaram 29.501 KB. A UF que apresentou o maior tamanho foi São Paulo com 844.782 páginas Web e 19,9 GB. Em termos de tamanho médio, o menor valor também foi o de Roraima com 12,23 KB e o maior foi Rondônia com 44,72 KB.

Tabela A2-2: Tamanho da Web governamental brasileira (.gov.br) e medidas resumo por unidades federativas

UF	Número de páginas	Tamanho (KB)					
		Total	Mínimo	Mediana	Média	Máximo	CV
.GOV.BR	1.531.651	41.235.927,04	0,00	18,36	26,92	488,28	1,19
.outros	2.059	45.434,88	0,00	11,98	22,07	262,66	1,14
AC	44.699	1.062.522,88	0,00	16,04	23,77	487,92	1,07
AL	268.250	10.129.745,92	0,00	36,76	37,76	485,94	0,42
AM	35.901	856.258,56	0,00	20,57	23,85	474,87	0,77
AP	13.197	162.631,68	0,00	10,91	12,32	260,89	0,59
BA	319.622	8.865.433,60	0,00	21,93	27,74	488,24	0,92
CE	292.353	9.807.196,16	0,00	35,54	33,55	487,81	0,46
DF	118.800	3.295.303,68	0,00	23,38	27,74	488,25	1,13
ES	57.693	1.411.952,64	0,00	20,88	24,47	488,09	0,86
GO	84.496	2.435.717,12	0,00	25,91	28,83	487,84	0,84
MA	43.861	789.534,72	0,00	19,15	18,00	488,24	1,33
MG	590.693	13.551.656,96	0,00	19,54	22,94	488,13	1,01
MS	71.261	1.508.628,48	0,00	19,60	21,17	255,38	0,63
MT	263.501	10.656.675,84	0,00	34,45	40,44	488,28	0,76
PA	76.089	1.811.384,32	0,00	24,40	23,81	487,18	0,48
PB	65.519	1.841.838,08	0,00	24,20	28,11	488,22	1,45
PE	85.963	2.376.775,68	0,00	22,08	27,65	488,18	0,66
PI	48.375	1.065.512,96	0,00	22,89	22,03	288,94	0,47
PR	585.172	13.559.541,76	0,00	19,50	23,17	488,27	0,87
RJ	276.762	6.220.052,48	0,00	19,57	22,47	488,19	0,86
RN	53.783	2.325.504,00	0,00	20,79	43,24	487,07	2,26
RO	54.652	2.444.257,28	0,00	29,30	44,72	488,27	0,98
RR	2.413	29.501,44	0,00	9,94	12,23	146,09	0,69
RS	433.531	9.152.481,28	0,00	18,60	21,11	488,10	1,00
SC	406.602	8.614.727,68	0,00	19,52	21,19	488,22	0,93
SE	158.177	3.730.749,44	0,00	21,00	23,59	488,27	0,78
SP	844.782	20.832.737,28	0,00	18,66	24,66	488,28	1,10
TO	45.069	902.492,16	0,00	13,02	20,02	485,77	1,00

- **Análise**

Os resultados mostram tamanho médio das páginas Web igual a 26,29 KB e, se observado por UF, este tamanho ficou entre 12,23 KB (Roraima) e 44,72 KB (Rondônia). A UF que apresentou maior tamanho absoluto, São Paulo, não reflete este tamanho quando observado o seu valor relativo de 24,66 KB.

Ao comparar o tamanho da Web, relativa ao número de páginas e o tamanho da população (dados do Censo 2010, IBGE), observa-se proporções ligeiramente diferentes. Sendo as Regiões Norte e Sudeste com proporção menor de páginas e a região Sul com proporção superior de páginas frente ao tamanho das respectivas populações. Segue dados sumarizados:

- Centro-Oeste (CO): 583.127 (11% das páginas Web); e 7% da população
- Norte (N): 226.951 (4% das páginas Web) e 8% da população
- Nordeste (NE): 1.335.903 (25% das páginas Web) e 28% da população

- Sul (S): 1.425.305 (27% das paginas Web) e 15% da população
- Sudeste (SE): 1.769.930 (33% das paginas Web) e 42% da população

Ao comparar com a distribuição de sítios por região do país, observa-se pequenas diferenças. As regiões Centro-Oeste e Nordeste apresentam um pouco mais de participação na distribuição de páginas em relação ao volume de sítios. A região Sul tem praticamente a mesma distribuição, enquanto que Norte e Sudeste tem participação um pouco menor em relação a sua distribuição de sítios.

De forma análoga ao resultado obtido para o indicador A1, pode-se notar que, de forma geral, os resultados da distribuição dos valores entre estados seguem aproximadamente os indicadores físicos, geopolíticos e sociais do país. E, a partir deles, pode-se desenvolver políticas que melhor atendam os usuários e geradores de conteúdo e que, ao mesmo tempo, promovam a produção de conteúdo digital relevante em regiões onde esta carência foi detectada.

- **Análise Comparativa com estudo anterior e considerações**

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que ocorreu um aumento de 148,4 GB para 172,4 GB, equivalente a 16,2%, no volume de dados coletados. E, o número de páginas Web coletadas aumentou de 6.334.054 para 6.874.926, o que equivale a um aumento de 8,5%. Em geral, observa-se crescimento em diferentes proporções nos indicadores de tamanhos.

B1: Proporção de sítios com endereço IPv6 em sua URL principal

1. Definição do indicador

Valor percentual de quantidade de sítios da Web governamental brasileira (.gov.br) que possuem endereço IPv6 registrado para seu nome de domínio principal no Sistema de Nomes de Domínios (DNS).

2. Propósito

Obter uma medida da atual quantidade de sítios da Web governamental brasileira (.gov.br) que divulgam endereços IPv6 em seus domínios principais.

3. Metodologia

Para obter o endereço IPv6 dos sítios encontrados na coleta, foram realizadas consultas DNS do tipo "quad A" a cada um deles com a utilização do comando `/dig/`, disponível para plataformas Linux. A resposta foi, então, analisada e, caso o endereço IPv6 fosse informado, ele era registrado e contabilizado.

4. Apresentação dos resultados

Um resumo descritivo dos dados obtidos está apresentado em duas tabelas (B1-1 e B1-2).

Observa-se que 250 dos 18.229 sítios, o que representa 1,37%, possuem endereço IPv6. Observa-se que, segundo a definição do presente projeto, todos os sítios coletados a partir da semente inicial são considerados como pertencentes a esse subconjunto da Web e, portanto, os sítios listados a seguir que não são do subdomínio .gov.br também são considerados na análise uma vez que foram alcançados por redirecionamentos pelo coletor utilizado.

Tabela B1-1: Quantidade e percentual de sítios da Web governamental brasileira (.gov.br) com endereço IPv6 em sua URL principal

Sítios com endereço IPv6	Número total de sítios (Tab. A1-1)	% de sítios com endereço IPv6
250	18.229	1,37%

A Tabela B1-2 apresenta a relação entre as unidades federativas e o número de sítios com endereço IPv6.

Tabela B1-2: Quantidade e percentual de sítios da Web governamental brasileira (.gov.br) com endereço IPv6 em sua URL principal segundo as unidades federativas (UF)

UF	Nº de sítios utilizando IPv6	Total de sítios (Tab. A1-3)	Proporção
.GOV.BR	5	3.104	0,16%
.outros	96	2.260	4,25%
AC	0	68	0,00%
AL	3	199	1,51%
AM	0	225	0,00%
AP	0	77	0,00%
BA	7	648	1,08%
CE	1	527	0,19%
DF	0	232	0,00%
ES	13	384	3,39%
GO	3	285	1,05%
MA	1	176	0,57%
MG	42	1.181	3,56%
MS	2	275	0,73%
MT	3	233	1,29%
PA	3	262	1,15%
PB	2	271	0,74%
PE	4	289	1,38%
PI	0	152	0,00%
PR	6	1.648	0,36%
RJ	15	782	1,92%
RN	2	260	0,77%
RO	0	131	0,00%
RR	0	61	0,00%
RS	4	803	0,50%
SC	9	992	0,91%
SE	0	242	0,00%
SP	25	2.267	1,10%
TO	4	195	2,05%

5. Análise

A Web governamental brasileira (.gov.br) não adota de forma consistente o IPv6. Está presente em apenas 1,37% dos domínios pesquisados. Podemos observar também que aproximadamente 60% dos sítios com suporte são de fato domínios .gov.br, pertencendo quase em sua totalidade à estados da região Sudeste (ES, MG, SP, RJ), com ênfase para os estados de Minas Gerais e Espírito Santo, que mostram porcentagens significativamente acima da média geral.

Conforme exibido na Tabela B1-2, pode-se notar que parte dos domínios que suportam IPv6 são domínios que não contêm a extensão .gov.br e que, portanto, foram alcançados a partir de redirecionamento dos sites governamentais.

Para completar essa análise é importante observar que os indicadores B1 a B4, relacionados ao IPv6, avaliam o nível de suporte ao uso do protocolo IP v6 de diferentes formas. No indicador B1, é verificada a existência de endereço IPv6 associado a URL principal do sítio. No indicador B2, é verificada a existência de endereços IPv6 associados a URLs alternativas de um sítio Web formados a partir da adição de prefixos, como “v6” ou “www6”, ao nome de domínio desse sítio. No B3, cada um dos

sítios com endereço IPv6, alternativo ou principal, é avaliado quanto à sua resposta a uma requisição PING6, que verifica se o servidor identificado realmente responde a uma sonda de rede (usando o protocolo ICMP6). Por fim, no indicador B4, consulta-se os servidores com endereço IPv6 via HTTP GET na porta 80 usando-se o protocolo IPv6 para verificar se o servidor Web está realmente ativo também em IPv6. Dessa forma, pode-se afirmar que cada indicador verifica a aderência ao protocolo IPv6 de forma mais restritiva e direta que o anterior. Uma análise comparativa desses indicadores permite uma melhor avaliação do estado em evolutivo da implantação do protocolo.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que houve um crescimento no uso do IPv6, que passou de 0,031% para 1,37%.

Mesmo essa quantidade ainda sendo pequena, somente 250 sítios, ela representa um avanço significativo em relação ao resultado anterior e mostra uma tendência à adoção do IPv6. O ano de 2012 marca o início de diversas campanhas massivas de divulgação da necessidade de configuração do IPv6. E, por isso a importância de se verificar a evolução do suporte a esse novo protocolo.

B2: Proporção de sítios com endereço IPv6 em URLs alternativas

1. Definição do indicador

Valor percentual de sítios da Web governamental brasileira (.gov.br) que possuem nomes de domínios alternativos registrados com endereços IPv6 no Sistema de Nomes de Domínios (DNS). Por nomes de domínios alternativos entende-se adição, à parte da URL que define o sítio, de alguns prefixos comumente utilizados para designar a versão de um site com suporte a esse novo protocolo. No caso, foram utilizados os prefixos "ipv6", "v6", "www.v6", "www.ipv6" e "www6" para a realização dos testes.

2. Propósito

Obter uma medida da quantidade de sítios da Web governamental brasileira (.gov.br) que estão se preparando para a adoção do IPv6 com a utilização de mecanismos preliminares de testes com domínios esse mecanismo de denominação explícita para o IPv6.

3. Metodologia

Para obter esse indicador, o programa ConNeCTOR foi usado para realizar consultas DNS do tipo "quad A" para cada uma das URL's formadas com a união dos prefixos citados aos nomes de cada um dos sítios encontrados durante a coleta. Caso mais de um endereço IPv6 fosse encontrado, apenas o último era registrado.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre sítios da Web governamental brasileira (.gov.br) com endereço IPv6 em URLs alternativas está apresentado em duas tabelas (B2-1 e B2-2).

Obteve-se como resultado que 250 dos 18.229 sítios coletados utilizam domínio alternativo IPv6. Observa-se que, segundo a definição do presente projeto, todos os sítios coletados a partir da semente inicial são considerados como pertencentes a esse subconjunto da Web e, portanto, os sítios que não são do subdomínio .gov.br também são considerados na análise uma vez que foram alcançados por redirecionamentos pelo coletor utilizado.

Tabela B2-1: Quantidade e percentual de sítios da Web governamental brasileira (.gov.br) com endereço IPv6 em URLs alternativas

Número total de sítios (Tab. A1-1)	Número de sítios com URLs alternativas para IPv6	% de Hosts
18.229	250	1,37

A Tabela B2-2 apresenta a relação entre as unidades federativas (UF) e o número de sítios com endereço IPv6 em URL alternativa. As UFs que não continham sítios com tal característica foram omitidas para facilitar a visualização.

Tabela B2-2: Quantidade de sítios da Web governamental brasileira (.gov.br) com endereço IPv6 em URLs alternativas segundo as unidades federativas (UF)

UF	Nº de sítios utilizando domínio alternativo
.GOV.BR	5
.outros	96
AL	3
BA	7
CE	1
ES	13
GO	3
MA	1
MG	42
MS	2
MT	3
PA	3
PB	2
PE	4
PR	6
RJ	15
RN	2
RS	4
SC	9
SP	25
TO	4

5. Análise

Na Web governamental brasileira (.gov.br) existe um volume pequeno de sítios com endereço IPv6 em URLs alternativa. Como mostrado na Tabela B2-1, apenas 1,37% dos sítios utilizam algum domínio alternativo para prover suporte IPv6. Cabe ainda ressaltar que em torno de 40% desses domínios não são .gov.br, mas sim foram alcançados por redirecionamento durante a coleta.

Observa-se, com isso, uma tendência no uso de URL alternativa para a realização de testes de acesso via IPv6. É interessante notar que os mesmos 250 sítios possuem também endereço IPv6 em sua URL principal, o que indica maior maturidade no desenvolvimento dessas soluções.

Os estados que mais possuem mais domínios alternativos com endereços IPv6 são Minas Gerais (MG), São Paulo (SP), Rio de Janeiro (RJ) e Espírito Santo (ES) – estados da região Sudeste, que juntos possuem 95 dos 154 domínios pertencentes à esfera governamental, o que equivale a 61,69%.

Há ainda 5 domínios que pertencem a sítios do .gov.br ligados à autarquia federal, que são os seguintes: www.ctir.gov.br, www.dpf.gov.br, www.serpro.gov.br, www.portalbrasil.gov.br, e www.dner-ce.gov.br.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que se iniciou a utilização desse tipo de técnica, pois antes nenhum sítio adotava IPv6.

B3: Proporção de sítios que respondem a ping IPv6

1. Definição do indicador

Valor percentual de sítios que respondem a um comando PING em seu endereço IPv6.

2. Propósito

Ter uma medida da quantidade de servidores Web encontrados que possuem na prática uma configuração mínima para operação com IPv6.

3. Metodologia

Para obter esses dados foi utilizado o programa ConNeCTOR que possui métodos para realizar consultas do tipo PING6 a servidores Web.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre sítios que respondem a PING IPv6 está apresentado em três tabelas (B3-1, B3-2 e B3-3). Observa-se que 11 dos 18229 sítios coletados, o que representa 0,06%, respondem ao protocolo IPv6. A Tabela B3-1 apresenta a relação de hosts desses sítios com suas URLs.

Tabela B3-1: Sítios que responderam ao comando PING em seu endereço IPv6

UF	URL
.outros	www.fmb.unesp.br
RJ	www.riodasostras.rj.gov.br
RJ	pmro.rj.gov.br

Tabela B3-2: Percentual dos sítios que responderam a um PING em seus endereços IPv6

% de Sítios que responderam a PING via IPv6
0,0165

A Tabela B3-3 apresenta a relação entre as unidades federativas e o número de sítios que respondem à PING em IPv6. Resultados dos estados com valores zerados foram omitidos para facilitar a visualização.

Tabela B3-3: Número de sítios que responderam à ping IPv6 agrupados por UF

UF	Nº de sítios que respondem à ping IPv6
.GOV.BR	0
.outros	1
RJ	2

5. Análise

Na Web Governamental Brasileira somente 3 sítios estavam configurados para responder a consultas mínimas (PING), usando o protocolo ICMP6 sobre IPv6. Esse valor corresponde a 0,0165% dos sítios. Cabe ainda ressaltar que um deles não é .gov.br, mas sim foi alcançado por redirecionamento durante a coleta.

Observa-se que praticamente não há resposta PING ao endereço IPv6, já que somente 2 sítios governamentais apresentaram resultado. Isso é um indicativo de que os endereços IPv6 apontados pelo DNS não estão de fato alocados nos servidores. Entretanto, cabe ressaltar que tal observação não é absoluta, pois a consulta PING através do protocolo ICMPv6 pode estar sendo bloqueado por algum sistema de segurança (firewall), já que essa prática, apesar de não ser recomendada na Internet em geral, é largamente adotada por administradores de rede.

É importante enfatizar que a verificação da resposta ao PING ICMPv6 no endereço IPv6 indica que o serviço Web do sítio possui suporte ao protocolo IPv6. A ausência de resposta indica que o servidor não estava acessível sobre o protocolo IPv6, seja pelo próprio servidor não estar configurado para usar o protocolo ou por algum elemento de rede no caminho para aquele servidor não estar configurado para aceitar esse protocolo.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que houve um aumento no número de sítios que respondem ao comando PING em seus endereços IPv6, pois antes apenas quatro sítios (todos oriundos de redirecionamento para outros domínios a partir do gov.br) respondiam ao comando PING IPv6, sendo que nesta nova análise foram 3, dos quais 2 são de extensão gov.br. Conforme já relatado em outras análises do indicador de IPv6, observa-se que o protocolo IPv6 ainda não está, de fato, implantado nos sistemas de rede e servidores da Web governamental brasileira.

B4: Proporção de sítios que respondem a requisições HTTP GET em seu endereço IPv6

1. Definição do indicador

Valor percentual de sítios da Web Governamental Brasileira questão que respondem a requisições do tipo HTTP GET na porta 80 de seu endereço IPv6

2. Propósito

Obter uma medida da atual quantidade de sítios que podem ser acessados via IPv6.

3. Metodologia

Para obter esse indicador, o programa ConNeCTOR realizou uma conexão HTTP com uma requisição do tipo HTTP HEAD diretamente ao endereço IPv6 dos sítios identificados com endereços IPv6. Caso o servidor respondesse, o sítio era considerado como conectável via IPv6, independentemente do código enviado como resposta. Na prática, esse procedimento verifica se o servidor possui a infra-estrutura básica para responder requisições Web via IPv6. O método HTTP HEAD foi escolhido por ser mais rápido do que o método HTTP GET e, assim, permitir que os testes fossem mais eficientes. Apesar dessa mudança, do ponto de vista do teste de conectividade Web, os resultados não são influenciados.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre sítios com suporte Ipv6 em domínios alternativos está apresentado em três tabelas (B4-1, B4-2 e B4-3). Somente 3 dos 18.229 sítios da Web Governamental, o que representa 0,0165%, responderam a requisições Web via IPv6. A Tabela B4-1 apresenta uma listagem desses sítios.

Tabela B4-1: Sítios que responderam a requisições HTTP GET em seu endereço IPv6

UF	URL
.outros	www.fmb.unesp.br
RJ	www.riodasostras.rj.gov.br
RJ	pmro.rj.gov.br

Tabela B4-2: Percentual de Sítios da Web Governamental Brasileira que responderam à requisição HTTP GET porta 80 de endereço IPv6

% de Sítios – HTTP GET IPv6
0,0165

A Tabela B4-3 apresenta a relação entre as unidades federativas e o número de sítios que responderam à requisição HTTP GET em seu endereço IPv6. Resultados com

valores zerados foram omitidos para facilitar a visualização.

Tabela B4-3: Número de sítios que responderam ao comando GET no endereço IPv6 agrupados por UF

UF	Nº de sítios que respondem à ping IPv6
.GOV.BR	0
.outros	1
RJ	2

5. Análise

Na Web governamental brasileira somente 3 sítios responderam ao comando GET no endereço IPv6. Esse valor corresponde a 0,0165% dos sítios. Cabe ainda ressaltar que 1 deles não é gov.br, mas sim foi alcançado por redirecionamento durante a coleta.

Observa-se que praticamente não há resposta à requisição HTTP GET ao endereço IPv6, já que somente os mesmos 3 sítios governamentais referenciados no indicador B3 apresentaram resultado. Isso sugere a hipótese de que endereços IPv6 apontados pelo DNS não estão de fato alocados nos servidores, apesar de haver ainda a possibilidade de um servidor estar configurado mas a rede de acesso a ele não estar configurada para transportar mensagens no protocolo IPv6. Todos os servidores que foram identificados como configurados com IPv6 (indicador B3) realmente continham um servidor HTTP configurado para atender requisições nessa porta (indicador B2).

Considerando-se os resultados para os indicadores B1 e B2, especialmente para o primeiro, pode-se afirmar que, mesmo com 250 servidores anunciando que possuem endereços IPv6 configurados, apenas 3 deles podiam ser acessados utilizando-se esse protocolo.

É importante enfatizar que a verificação da resposta ao comando GET no endereço IPv6 indica que o serviço Web do sítio possui suporte ao HTTP sobre o protocolo IPv6. A ausência de resposta indica que o servidor não está acessível através daquele protocolo, seja pelo próprio servidor não estar configurado para usar IPv6 ou por algum elemento de rede no caminho para aquele servidor não estar configurado para aceitar aquele protocolo.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que houve um aumento no número de sítios que atendem realmente ao protocolo IPv6, pois antes apenas 4 sítios (todos oriundos de redirecionamento a partir do gov.br, ou seja nenhum sítio .gov.br) respondiam ao comando PING IPv6, sendo que nesta nova análise foram 3, dos quais 2 são de extensão gov.br.

C1: Distribuição de idiomas das páginas Web

1. Definição do indicador

Distribuição dos idiomas das páginas da Web governamental brasileira (.gov.br).

2. Propósito

Obter informações sobre a utilização de diferentes idiomas na Web governamental brasileira (.gov.br).

3. Metodologia

A metodologia de obtenção desse indicador consiste na análise de cada uma das páginas coletadas de modo a comparar seus conteúdos com dicionários de stop words (palavras comuns em um idioma) de algumas línguas preestabelecidas. Essa classificação consiste em contar o número de palavras que uma página tem de cada um dos idiomas e, aquele que tiver o maior número será considerado o idioma das páginas. Considera-se os seguintes idiomas pré-definidos: português, inglês, espanhol e francês. Caso a página não possua palavras suficientes (50 palavras) ou caso a diferença entre o número de palavras dos idiomas mais prováveis não seja suficiente para que se possa classificá-la, seu idioma é classificado como "outros". O número de palavras suficientes para ser classificado com um idioma foi 15.

4. Apresentação dos resultados

Um resumo descritivo da distribuição dos idiomas das páginas Web está apresentado em duas tabelas (C1-1 e C1-2).

Cabe ressaltar que existe um número significativo de documentos HTML para os quais a técnica utilizada não foi capaz de identificar o idioma, esse grupo corresponde a 3.297.589 páginas (48% do total). Assim, a Tabela C1-1 apresenta a distribuição dos quatro idiomas pré-definidos (Português, Inglês, Espanhol, e Francês) e a Tabela C1-2 apresenta a distribuição desses mesmos idiomas por Unidade Federativa (UF).

Considerando as páginas nas quais foi possível a identificação, a grande maioria (97,25%) foi identificada com o idioma português, o que mostra o caráter de levar informações aos cidadãos brasileiros em sua língua nativa. Os demais idiomas possuem percentuais menores de páginas (Inglês, Espanhol e Francês).

Tabela C1-1: Distribuição dos principais idiomas das páginas Web no domínio .gov.br.

Português	% Português	Inglês	% Inglês	Espanhol	% Espanhol	Francês	% Francês
3.478.879	97,25	81.795	2,29	15.683	0,44	980	0,03

Ao observar a distribuição dos quatro idiomas pré-definidos, as UF que apresentam menor participação do idioma português em suas páginas, comparativamente as outras UF, são: Minas Gerais (MG), os sítios diretamente organizados pelo Governo Federal

(.gov.br) e Rio Grande do Norte (RN), com 95,78%, 93,71%, e 87,24%, respectivamente.

Tabela C1-2: Distribuição dos principais idiomas das páginas Web no domínio .gov.br. por UF

UF	Português	% Português	Inglês	% Inglês	Espanhol	% Espanhol	Francês	% Francês
.GOV.BR	881.737	93,71	48.982	5,21	9.575	1,02	621	0,07
.outros	680	73,91	180	19,57	50	5,43	10	1,09
AC	17.513	99,58	6	0,03	68	0,39	0	0,0
AL	131.585	98,53	1.959	1,47	9	0,01	0	0,0
AM	11.491	97,27	315	2,67	0	0,0	8	0,07
AP	11.081	99,95	4	0,04	1	0,01	0	0,0
BA	151.533	97,82	3.348	2,16	29	0,02	2	0,0
CE	99.565	99,08	877	0,87	8	0,01	35	0,03
DF	65.549	99,12	503	0,76	1	0,0	78	0,12
ES	21.021	98,22	338	1,58	42	0,2	1	0,0
GO	39.636	99,79	80	0,2	3	0,01	0	0,0
MA	28.501	98,7	373	1,29	0	0,0	3	0,01
MG	277.947	95,78	10.648	3,67	1.587	0,55	22	0,01
MS	41.664	99,18	344	0,82	0	0,0	0	0,0
MT	189.219	99,78	370	0,2	37	0,02	9	0,0
PA	63.164	99,36	290	0,46	106	0,17	8	0,01
PB	22.782	98,44	360	1,56	0	0,0	0	0,0
PE	47.598	99,83	81	0,17	2	0,0	0	0,0
PI	29.164	99,97	8	0,03	0	0,0	1	0,0
PR	247.038	99,41	1.039	0,42	409	0,16	9	0,0
RJ	156.487	98,96	1.159	0,73	465	0,29	20	0,01
RN	12.514	87,24	1.829	12,75	1	0,01	1	0,01
RO	26.644	99,98	5	0,02	1	0,0	0	0,0
RR	555	99,82	1	0,18	0	0,0	0	0,0
RS	203.444	99,41	1.068	0,52	139	0,07	7	0,0
SC	141.034	98,62	1.835	1,28	107	0,07	32	0,02
SE	85.436	99,08	744	0,86	45	0,05	1	0,0
SP	453.785	98,33	4.676	1,01	2.950	0,64	63	0,01
TO	20.512	97,76	373	1,78	48	0,23	49	0,23

5. Análise

Em relação as páginas dos sítios vinculados ao Governo Federal (.gov.br), além do português, o inglês foi bastante observado nos sítios bd.camara.gov.br e saaldemo.interlegis.gov.br. Em relação à UF de MG, nesse mesmo quesito, destacaram-se os sítios da cidade de Dom Silvério (www.domsilverio.mg.gov.br), da cidade de Belo Horizonte (www.belo Horizonte.mg.gov.br) e das secretarias do governo do estado de Minas Gerais acessadas por dois domínios (www.desenvolvimento.mg.gov.br e www.sede.mg.gov.br). E, quanto ao governo do RN, observa-se maior número de páginas em inglês no sítio da cidade de São José de Mipibu (www.saojosedemipibu.rn.gov.br).

Considerando-se as páginas em Espanhol, nota-se maior presença nos sítios da cidade de Belo Horizonte (www.belo Horizonte.mg.gov.br) e da Secretaria de

Desenvolvimento Econômico de Minas Gerais (www.desenvolvimento.mg.gov.br).

Cabe ressaltar que para conteúdos de páginas Web coletadas de outros domínios (em função de redirecionamento), existe maior incidência de conteúdo em idioma estrangeiro, cujo valor é 26,09%.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), tem-se um cenário similar. A análise anterior havia identificado o idioma de 47,8% das páginas Web. Comparando-se a distribuição por idiomas os percentuais também são bastante próximos: português 97,05% (2010) e 97,25%; inglês 2,54% (2010) e agora há 2,29% de páginas; espanhol estava presente em 0,33% das páginas, e nesta edição em 0,44; por fim, o francês estava com 0,08% das páginas e agora está com 0,027%.

D1: Localização geográfica dos servidores de sítios

1.1 Definição do indicador

Distribuição da localização geográfica dos servidores que hoperam sítios da Web governamental brasileira (.gov.br).

1.2 Propósito

Obter uma distribuição percentual dos sítios da Web governamental brasileira (.gov.br) de acordo com o país onde está localizado o servidor desses sítios.

1.3 Metodologia

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

1.4 Apresentação dos resultados

Um resumo descritivo dos dados sobre localidade geográfica está apresentado em duas tabelas. (D1-1 e D1-2)

A Tabela D1-1 consolida essas informações de geolocalização, exibindo as 10 localidades mais frequentes. A Tabela D1-2 apresenta a distribuição percentual das principais localidades geográficas por unidade federativa (UF).

Tabela D1-1 - Lista de informações das 10 localidades geográficas mais frequentes dos sítios

Geolocalização	Nº de sítios	% do total
BR	14.499	87,38
US	1.461	8,80
CA	196	1,18
AR	40	0,24
DE	39	0,24
ES	29	0,17
GB	22	0,13
FR	19	0,11
IT	18	0,11
PT	14	0,08

Tabela D1-2 - Distribuição percentual das principais localidades geográficas dos sítios por UF

Estado	Total	% BR	% US	% CA
.GOV.BR	2.413	98,14	1,74	0,04
.outros	2144	53,87	25,33	1,21
AC	56	98,21	1,79	0
AL	185	95,68	4,32	0
AM	202	97,52	1,98	0,5
AP	70	95,71	4,29	0
BA	589	92,87	6,79	0,17
CE	492	92,89	6,3	0,61
DF	182	98,9	1,1	0
ES	370	91,35	8,38	0,27
GO	264	90,53	8,33	0,76
MA	163	92,64	4,29	0,61
MG	1.099	84,62	14,38	0,82
MS	263	93,54	6,46	0
MT	209	91,39	8,61	0
PA	247	93,52	5,67	0
PB	238	86,13	13,45	0,42
PE	270	89,63	9,26	0,37
PI	133	98,5	1,5	0
PR	1.613	89,71	2,29	8,00
RJ	705	90,21	8,65	1,13
RN	229	89,52	9,61	0,87
RO	123	97,56	2,44	0
RR	56	96,43	3,57	0
RS	762	95,93	4,07	0
SC	940	95,32	4,68	0
SE	233	94,85	4,72	0,43
SP	2.162	88,25	11,19	0,42
TO	182	95,6	4,4	0

1.5 Análise

Os dados foram obtidos e validados para 16.594 sítios, o que equivale a 91% do total de 18.229 sítios da Web governamental brasileira (.gov.br). A análise das informações de localização geográfica mostra que a grande maioria dos sítios estão hospedados em território brasileiro(87,4%). Há uma outra parcela em território norte-americano (8,8%). Em seguida aparece o Canadá com 1,2%. Esses 3 mais frequentes, representam 97,4% dos sítios. Entre os 10 mais frequentes ainda aparecem Argentina, Alemanha, Espanha, Grã-Bretanha, França, Itália e Portugal. Esses 10 países hospedam 98,44% da web brasileira, deixando 1,56% dos sítios residentes em outros países.

A avaliação por unidade federativa mostra que em 6 estados brasileiros existem mais de

10% dos sítios hospedados em território estrangeiro. O maior volume de sítios hospedado fora do Brasil pertence a Minas Gerais, com 15,38%, cuja maioria absoluta (14,38%) está hospedada nos Estados Unidos. Os outros estados com mais de 10% dos sítios hospedados fora do país são Paraíba (13,87%), São Paulo (11,75%), Rio Grande do Norte (10,48%), Pernambuco (10,37%) e Paraná (10,29%). Nesse grupo todos possuem a grande maioria de seus sítios hospedados em território norte-americano, exceto Paraná, que possui 8% dos sítios hospedados no Canadá.

Após realizar uma análise detalhada dos sítios nos 6 estados brasileiros que se destacaram por ter volume significativo de seus servidores em território estrangeiro, pode-se observar o seguinte:

- No caso do estado de Minas Gerais (MG), são 158 sítios hospedados nos Estados Unidos. A grande parte se refere a municípios do estado de MG (como, por exemplo, miradouro.mg.gov.br e www.manhumirim.mg.gov.br) e também sítios de câmaras legislativas (como, por exemplo: camaraconceicao.mg.gov.br e www.camaraabaete.mg.gov.br). Há ainda servidores de *email* (*webmails*) e alguns sítios temporários, como oficinadetravessias.mg.gov.br.
- No caso do estado da Paraíba (PB), são 32 sítios hospedados nos Estados Unidos, sendo grande parte de municípios do estado. Há ainda alguns órgãos, como cmc.pb.gov.br e www.proconcg.pb.gov.br.
- No caso do estado de São Paulo (SP), são 242 sítios hospedados no Estados Unidos, sendo boa parte de câmaras municipais (como, por exemplo: www.camaraaguasdesaopedro.sp.gov.br, www.camaraaluminio.sp.gov.br e www.camaraandalia.sp.gov.br) e alguns municípios (como, por exemplo: www.itaoca.sp.gov.br, www.itapecerica.sp.gov.br, www.itapevi.sp.gov.br, www.itapui.sp.gov.br, www.jacarei.sp.gov.br e www.jau.sp.gov.br). Além disso, ainda existem alguns outros sítios, como www.dadsnh.sp.gov.br e www.cbpm.sp.gov.br. E existem ainda vários serviços de *webmail*.
- No caso do estado do Rio Grande do Norte (RN), são 22 sítios, são municípios principalmente, além do sítio principal do estado (www.rn.gov.br) e outros sítios estaduais, como www.rncontradengue.rn.gov.br e www.agenciarn.rn.gov.br.
- No caso do estado de Pernambuco (PE), são 25 sítios, a maioria são municípios do estado, sendo alguns sítios pertencentes ao estado de Pernambuco, como www.caspm.pe.gov.br. Há sítios temporários como www.fenearte.pe.gov.br e www.carnaval.olinda.pe.gov.br.
- Por fim, no caso do estado do Paraná (PR), são 37 sítios em território norte-americano, sendo a grande maioria de municípios paranaenses e de câmaras municipais dos municípios. Mas merece destaque que há 129 sítios, cujos servidores estão hospedados no Canadá. E muitos são municípios, sendo que hospedam seu sítios principal e servidores de nomes auxiliares, usando prefixo ns1 e ns2, como, por exemplo: ns1.assischateaubriand.pr.gov.br, ns1.brasilandiosul.pr.gov.br, ns1.cafelandia.pr.gov.br, ns1.california.pr.gov.br e ns1.doutorcarnargo.pr.gov.br.

É difícil justificar a razão pela qual se verifica esse volume significativo de servidores de sítios hospedados em território estrangeiro. Uma possibilidade pode ser pelo uso de serviços Web disponibilizados na Web estrangeira (ferramentas ou tecnologias) que está sendo adotado em sítios governamentais brasileiros. Outras possibilidades são o custo do hospedagem do sítio que pode ser menor no exterior do que no Brasil, e a

conectividade ao restante da internet, que pode ser melhor no exterior, traduzindo em tempos de acessos menores e mais consistentes.

1.6 Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que a proporção de sítios .gov.br hospedados em território brasileiro reduziu de 93% para 87,4%, dobrando o percentual hospedado em território estrangeiro (de aproximadamente 6% para 12%).

E1: Proporção de páginas Web aderentes aos padrões HTML do W3C

1. Definição do indicador

Valor percentual de páginas HTML da Web governamental brasileira (.gov.br) que atendem aos padrões W3C.

2. Propósito

Avaliar a qualidade das páginas HTML da Web governamental brasileira (.gov.br) segundo sua conformidade com os padrões HTML e suas variantes especificadas pelo W3C.

3. Metodologia

Foi utilizada uma instância local do validador de páginas disponibilizado como software livre pelo W3C para identificar o tipo de documento e informar o número de erros encontrados. Com ele foi possível validar a maior parte das linguagens de marcação utilizadas na Web, o que inclui as famílias HTML e XHTML, MathML, SMIL e SVG (1.0 and 1.1). Como resposta a uma validação, o sistema fornece o número total de erros e avisos encontrados, além de suas respectivas descrições, as quais não foram utilizadas no presente projeto. Além da quantidade de erros por páginas, algumas páginas Web não puderam ser verificadas, em geral devido a uma extensão indevida ou a erros que impossibilitaram a identificação da codificação dos caracteres em algum ponto do documento. O valor zero (0) indica página válida, na qual nenhum erro ou resultados inesperados foram encontrados, ou seja, é uma página que passou na validação de HTML W3C.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre validação W3C das páginas Web está apresentado em quatro tabelas (E1-1, E1-2, E1-3, e E1-4).

A Tabela E1-1 apresenta os resultados gerais obtidos, pode-se observar que apenas 6,38% das páginas estão isentas de erros.

Tabela E1-1: Distribuição das páginas segundo quantidade de erros

Não Verificável	0 (%)	>0 (%)
7,17	6,38	86,45

A Tabela E1-2 apresenta uma distribuição da quantidade de erros por página,

considerando faixas de valores para propiciar uma melhor visualização de seu grau de incorreção. A maior parte das páginas apresentaram até 10 erros (30,89%).

Tabela E1-2: Distribuição das páginas segundo faixa de quantidade de erros (percentual)

> 0 e ≤10	≥10 e <20	≥20 e <30	≥30 e <40	≥40 e <50	≥ 50 e < 60	≥60 e <70	≥70 e <80	≥80 e <90	≥90 e <100	≥100
30,89	11,01	8,59	4,99	3,78	3,5	2,91	2,79	2,17	1,65	14,17

A Tabela E1-3 apresenta estatísticas gerais acerca do indicador de validação W3C das páginas da Web. A média de erros encontrados nas páginas foi de 63,11 e o número máximo de erros encontrados foi de 27.665.

Tabela E1-3: Medidas resumo para a quantidade de erros

Média	Min	Max	Mediana	Desvio Padrão	CV
63,11	0	27.665	22	198,02	3,14

A Tabela E1-4 apresenta um recorte da validação W3C por unidades federativas (UF). Os estados de Minas Gerais, Maranhão e Tocantins merecem destaque por serem os 3 que apresentaram melhor conformidade, com percentual de páginas válidas acima de 10%.

Tabela E1-4: Conformidade das páginas por Unidade Federativa (UF)

UF	Não Conformidade	% Não Conformidade	Conformidade	% Conformidade
.GOV.BR	1.436.451	93,78	95.200	6,22
.outros	2.006	97,43	53	2,57
AC	42.818	95,79	1.881	4,21
AL	266.883	99,49	1.367	0,51
AM	35.546	99,01	355	0,99
AP	13.092	99,2	105	0,8
BA	305.152	95,47	14.470	4,53
CE	272.394	93,17	19.959	6,83
DF	118.332	99,61	468	0,39
ES	54.207	93,96	3.486	6,04
GO	78.164	92,51	6.332	7,49
MA	36.765	83,82	7.096	16,18
MG	491.694	83,24	98.999	16,76

UF	Não Conformidade	% Não Conformidade	Conformidade	% Conformidade
MS	66.669	93,56	4.592	6,44
MT	261.174	99,12	2.327	0,88
PA	75.000	98,57	1.089	1,43
PB	61.141	93,32	4.378	6,68
PE	83.757	97,43	2.206	2,57
PI	46.226	95,56	2.149	4,44
PR	548.177	93,68	36.995	6,32
RJ	252.704	91,31	24.058	8,69
RN	48.991	91,09	4.792	8,91
RO	51.501	94,23	3.151	5,77
RR	2.388	98,96	25	1,04
RS	410.423	94,67	23.108	5,33
SC	380.711	93,63	25.891	6,37
SE	154.559	97,71	3.618	2,29
SP	799.965	94,69	44.817	5,31
TO	39.446	87,52	5.623	12,48

5. Análise

Como pode-se observar nos resultados apresentados, o volume de páginas com erros é elevado e, apenas 6,38% do total foram consideradas páginas validadas sem erros. Nota-se, também, que existe uma quantidade significativa de páginas com menos de 10 erros (37,27%), entretanto ainda existem outros 14,17% que apresentaram mais de 100 erros no procedimento de validação. Isso sugere que existe um montante significativo de páginas que podem ser mais facilmente corrigidas para ficarem compatíveis com os padrões do W3C. Destacam-se, em termos de maior taxa de conformidade, os estados do Maranhão, Tocantins e Minas Gerais. No estado do Maranhão, os sítios com maior número de páginas que atendem aos critérios de conformidade são da secretaria de Cultura (www.cultura.ma.gov.br/ - 2194 páginas) e da cidade de Caxias (caxias.ma.gov.br/ - 723 páginas), enquanto que em Minas Gerais temos a secretaria de Ciência e Tecnologia (www.tecnologia.mg.gov.br/ - 6927) e secretaria de Desenvolvimento Econômico (www.tecnologia.mg.gov.br/ - 6359). Em Tocantins se destaca a Secretaria de Saúde, que é acessível por 3 sítios: visa.to.gov.br, saude.to.gov.br e atencaprimaria.to.gov.br.

É importante destacar que os erros apontados não significam que os sítios não funcionem, mas indicam que eles não estão preparados para uma permitir um melhor desempenho das aplicações de visualização de conteúdo. Isso sugere a importância de se criar políticas e incentivos ao aumento do grau de aderência aos padrões Web.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira

(2010), constata-se que ocorreu diminuição no volume médio de erros por página, antes 82,23 e agora foi de 63,11. O que também pode ser confirmado pelo valor da mediana que passou de 31 para 22. A quantidade de erros medida na validação das páginas da Web foi elevada, porém não foi possível a contabilização dos tipos de erros mais frequentes, o que permitiria uma análise mais aprofundada das principais dificuldades na padronização das páginas. Além disso, a validação é restritiva, de forma que um erro pode se propagar ao longo de uma página, causando o aparecimento de novos erros. Isso é uma característica intrínseca do validador utilizado e, conseqüentemente, das linguagens de marcação.

Uma possível explicação para a melhora na aderência aos padrões HTML do W3C é a adoção de ferramentas de publicação e gestão de conteúdo pelas entidades governamentais, o que possibilita uma melhor padronização e um melhor uso dos recursos tecnológicos de forma a atender às recomendações definidas pelo W3C.

F1: Proporção de Páginas Web aderentes aos padrões de acessibilidade E-MAG

1. Definição do indicador

Valor percentual de páginas Web em conformidade com os padrões determinados pelo E-MAG testados através de uma versão modificada da ferramenta ASES. Esse teste considera os níveis de prioridade 1, 2 e 3 baseados nos níveis de conformidade A, AA e AAA do E-MAG 2.0 que, por sua vez, foram adaptados do WCAG 1.0 .

2. Propósito

Avaliar a qualidade das páginas HTML da Web governamental brasileira (.gov.br) em relação à sua acessibilidade.

3. Metodologia

Para avaliar a acessibilidade dos sítios da Web governamental brasileira (.gov.br) executou-se, para cada uma das páginas coletadas, o validador ASES que, entre outras coisas, verifica a conformidade de um documento com o padrão de acessibilidade de páginas do governo brasileiro E-MAG.

Segundo os níveis de conformidade do W3C, um sítio pode ser classificado em quatro níveis de acessibilidade: não acessível, A, AA ou AAA. Sendo que o nível 'A' possui os pontos que os criadores de conteúdo Web devem satisfazer inteiramente. Se não o fizerem, um ou mais grupos de usuários ficam impossibilitados de acessar as informações contidas no documento. A satisfação desses pontos é um requisito básico para que determinados grupos de pessoas sejam capazes de acessar documentos disponíveis na Web. O nível 'AA' contém os pontos que os criadores de conteúdos na Web deveriam satisfazer. Se não o satisfizerem, um ou mais grupos de usuários terão dificuldades em acessar as informações contidas no documento. A satisfação desses pontos promove a remoção de barreiras significativas ao acesso a documentos disponíveis na Web. E, o 'AAA', contém os pontos que os criadores de conteúdos na Web podem satisfazer. Se não o satisfizerem, um ou mais grupos de usuários poderão se deparar com algumas dificuldades em acessar informações contidas nos documentos. A satisfação deste tipo de pontos melhora o acesso a documentos armazenados na Web. Contudo, para que um site possa ser classificado, todas as suas páginas devem passar pelo processo de validação do ASES, que fornece a quantidade de erros para funcionalidade de 3 níveis de prioridades, que representam os níveis de conformidade do EMAG 2.0 (prioridade 1 - P1, prioridade 2 - P2, e prioridade 3 - P3). Assim, caso todas as suas páginas não apresentem erros no primeiro nível de prioridade, ele é considerado um site 'A'. Se, todas as páginas não apresentarem erros para os níveis de prioridade 1 e 2, o sítio é considerado 'AA'. E, caso todas as páginas não contenham erros em todos os três níveis de prioridade o sítio é classificado como 'AAA'. Contudo, se qualquer das páginas do sítio apresentar erros, ele é classificado como não acessível.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre a aderência das páginas Web aos padrões de acessibilidade E-MAG está apresentado em duas tabelas (F1-1 e F1-2).

A Tabela F1-1 apresenta os resultados para a avaliação das páginas HTML analisadas em relação à acessibilidade, de acordo com a metodologia aplicada.

Tabela F1-1: Nível de Acessibilidade

sem erros	% sem erros	não atende P3	% não atende P3	não atende P3 e P2	% não atende P3 e P2	não conformidade	% não conformidade
67.166	0,98	9.147	0,13	255.326	3,71	6.543.287	95,18

A Tabela F1-2 apresenta os resultados obtidos para avaliação da acessibilidade das páginas Web analisadas agrupadas por Unidade Federativa (UF).

Tabela F1-2: Nível de Acessibilidade (Páginas Web) - Recorte por Unidade Federativa (UF)

UF	sem erros ASES	não atende P3	não atende P3 e P2	Não Conformidade	% Não Conformidade
.GOV.BR	14.911	5.118	82.342	1.424.981	93,04
.outros	13	0	50	1.993	96,79
AC	0	192	947	43.541	97,41
AL	733	1	1.076	266.431	99,32
AM	10	25	19	35.841	99,83
AP	0	0	91	13.092	99,2
BA	7	0	3.038	315.280	98,64
CE	786	1	6.116	285.435	97,63
DF	1.021	664	5.502	111.535	93,88
ES	11	6	1.871	55.657	96,47
GO	8	14	2.791	81.608	96,58
MA	124	4	180	40.631	92,64
MG	1.014	296	22.697	566.178	95,85
MS	1	0	31	71.181	99,89
MT	612	0	4.899	257.704	97,8
PA	10.443	7	11.086	54.539	71,68
PB	285	0	1.323	63.760	97,32
PE	157	235	1.064	84.495	98,29
PI	53	0	298	48.014	99,25
PR	8.473	459	55.656	518.721	88,64
RJ	333	20	2.338	273.935	98,98
RN	12	0	106	53.663	99,78
RO	10	0	746	53.865	98,56

UF	sem erros ASES	não atende P3	não atende P3 e P2	Não Conformidade	% Não Conformidade
RR	2	0	21	2.389	99,01
RS	3.373	1.496	7.469	420.892	97,08
SC	1.779	6	9.580	392.271	96,48
SE	15	0	409	157.745	99,73
SP	3.696	601	31.152	805.348	95,33
TO	3	2	2.428	42.562	94,44

5. Análise

A análise da acessibilidade para o universo de domínios da Web governamental brasileira (.gov.br) mostra que ainda existe muito a ser feito para que se aperfeiçoe seu nível de acessibilidade, pois apenas aproximadamente 4,8% das páginas atendem a algum dos níveis de acessibilidade (níveis de acessibilidade A, AA ou AAA), segundo os critérios do padrão E-MAG.

Em relação aos resultados com recorte por Unidade Federativa, observa-se que a grande maioria dos Estados possuem um nível de não conformidade em relação à acessibilidade próximo à média geral. Entretanto, 2 estados se sobressaem nesse indicador: o Paraná (PR) que apresenta 88,64% de não conformidade e o Pará (PA), que apresenta 71,68%.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que houve uma melhora no nível de acessibilidade das páginas Web, já que o nível de não conformidade passou de 98% para 95,18%. Poderá ser de grande utilidade identificar que sítios estão alcançando melhor acessibilidade a fim de disseminar esses exemplos para que o nível de acessibilidade possa melhorar a cada ano.

F2: Proporção de sítios aderentes aos padrões de acessibilidade E-MAG

1. Definição do indicador

Valor percentual de sítios da Web governamental brasileira (.gov.br) em conformidade com os padrões determinados pelo E-MAG testados através de uma versão modificada da ferramenta ASES. Esse teste considera os níveis de conformidade A, AA, AAA para sítios e os níveis de prioridade 1, 2 e 3 para as páginas Web.

2. Propósito

Avaliar a qualidade dos sítios da Web governamental brasileira (.gov.br) em relação à sua acessibilidade.

3. Metodologia

Para avaliar a acessibilidade dos sítios da web governamental brasileira executou-se, para cada uma das páginas coletadas, o validador ASES que, nesse caso, verificou a conformidade de um documento com o padrão de acessibilidade de páginas do governo brasileiro E-MAG.

Segundo esse padrão, um sítio Web pode ser classificado em quatro níveis de acessibilidade: não acessível, A, AA ou AAA. Sendo que o nível 'A' indica que o sítio possui as características mínimas para que pessoas com algum tipo de deficiência consigam acessar o sítio. O 'AA', que o sítio possui características que tornem sua utilização facilitada por pessoas com certos tipos de deficiências. E, o 'AAA', que o sítio possui características que tornem sua utilização facilitada por pessoas com qualquer tipo de deficiência. Contudo, para que um sítio possa ser classificado com algum desses níveis, todas as suas páginas devem passar pelo processo de validação que fornece a quantidade de erros para funcionalidade de 3 níveis de prioridades (prioridade 1 - P1, prioridade 2 - P2, e prioridade 3 - P3). Assim, caso todas suas páginas não apresentem erros no primeiro nível de prioridade, ele é considerado um sítio 'A'. Se, todas as páginas não apresentarem erros para os níveis de prioridade 1 e 2, o sítio é considerado 'AA'. E, caso todas as páginas não contenham erros em todos os três níveis de prioridade o sítio é classificado como 'AAA'. Finalmente, se qualquer das páginas do sítio apresentar erros, ele é classificado como não acessível.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre a aderência dos sítios aos padrões de acessibilidade E-MAG está apresentado em duas tabelas (F2-1 e F2-2).

A Tabela F2-1 apresenta os resultados para a avaliação dos sítios em relação à acessibilidade, de acordo com a metodologia aplicada.

Tabela F2-1: Nível de Acessibilidade

nível AAA	% AAA	nível AA	% AA	nível A	% A	não conformidade	% não conformidade
32	0,18	2	0,01	49	0,27	18.146	99,54

A Tabela F2-2 apresenta os resultados obtidos para avaliação da acessibilidade dos sítios analisados por Unidade Federativa (UF).

Tabela F2-2: Indicador F2 - Nível de Acessibilidade (Sítios) - Recorte por Unidade Federativa (UF)

UF	nível AAA	nível AA	nível A	Não Conformidade	% Não Conformidade
.GOV.BR	4	1	10	3.089	99,52
.outros	11	0	8	2.241	99,16
AC	0	0	0	68	100
AL	0	0	0	199	100
AM	0	0	0	225	100
AP	0	0	0	77	100
BA	1	0	0	647	99,85
CE	0	0	0	527	100
DF	0	0	0	232	100
ES	0	0	0	384	100
GO	0	0	0	285	100
MA	0	0	0	176	100
MG	0	0	0	1.181	100
MS	0	0	0	275	100
MT	0	0	0	233	100
PA	0	0	2	260	99,24
PB	0	0	0	271	100
PE	0	0	1	288	99,65
PI	0	0	0	152	100
PR	9	1	15	1.623	98,48
RJ	1	0	2	779	99,62
RN	0	0	0	260	100
RO	0	0	2	129	98,47
RR	0	0	0	61	100
RS	1	0	1	801	99,75
SC	3	0	1	988	99,6
SE	0	0	1	241	99,59
SP	2	0	6	2.259	99,65
TO	0	0	0	195	100

5. Análise

A análise da acessibilidade mostra que ainda existe muito a ser feito para que se aperfeiçoe a acessibilidade da Web governamental brasileira (.gov.br), pois somente 0,46% dos sítios atenderam a algum dos níveis de acessibilidade (A, AA, e AAA) do padrão E-MAG. Esse resultado é mais restritivo do que o para as páginas Web, uma vez que um sítio perde sua classificação de acessibilidade caso qualquer de suas páginas não seja considerada acessível. Em relação aos resultados por UF, observa-se que a quase totalidade dos Estados possui uma porcentagem de não conformidade em relação à acessibilidade próxima à média geral. Somente 2 Estados se destacam neste grupo, mas com valor apenas um pouco melhor, que são Rondônia (RO) com 98,47% de não conformidade e Paraná (PR), com 98,48%. No caso do Paraná, atenderam

todos os níveis de prioridade (AAA) os sítios www.comic.pr.gov.br, www.conpen.pr.gov.br, www.paranavai.pr.gov.br, www.comel.pr.gov.br, www.comem.pr.gov.br, www.doesuavoz.pr.gov.br, entre outros.

Uma possível explicação para se ter uma maior proporção de conformidade em relação aos padrões de acessibilidade é a adoção de ferramentas de publicação e gestão de conteúdo pelas entidades governamentais, o que possibilita uma melhor padronização e uso dos recursos tecnológicos de forma a atender às recomendações definidas para atender aos quesitos de acessibilidade.

Porém observa-se claramente que existe uma grande necessidade de criar mecanismos que orientem e favoreçam com que as entidades governamentais melhor avaliem seus sítios e adotem metodologias para elaboração de seus conteúdos com mais qualidade em termos de acessibilidade.

6. Análise Comparativa com estudo anterior e considerações

Na pesquisa anterior da Web governamental brasileira (2010) não foi avaliada a acessibilidade para os sítios, então esse resultado fica sendo a primeira análise divulgada para esse indicador.

G1: Distribuição de tipos de objetos na Web

- **Definição do indicador**

Valor percentual dos tipos de objetos usados nas páginas da Web governamental brasileira (.gov.br), de acordo com uma classificação categórica (imagens, scripts, vídeos etc.).

- **Propósito**

Obter uma distribuição dos tipos de objetos usados nas páginas da Web governamental brasileira (.gov.br) de acordo com uma categoria pré-definida (imagens, scripts, vídeos, etc.).

- **Metodologia**

Páginas Web foram coletadas com a utilização do Web crawler, considerando tanto a URL de cada página coletada quanto as URLs presentes em cada página coletada. Todas as extensões foram convertidas para os grupos listados na Tabela G1-1.

A taxonomia de tipos de documentos foi extraída da e-Ping, Padrões de interoperabilidade de Governo Eletrônico, Documento de Referência Versão 2.0, 11 de Dezembro de 2009.

- **Apresentação dos resultados**

Um resumo descritivo dos dados sobre distribuição dos tipos de objetos na Web governamental brasileira (.gov.br) está apresentado em três tabelas (G1-1, G1-2 e G1-3).

A Tabela G1-1 apresenta a quantidade e percentual de objetos nas páginas HTML, de acordo com algumas extensões principais separadas por tipo ou grupo de documentos (apresentação, áudio e vídeo, documento, etc.). Como pode ser observado, a extensão mais frequente para documento tipo apresentação é PPT (e PPTX), que somam 97,7%. Já para áudio e vídeo, MP3 e MP4 somam mais de 95%. Já para documentos, quase 85% destes são PDF. Em relação a imagens (gráficos), GIF, JPG e PNG representam 52,6%, 27,7% e 19,5%, respectivamente. Para hipertexto, htm e html somam mais de 98%. E a parte de planilhas tem mais de 99% com uso de XLS.

Tabela G1-1 - Quantidade e percentual de objetos nas páginas Web, por tipos de documentos

Grupo	Extensão	Quantidade	% de uso da extensão no grupo
Apresentação	Odp	348	2,2815
Apresentação	Pptx	266	1,7439
Apresentação	Ppt	14.639	95,9746
Áudio e Vídeo	Avi	8.481	0,6379
Áudio e Vídeo	Mid	19	0,0014

Grupo	Extensão	Quantidade	% de uso da extensão no grupo
Áudio e Vídeo	mp3	203.009	15,2702
Áudio e Vídeo	Mpg	3.591	0,2701
Áudio e Vídeo	Ogg	413	0,0311
Áudio e Vídeo	Swf	1.047.703	78,8075
Áudio e Vídeo	Wav	652	0,049
Áudio e Vídeo	mp4	65.578	4,9327
Banco de Dados	Csv	2.527	100,0
Documento	Doc	417.941	8,6998
Documento	Xsl	36	0,0007
Documento	Txt	20.943	0,4359
Documento	Rtf	13.209	0,275
Documento	Pdf	4.345.821	90,4619
Documento	Odt	3.005	0,0626
Documento	Docx	3.078	0,0641
Gráficos	Tif	5.086	0,0048
Gráficos	Svg	5.463	0,0052
Gráficos	Png	20.524.638	19,4992
Gráficos	Odg	15	0,0
Gráficos	Jpg	29.147.114	27,6909
Gráficos	Jpeg	109.350	0,1039
Gráficos	Gif	55.407.835	52,6396
Gráficos	Bmp	59.315	0,0564
Hipertexto	XML	331.168	2,2783
Hipertexto	Xhtml	12.458	0,0857
Hipertexto	Shtml	164.108	1,129
Hipertexto	RSS	60.512	0,4163
Hipertexto	Mht	1.615	0,0111
Hipertexto	HTML	10.419.371	71,6822
Hipertexto	Htm	3.546.272	24,3973
Planilha	Ods	454	0,3234
Planilha	Xls	139.699	99,5255
Planilha	Xlsx	212	0,151

A Tabela G1-2 apresenta a distribuição de objetos em termos de seus grupos ou tipos. Observa-se que uma grande parte dos objetos são de tipo gráfico (84,2%), seguido de hipertexto (HTMLs) com 11,4% e em seguida documentos (4,1%).

Tabela G1-2 - Distribuição de objetos por grupos ou tipos

Grupos ou tipo	Quantidade	% de uso do Grupo
Apresentação	15.253	0,0121
Áudio e Vídeo	1.329.446	1,0544
Banco de Dados	2.527	0,002
Documento	4.804.033	3,8101
Gráficos	105.258.816	83,4818
Hipertexto	14.535.504	11,5283

Grupos ou tipo	Quantidade	% de uso do Grupo
Planilha	140.365	0,1113

A Tabela G1-3 apresenta a distribuição dos tipos de objetos por Unidade Federativa (UF).

Tabela G1-3 - Distribuição para cada tipo de objeto por Unidade Federativa (UF)

Estado	Apresentação	Áudio e Vídeo	Banco de Dados	Documentos	Gráficos	Hipertexto	Planilha
.GOV.BR	0,02	0,43	0,01	4,22	82,35	12,65	0,33
.outros	0,01	1,29	0,00	2,98	69,34	26,29	0,09
AC	0,00	1,21	0,00	2,97	91,27	4,55	0,01
AL	0,01	0,48	0,00	3,69	95,67	0,14	0,01
AM	0,00	0,74	0,00	3,51	60,96	34,78	0,01
AP	0,00	2,12	0,00	18,44	64,15	15,28	0,00
BA	0,00	1,57	0,00	2,87	89,57	5,97	0,02
CE	0,03	0,67	0,00	8,46	85,18	5,61	0,06
DF	0,01	0,46	0,00	3,51	94,00	1,98	0,05
ES	0,01	1,23	0,00	2,97	67,43	28,35	0,02
GO	0,00	0,51	0,00	6,93	85,11	7,34	0,11
MA	0,01	0,73	0,00	5,72	88,12	5,40	0,02
MG	0,01	0,62	0,00	4,63	64,22	30,50	0,03
MS	0,01	1,97	0,00	1,44	89,11	7,47	0,00
MT	0,00	1,78	0,00	2,12	93,09	2,96	0,04
PA	0,26	0,53	0,00	4,62	90,73	3,86	0,01
PB	0,01	3,58	0,00	2,60	78,81	14,91	0,09
PE	0,00	2,79	0,00	2,99	81,87	12,32	0,02
PI	0,00	15,53	0,00	4,72	75,03	4,71	0,01
PR	0,00	1,25	0,00	2,67	94,80	1,25	0,02
RJ	0,01	0,55	0,00	4,28	84,34	10,79	0,02
RN	0,01	0,10	0,00	2,55	59,20	38,13	0,01
RO	0,00	1,16	0,00	2,17	69,40	27,23	0,02
RS	0,01	0,83	0,00	3,51	91,54	4,05	0,05
SC	0,02	1,89	0,00	3,24	89,49	5,29	0,08
SE	0,00	0,65	0,00	0,73	79,83	18,79	0,00
SP	0,01	0,90	0,00	4,59	72,50	21,83	0,18
TO	0,00	0,49	0,00	3,85	89,58	6,07	0,00

- **Análise**

As análises realizadas permitem entender melhor o perfil mais comum de páginas da Web governamental brasileira (.gov.br) de acordo com a composição de tipos de objetos que estas apresentam. É importante ressaltar a diversidade tanto em termos de padrão tecnológico quanto de distribuição de tipos de documentos, o que permite concluir apenas que o nível de heterogeneidade observado nos impede de extrair regras gerais que sintetizem o cenário.

Cabe ressaltar que existe predominância de objetos do tipo **Gráfico** tanto nas Unidades Federativas (UF) quanto em sítios do Governo Federal (com variação de 59,2% – RN - a 95,7% - AL), seguido de **Hipertexto**, cuja variação 0,14% (AL) a 38,1% (RN). Além desses aspectos, casos que merecem destaque são Piauí com 15,5% de objetos do tipo **Áudio e Vídeo**, e Amapá com 18,4% de objetos do tipo **Documento**.

- **Análise Comparativa com estudo anterior e considerações**

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (.gov.br) realizada em 2010, observa-se algumas pequenas mudanças, em especial: aumento do volume de HTM / HTML em tipo hipertexto de 97 para 96%; o uso de ODP no tipo “apresentação” subiu de 0,58 para 2,28%, apesar da grande maioria de uso de PPT ainda; na parte de documentos, o uso de objetos PDF subiu de 80,2 para 90,5%; na parte de áudio e vídeo, o uso de MP3 e MP4 diminuiu de 97,8% para 20,2%, mas cabe explicar que passou a ser contabilizado o formato SWF, que por não ser do padrão e-Ping antes não era considerado no cálculo, mas representa 78,8% dessa categoria de objeto; por fim, na parte de gráficos, observa-se que houve uma mudança significativa - o uso de JPG se manteve estável (aproximadamente 27%), mas dobrou o uso de PNG (passando de 9,5% para 19,5%) com diminuição no volume do uso de GIF.

Com base nos resultados apresentados, pode-se estabelecer algumas conclusões importantes em termos de tendências observadas na comparação das pesquisas:

- O uso de tecnologias software livre no meio governamental tem crescido um pouco, como observado em relação a objetos de apresentação e documentos;
- A adoção do formato PDF também apresentou crescimento, podendo observar que esse tipo de documento tem representativa adoção para divulgação de documentos na Web;
- O uso do formato gráfico PNG cresceu significativamente, seguindo a tendência de adoção de formatos mais leves em ambiente Web.

G2: Distribuição de tipos de tecnologias adotadas nas páginas Web

- **Definição do indicador**

Valor percentual dos tipos de tecnologias usados nas páginas da Web governamental brasileira (.gov.br).

- **Propósito**

Obter uma distribuição dos tipos de tecnologias usados nas páginas da Web governamental brasileira (.gov.br).

- **Metodologia**

A determinação de tecnologias usadas é um desafio porque uma página coletada não possui obrigatoriamente informações sobre a tecnologia que a gerou. A opção adotada foi a de se basear nas eventuais extensões de arquivo presentes na URL. Assim, partise, de uma dicionário de 406 extensões de arquivos e processamos o arquivo de páginas válidas (OK), verificando em cada URL listada naquele arquivo a ocorrência de uma extensão válida. Uma extensão válida deve ocorrer antes da primeira “?” da URL e a partir da última “/” que antecede esse sinal. O processo verifica então, por casamento de padrões, a ocorrência das extensões nessa subcadeia delimitada por “/” e “?” da URL. Para extensões que tenham o mesmo radical (por exemplo, asp e aspx) é considerada a mais longa.

O ponto de partida da metodologia foi apurar quais as possíveis extensões, como medida das tecnologias utilizadas. Nesse caso, buscou-se uma lista de 397 extensões de arquivos a partir do sítio <http://www.file-extensions.org> e com base nelas foi feita a análise da URL de cada página coletada de forma a identificar quais extensões ocorriam na URL.

Feita a identificação de extensões, temos 3 casos possíveis.

1- Nenhuma extensão foi encontrada na URL: Que impede a possibilidade de identificar a tecnologia utilizada.

2- Encontra-se exatamente uma extensão, caso no qual a tecnologia, se for o caso, é associada diretamente.

3-Mais de uma extensão é associada à URL e é necessário estimar qual das extensões detectadas é a mais pertinente. Utilizou-se dois critérios para detectar a extensão mais pertinente. O primeiro critério foi a posição onde a extensão ocorre na URL, com base na premissa que a extensão do primeiro arquivo que ocorre identifica a sua tecnologia base. Esse critério foi validado em uma porção significativa dos casos. O segundo critério foi, para extensões que ocorram na mesma posição, escolher a maior, por ser naturalmente mais discriminativa. Por exemplo, considere que as extensões php e php3 são detectadas a partir da mesma posição em uma URL, o que explica pelo fato de php ser parte de php3. Neste caso, a extensão selecionada será php3, pois ela é maior e

mais discriminativa.

O último passo da metodologia foi selecionar, dentre as extensões identificadas, aquelas associadas a tecnologias. Esse processo foi feito manualmente, verificando as extensões que efetivamente ocorrem e as suas respectivas descrições.

- **Apresentação dos resultados**

Um resumo descritivo dos dados sobre distribuição dos tipos de objetos na Web governamental está apresentado em quatro tabelas (G2-1A, G2-1B, G2-2A e G2-2B).

As Tabelas G2-1A e G2-1B apresentam as tecnologias adotadas nas páginas Web, separadas por Linguagens de Programação e demais tipos de extensão, respectivamente.

Tabela G2-1A: Distribuições das tecnologias adotadas - Linguagens de Programação

Tecnologia	Quantidade	%	Descrição
asp	541.407	7,88	ASP script, page
aspx	106.470	1,55	ASP.NET script, page
cgi	8.798	0,13	Common Gateway Interface
jsp	22.931	0,33	JAVA Server page
php	2.598.249	37,79	PHP script page
py	172	0,00	Python

Tabela G2-1B: Distribuições das tecnologias adotadas - demais tipos de extensão

Tecnologia	Quantidade	%	Descrição
.Indefinido	3.556.891	51,74	Indefinido
cfm	18.008	0,26	Cold Fusion Markup
com	33	0,00	Common Object Module
dll	7.328	0,11	Dynamic Link Library file
do	12.602	0,18	Oracle Application Server
exe	4	0,00	Executable file
nsf	35	0,00	IBM Notes
sql	20	0,00	Structured Query Language Data SQL
wsp	1.971	0,03	SharePoint Services Solution

As Tabelas G2-2A e G2-2B apresentam como está a adoção de tais tecnologias adotadas nas unidades federativas, em sítios da Web governamental brasileira (.gov.br) que não estão associados diretamente a uma dessas unidades federativas e outros sítios, obtidos em razão de acesso via redirecionamento de um site governamental.

Tabela G2-2A: Distribuições das tecnologias adotadas por unidade federativa - Linguagens de Programação

Estado	asp	aspx	cgi	jsp	php	php3	py
.GOV.BR	7,23	1,84	0,00	0,35	23,18	0,00	0,01
.outros	2,23	0,78	0,05	17,48	4,42	0,05	0,00
AC	0,01	0,00	0,00	4,81	78,11	0,00	0,00
AL	1,55	2,17	0,00	0,06	10,00	0,00	0,00
AM	3,80	0,01	0,00	0,03	16,34	0,00	0,00
AP	0,00	0,00	0,00	5,21	10,64	0,00	0,00
BA	5,04	1,67	0,00	0,00	34,70	0,00	0,00
CE	1,66	0,04	0,00	0,03	49,32	0,00	0,00
DF	71,36	0,35	0,00	0,00	16,14	0,00	0,00
ES	6,38	7,08	0,00	0,00	42,51	0,00	0,00
GO	1,40	0,70	0,00	1,28	55,83	0,00	0,00
MA	2,49	1,30	0,00	0,12	60,34	0,00	0,00
MG	4,20	0,66	0,03	0,01	27,87	0,00	0,00
MS	1,13	6,19	0,00	0,04	23,88	0,00	0,00
MT	2,96	0,09	0,00	0,01	15,76	0,00	0,00
PA	4,71	0,14	0,00	0,02	20,58	0,00	0,00
PB	0,94	0,19	0,18	0,02	63,36	0,00	0,00
PE	4,36	0,58	0,00	0,03	32,30	0,00	0,00
PI	0,31	0,00	0,00	2,26	68,39	0,00	0,00
PR	2,98	0,25	0,00	0,16	73,64	0,00	0,00
RJ	22,50	0,30	0,00	1,12	41,67	0,00	0,00
RN	58,79	0,39	0,00	0,00	21,81	0,00	0,00
RO	12,22	0,29	0,09	0,10	39,22	0,00	0,00
RR	48,40	0,00	0,00	0,12	41,24	0,00	0,00
RS	12,70	0,47	0,00	1,28	45,31	0,00	0,00
SC	2,54	0,12	2,08	0,35	46,61	0,00	0,00
SE	2,26	0,13	0,00	0,29	73,85	0,00	0,00
SP	9,91	5,49	0,00	0,01	41,31	0,00	0,00
TO	0,08	0,72	0,00	0,04	62,76	0,00	0,00

Tabela G2-2B: Distribuição das tecnologias adotadas por unidade federativa - demais tipos de extensão

Estado	INDEFINIDO	cfm	com	dbc	dll	do	exe	js	nsf	sql	wsp
.GOV.BR	65,66	0,89	0,00	0,00	0,03	0,67	0,00	0,00	0,00	0,00	0,13
.outros	74,75	0,19	0,00	0,00	0,00	0,05	0,00	0,00	0,00	0,00	0,00
AC	17,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
AL	85,90	0,31	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
AM	79,71	0,00	0,00	0,00	0,00	0,09	0,00	0,00	0,00	0,00	0,00
AP	84,16	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
BA	56,42	0,25	0,00	0,00	1,89	0,00	0,00	0,00	0,00	0,00	0,02
CE	48,93	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00
DF	12,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ES	39,19	4,02	0,00	0,00	0,00	0,83	0,00	0,00	0,00	0,00	0,00
GO	40,78	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MA	35,72	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00
MG	67,23	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MS	68,73	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,02	0,00	0,00
MT	81,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PA	74,52	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PB	35,29	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00
PE	62,72	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PI	28,75	0,00	0,00	0,00	0,00	0,30	0,00	0,00	0,00	0,00	0,00
PR	22,83	0,00	0,00	0,00	0,00	0,14	0,00	0,00	0,00	0,00	0,00
RJ	34,24	0,15	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00
RN	19,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RO	48,08	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RR	10,24	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RS	40,21	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00
SC	48,28	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00
SE	23,01	0,00	0,01	0,00	0,45	0,00	0,00	0,00	0,00	0,00	0,00
SP	43,22	0,00	0,00	0,00	0,00	0,06	0,00	0,00	0,00	0,00	0,00
TO	36,39	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

- **Análise**

Primeiramente é importante destacar que não foi possível identificar a tecnologia adotada em 51,74% das páginas Web porque existe um volume significativo de páginas Web que não apresentam extensão. No restante das páginas, observa-se predominância do uso da tecnologia PHP, presente em 37,79% das páginas Web. Em

seguida estão as tecnologias ASP (7,88%) e ASPX (1,55%). O uso da linguagem Java (através das tecnologias JSP, JSF etc.) e Python (.py) foi muito pequeno, provavelmente porque o uso de tais tecnologias fica oculto nas URLs.

Em relação às unidades federativas (UF), a explicação dos resultados é difícil de ser feita, uma vez que há grande variação na identificação dos tipos de objetos. Exemplos são os valores de indefinição de tipo de objetos de 10,24% (RR) a 85,9% (AL). Ainda assim, pode-se destacar alguns pontos:

- Em Roraima (RR): há quase 90% dos tipos de objetos definidos, com destaque para adoção de ASP (48,4%) e PHP (41,2%);
 - No Distrito Federal (DF), há quase 88% dos tipos de objetos definidos, com destaque para grande adoção de tecnologia ASP (71,4%), contra 16,1% de PHP;
 - Já no Acre (AC), há quase 83% dos objetos identificados, com predominância no uso de PHP (78,1%), seguido de JSP (4,8%);
 - No caso dos sítios da esfera federal (.gov.br), 65,6% dos objetos não são identificáveis, sendo que PHP é adotado em 23,2%, ASP em 7,23% e ASPX em 1,84%
-
- **Análise Comparativa com estudo anterior e considerações**

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (.gov.br) realizada em 2010, observa-se que a quantidade de objetos em que foi possível identificar a tecnologia adotada diminuiu quase 10%. O uso da tecnologia PHP aumentou em 4,6% (passando de 2.483.013 para 2.598.249). O uso de ASP caiu 37,7% (passando de 868.183 para 451.407), já a tecnologia ASPX apresentou variação positiva de 13,3% (de 94.017 para 106.470). Pode-se observar também que o uso de tecnologias vinculadas à linguagem Java reduziu de forma significativa (antes tinha-se 1,49% do uso de JSP e na atual pesquisa o valor reduzir para 0,33%).

Com base nos resultados apresentados, podemos concluir que não se pode afirmar ao certo sobre a adoção real de certas tecnologias nas páginas Web, visto que existe uma tendência, adotada mais recentemente, de omitir o formato do objeto nas páginas Web, dificultando a identificação da tecnologia adotada.

H1: Idade média das páginas Web

1. Definição do indicador

Valor médio da idade das páginas da Web governamental brasileira (.gov.br), considerando a data de sua última atualização.

2. Propósito

Obter a idade média das páginas da Web governamental brasileira (.gov.br).

3. Metodologia

O cálculo da idade de uma página pode ser realizado a partir da diferença entre o instante de tempo em que ela é coletada e o instante de sua última modificação. Tal informação é obtida como parte da resposta à requisições às páginas Web, porém ela não é obrigatoriamente enviada, o que deixa defasados os dados para esse indicador. Outras limitações da metodologia dizem respeito a erros na configuração da hora nos servidores Web.

4. Apresentação dos resultados

Apresenta-se um resumo descritivo sobre idade média das páginas da Web governamental brasileira (.gov.br) em duas tabelas (H1-1 e H1-2).

O resultado de 646,47 dias apresentado na Tabela H1-1 mostra que as páginas Web do governo são atualizadas aproximadamente uma vez a cada 2 anos.

Tabela H1-1: Idade média das páginas da Web governamental brasileira (.gov.br)

Idade média das páginas
646,47

A Tabela H1-2 mostra, além da idade média das páginas das quais foi possível obter a informação de idade, o percentual de páginas segundo as unidades federativas (UF). Observa-se que a maior parte dos estados (23 dos 27) apresenta menos de 5% de proporção de páginas com informação de idade. Acima desse limiar estão apenas os estados do Maranhão (5,77%), São Paulo (10,15%), Rondônia (12,29%) e Alagoas (62,92%).

Tabela H1-2: Total de páginas da Web governamental brasileira (.gov.br) e idade média segundo as por unidade federativa

UF	Total Págs	Págs com idade	% Págs com idade	Idade média
----	------------	----------------	------------------	-------------

UF	Total Págs	Págs com idade	% Págs com idade	Idade média
.GOV.BR	1.531.651	112.557	7,35	843,54
.outros	2.059	541	26,27	696,06
AC	44.699	1.676	3,75	1309,14
AL	268.250	168.794	62,92	352,2
AM	35.901	920	2,56	374,85
AP	13.197	657	4,98	377,81
BA	319.622	6.386	2	676,54
CE	292.353	8.423	2,88	480,82
DF	118.800	1.222	1,03	1923,07
ES	57.693	1.512	2,62	373,03
GO	84.496	1.380	1,63	789,93
MA	43.861	2.530	5,77	2086,75
MG	590.693	17.993	3,05	264,95
MS	71.261	529	0,74	510,02
MT	263.501	2.359	0,9	1916,42
PA	76.089	1.069	1,4	759,63
PB	65.519	886	1,35	1183,93
PE	85.963	3.052	3,55	834,38
PI	48.375	383	0,79	1207,79
PR	585.172	14.981	2,56	457,58
RJ	276.762	10.447	3,77	575,94
RN	53.783	785	1,46	1555,65
RO	54.652	6.718	12,29	328,29
RR	2.413	117	4,85	987,98
RS	433.531	21.077	4,86	745,2
SC	406.602	10.542	2,59	1114,23
SE	158.177	543	0,34	1038,27
SP	844.782	85.754	10,15	907,6
TO	45.069	369	0,82	3037,7

5. Análise

A grande maioria das páginas coletadas, 6.390.724 (93% do total), não tiveram dados consistentes para esse campo. Isso geralmente se deve ao fato dessa informação não ser fornecida ou devido ao fato de que grande parte das páginas ser dinâmica e, conseqüentemente, esse tipo de campo não se aplicar. Além disso, na avaliação da sincronização do relógio dos servidores (apresentado no indicador I1), foi observado que existem servidores apresentaram relógios muito defasados em relação ao sistema de coleta, o que pode ter afetado os resultados obtidos neste indicador de idade das páginas Web.

A falta de atualização das páginas Web nem sempre pode ser considerada algo ruim, dependendo da natureza de seu conteúdo algumas páginas Web não devem mesmo ser modificadas ao longo do tempo, exemplos são documentos públicos e informações de leis ou decretos.

As diferentes proporções de páginas com informação de idade por UF podem estar relacionadas à própria configuração de seus servidores Web. Em relação à idade média das páginas por estado, observa-se uma variação de 264,95 dias (Minas Gerais) até 3037,7 dias (Tocantins).

Os problemas com a sincronização dos relógios dos servidores e a não obrigatoriedade

no envio da data de última modificação em respostas à requisições Web podem dificultar as análises. Cabe ressaltar que a falta de sincronização não afeta esses resultados, uma vez que a porcentagem de sítios com diferenças de tempo muito grandes é muito pequena. Como a média de idade (ver indicador I-1) é maior do que 2 anos, os relógios com alguns minutos ou até algumas horas de erro não são tão relevantes.

Uma possível alternativa para minimizar esses problemas pode ser a utilização de coletas periódicas para avaliar a idade das páginas Web, o que permite um melhor acompanhamento da evolução desse indicador.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (.gov.br) realizada em 2010, observa-se que o indicador sofreu pequena variação, considerando que a idade média era de 656 e agora foi de 646 dias.

H2: Proporção de páginas Web dinâmicas

1. Definição do indicador

Valor percentual de páginas consideradas dinâmicas na Web governamental brasileira (.gov.br).

No desenvolvimento de um sítio Web, uma página dinâmica, em geral, diz respeito ao uso de linguagens de programação de servidor, tal como PHP, ASP, JSP e ColdFusion. Tais linguagens permitem, no momento da execução da requisição ou solicitação feita, a criação do conteúdo estático que será enviado ao usuário, por isso a denominação utilizada.

2. Propósito

Obter a representatividade de páginas Web com conteúdo dinamicamente gerado na Web governamental brasileira (.gov.br).

3. Metodologia

O coletor utilizado, WIRE, possui um conjunto de regras internas para diferenciar páginas dinâmicas de estáticas. Essas regras levam em conta tanto o tipo de extensão da URL, por exemplo, extensões de arquivo como .jps ou .php são associadas a documentos dinâmicos, quanto a existência de parâmetros associados à URL da página Web. Com base nessa informação, o coletor armazena, para cada página consultada, a natureza do conteúdo a ela associado (estático ou dinâmico).

4. Apresentação dos resultados

Apresenta-se um resumo descritivo dos dados acerca da proporção de páginas dinâmicas na Web em três tabelas (H2-1, H2-2 e H2-3).

A Tabela H2-1 mostra que o percentual de páginas dinâmicas foi de 76,62%, ou seja, aproximadamente 3 a cada 4 páginas Web.

Tabela H2-1: Percentual de Páginas Web Dinâmicas na Web governamental brasileira (.gov.br)

Percentual de páginas dinâmicas
76,62

Apesar de a maioria das páginas Web coletadas serem dinâmicas, sua distribuição entre os sítios é bastante irregular. A partir da análise da tabela H2-2, pode-se notar que a maioria dos sítios se encontra nos extremos do espectro: 6.487 sítios (47,5%) possuem até 10% de páginas dinâmicas e 4.539 sítios (33,2%) possuem mais de 90% de páginas dinâmicas. Dada a grande porcentagem de páginas dinâmicas encontradas, fica claro que os maiores sítios são aqueles com as maiores porcentagens de páginas dinâmicas.

Tabela H2-2: Número de sítios pela proporção de uso de páginas dinâmicas

0	>0 e ≤10	>10 e ≤20	>20 e ≤30	>30 e ≤40	>40 e ≤50	>50 e ≤60	>60 e ≤70	>70 e ≤80	>80 e ≤90	>90 e ≤100
5.993	494	203	155	213	590	183	354	398	536	4.539

Na Tabela H2-3 os dados são apresentados por unidade federativa (UF). Observa-se que os tipos de conteúdos utilizados variam bastante entre os estados. Para auxiliar a análise do indicador foram também incluídas nessa tabela informações sobre o número e porcentagem de páginas dinâmicas das quais não se foi possível obter a idade conforme metodologia apresentada pelo indicador H1.

Tabela H2-3: Quantidade e porcentagem de páginas dinâmicas e porcentagem de páginas dinâmicas sem idade na Web governamental brasileira (.gov.br) segundo as unidades federativas (UF)

UF	Total págs	Nº págs dinâmicas	Nº págs dinâmicas sem idade	% págs dinâmicas	% págs dinâmicas sem idade
.GOV.BR	1.531.651	1.070.932	1.063.766	69,92	69,45
.outros	2.059	623	610	30,26	29,63
AC	44.699	41.916	41.902	93,77	93,74
AL	268.250	153.432	55.810	57,2	20,81
AM	35.901	18.181	18.180	50,64	50,64
AP	13.197	10.301	10.284	78,06	77,93
BA	319.622	241.283	241.227	75,49	75,47
CE	292.353	205.305	200.504	70,23	68,58
DF	118.800	112.460	112.457	94,66	94,66
ES	57.693	49.937	49.862	86,56	86,43
GO	84.496	72.304	72.273	85,57	85,53
MA	43.861	35.204	35.195	80,26	80,24
MG	590.693	457.800	456.488	77,5	77,28
MS	71.261	36.835	36.544	51,69	51,28
MT	263.501	112.583	112.575	42,73	42,72
PA	76.089	68.332	68.326	89,81	89,8
PB	65.519	54.405	54.360	83,04	82,97
PE	85.963	66.268	66.250	77,09	77,07
PI	48.375	41.845	41.800	86,5	86,41
PR	585.172	529.342	522.196	90,46	89,24
RJ	276.762	244.500	244.437	88,34	88,32
RN	53.783	44.653	44.593	83,02	82,91
RO	54.652	48.322	44.727	88,42	81,84
RR	2.413	2.174	2.172	90,1	90,01
RS	433.531	364.236	362.292	84,02	83,57
SC	406.602	340.365	340.344	83,71	83,7
SE	158.177	135.746	135.678	85,82	85,78
SP	844.782	678.618	665.219	80,33	78,74
TO	45.069	29.409	29.234	65,25	64,87

5. Análise

Com base nos resultados apresentados, pode-se notar que mais de 3 páginas Web de cada 4 (76,62%) adotam alguma tecnologia de geração dinâmica de conteúdo. Conclui-se também que estados com um conjunto menor de sítios, como Acre, Paraná e Roraima, têm uma porcentagem elevada de páginas dinâmicas, provavelmente por estarem mais concentrados ao redor de grandes portais. Entretanto, o número de sítios não é determinante nesse caso: os estados com menores porcentagens de páginas dinâmicas, Alagoas e Amazonas, também possuem número reduzido de sítios. Cabe ressaltar que alguns estados, como Mato Grosso, Amazonas e Mato Grosso do Sul, apresentam valores abaixo da média nacional para este indicador, provavelmente devido ao fato de alguns sítios da esfera do governo federal terem um predomínio de páginas estáticas.

Os dados indicam grande uso de páginas dinâmicas, algo que vem sendo observado na Web como um todo em função da utilização cada vez maior de sistemas de personalização e de gerência de conteúdo, como por exemplo, Joomla – (www.joomla.org), que simplificam o processo de atualização das informações. Esse resultado também está coerente com a adoção das tecnologias ASP, ASPX, JSP e, em especial, o PHP, que vem sendo amplamente adotado na esfera governamental (ver indicador G2 - Distribuição de tipos de tecnologias adotadas nas páginas Web).

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior sobre a Web governamental brasileira (2010), observa-se que o indicador sofreu pequena variação, passando de 74,8% para 76,6%.

I1: Sincronização de tempo dos servidores

1. Definição do indicador

Valor estimado da diferença entre a hora configurada nos servidores da Web governamental brasileira (.gov.br) e a Hora Legal Brasileira, definida a partir do UTC (Coordinated Universal Time).

2. Propósito

Obter medida do sincronismo de tempo dos servidores da Web governamental brasileira (.gov.br) em relação à Hora Legal Brasileira. Uma das importâncias da sincronização é que ela é fundamental para garantir o correto registro de transações realizadas na Web, bem como o correto funcionamento de diversas aplicações.

3. Metodologia

Existem algumas formas para se manter a hora de um dispositivo atualizada, uma delas é com a utilização de um servidor de tempo e um protocolo. Normalmente o protocolo adotado é o NTP (Network Time Protocol), que permite a sincronização dos relógios dos dispositivos de uma rede através de fontes de tempo confiáveis.

Para obtenção da estimativa do sincronismo dos servidores da Web governamental brasileira (.gov.br), foi realizada uma requisição do tipo HTTP HEAD ao servidor. A partir da resposta obtida, obteve-se o horário do relógio do servidor e o tempo de resposta da requisição. E, assim, pôde-se calcular o erro de sincronismo desse relógio contabilizando o módulo da diferença entre o tempo obtido e o da máquina que realizou o experimento somado da metade do tempo de resposta obtido. É importante observar que para confirmar a validade dos resultados, a máquina de testes estava sincronizada via NTP.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre o sincronismo dos servidores da Web governamental brasileira (.gov.br) está apresentado em duas tabelas (I1-1 e I1-2).

A Tabela I1-1 apresenta os dados consolidados e a Tabela I1-2 apresenta os dados segundo as unidades federativas (UF). Em relação aos dados consolidados, o valor mínimo foi igual a zero, o que indica que há servidores sincronizados, porém o valor máximo foi de 664.206.306 segundos (~21 anos), mostrando a existência de servidores com anos de erro em seu relógio. A média foi 154.656,26, o que representa 1,79 dias. Como indicado pela mediana, metade dos valores de sincronização dos sítios é inferior a 4 segundos.

Tabela I1-1: Medidas resumo (em segundos) da sincronização de tempo dos servidores da Web governamental brasileira (.gov.br)

Min	Mediana	Média	Max	CV
0	4	154.656,26	664.206.306	56,14

Tabela I1-2: Medidas resumo (em segundos) da sincronização de tempo dos servidores da Web governamental brasileira (.gov.br) segundo as unidade da federativas (UF)

UF	Min	Mediana	Média	Max	CV
.GOV.BR	0	1	2.018,92	1.448.559	21,47
.outros	0	0	676,84	443.052	16,23
AC	0	7	746,87	14.366	3,32
AL	0	17	3.629,94	54.480	1,98
AM	0	0	149,68	5.445	3,9
AP	0	7	119.843,06	7.629.294	7,89
BA	0	2	279,81	61.883	9,55
CE	0	317	2.796.981,22	664.206.306	15,38
DF	0	17	689,73	86.982	9,51
ES	0	0	229,87	21.735	6,12
GO	0	1	309,63	11.142	4,25
MA	0	64	598,46	15.939	3,69
MG	0	2	394,35	29.096	4,38
MS	0	19	957,34	28.801	4,64
MT	0	1	369,05	17.243	5,34
PA	0	46	278,34	11.843	4,16
PB	0	126	10.560,76	2.332.612	14,47
PE	0	3	2.775,98	490.764	11,29
PI	0	672	5.885,10	309.021	4,68
PR	0	0	662,91	413.986	22,15
RJ	0	1	474,05	58.768	6,15
RN	0	0	46,04	1.879	4,85
RO	0	7	220,53	4.804	3,03
RR	0	0	386,06	3.448	2,13
RS	0	0	484.078,04	366.269.309	27,48
SC	0	4	801.581,37	246.531.381	17,51
SE	0	8	567,61	15.717	3,09
SP	0	3	1.364,97	2.173.229	34,89
TO	0	7	1.184,42	86.829	5,75

5. Análise

Observa-se que há uma grande variabilidade quanto à estimativa de sincronização dos servidores dos sítios da Web Governamental Brasileira (.gov.br) dada a diferença entre a média e a mediana observadas. Em relação as UF o mesmo efeito pode ser observado, porém nota-se também uma variação grande entre estados.

Esses dados indicam que, em geral, cerca de 50% dos servidores estão sincronizados no intervalo de mais ou menos quatro segundos em relação à Hora Legal Brasileira. No entanto, a partir daí os valores crescem rapidamente, mostrando que existem servidores que estão bastante dessincronizados. Valores extremos chegam a cerca de vinte anos de diferença, o que provavelmente ocorre por erro na configuração dos servidores analisados. Tecnicamente isso implica que o coeficiente de variação do conjunto, calculado pela razão entre o desvio padrão e a média, é extremamente elevado.

Isso mostra que existe um importante trabalho a ser feito com o intuito de conscientizar administradores sistemas sobre a necessidade da sincronização dos relógios dos servidores e sua importância principalmente quanto ao segurança e bom funcionamento das aplicações.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (2010), observa-se que o indicador sofreu algumas mudanças significativas. Na última edição da análise metade dos servidores estavam sincronizados ou defasados em até 1 segundo, já agora são até 4 segundos. Já o valor médio não teve grande variação, apenas a elevação de 3%.

I2: Tempo de resposta dos servidores

1. Definição do indicador

Valor do tempo de resposta à requisições Web dos servidores da Web governamental brasileira (.gov.br).

2. Propósito

Obter valor do tempo de resposta à requisições Web dos servidores da Web governamental brasileira (.gov.br).

3. Metodologia

Para realizar a coleta da informação de tempo de resposta de um determinado sítio da Web governamental brasileira (.gov.br) foi feita uma consulta do tipo HTTP HEAD ao servidor de cada sítio encontrado, a partir da qual calculou-se medida com base nos instantes de envio da requisição e de resposta do servidor. A informação obtida foi considerada como o valor aproximado de tempo que é necessário para acesso ao servidor. Como sua precisão é da ordem de segundos, permite uma avaliação geral a cerca do desempenho dos servidores avaliados.

4. Apresentação dos resultados

Um resumo descritivo dos dados sobre o indicador de tempo de resposta para os sítios da Web governamental brasileira (.gov.br) pode ser observado em três tabelas (I2-1, I2-2 e I2-3).

A Tabela I2-1 apresenta os dados consolidados, a Tabela I2-2 apresenta os dados segundo as unidades federativas (UF) e por fim, a tabela I2-3 mostra a representatividade do número de sítios, categorizada segundo classes de tempo de resposta (menor ou igual a 10 milissegundos, entre 10 e menor ou igual a 50 e assim por diante). Assim, observa-se que os tempos variam entre 0 e 19.882 milissegundos, sendo a média 360 milissegundos (6 segundos). E que há grande variabilidade entre os tempos de resposta em relação aos dados consolidados por estado, pois alguns estados da região Norte apresentam as maiores médias de tempo de resposta como Amapá (AP) e Pará (PA). Além de haver distribuições variadas do tempo de resposta por estado.

Tabela I2-1: Medidas resumo (em milissegundos) do tempo de resposta dos servidores da Web governamental brasileira (.gov.br).

Min	Mediana	Média	Máx	CV
0	74	360,28	19.882	3,33

A análise do tempo de resposta médio dos servidores foi feita por unidade federativa e está apresentada na Tabela I2-2. Existe grande variação nos tempos de resposta entre elas. Alguns Estados do Norte apresentam as maiores médias de tempo de resposta como Amapá (AP), Pará (PA) e Roraima (RR).

Tabela I2-2: Medidas resumo (em milissegundos) do tempo de resposta dos servidores da Web governamental brasileira (.gov.br) segundo as unidades federativas (UF)

UF	Min	Mediana	Média	Máx	CV
.GOV.BR	1	37	254,54	17.832	4,22
.outros	1	157	411,75	18.198	2,91
AC	8	90	237,93	2.761	1,72
AL	2	100	765,57	18.318	2,85
AM	3	67	261,61	4.110	1,65
AP	29	628	1101,11	10.835	1,52
BA	1	153	592,15	11.467	2,25
CE	3	70	418,85	18.453	3,85
DF	17	143	574,8	12.397	3,15
ES	1	175	767,45	12.105	1,88
GO	1	65	434,79	6.922	2,4
MA	3	156	410,37	6.780	2,05
MG	1	79	471,84	19.882	3,58
MS	1	79	310,35	7.447	2,35
MT	1	176	236,55	3.043	1,55
PA	1	105	991,86	15.443	3,27
PB	2	118	631,96	13.729	2,39
PE	2	89	311,95	18.666	4,07
PI	5	82	137,21	1.183	1,43
PR	1	30	245,65	17.361	3,72
RJ	0	44	229,99	8.884	3,19
RN	3	116	213,14	4.173	1,99
RO	0	99	431,42	5.174	1,77
RR	5	559	781,21	11.490	2,07
RS	2	42	300,48	11.510	2,82
SC	1	41	239,04	7.808	2,18
SE	17	84	375,15	5.950	2,04
SP	0	25	285,19	17.351	3,67
TO	21	76	372,79	14.662	3,34

A Figura I2-1 apresenta a frequência de algumas faixas de valores de tempos de resposta. Dessa figura, pode-se notar a existência de uma alta frequência de tempos de resposta concentrados até 200 milissegundos (aproximadamente 3/4), sendo que a faixa com maior frequência é a de 10 a 50 milissegundos, com aproximadamente 30% dos sítios. Já em relação aos tempos superiores a 1 segundo, nota-se seu baixo percentual de ocorrências, porém crítico em algumas unidades federativas, como Amapá (AP) e Espírito Santo (ES). Em ambos os casos, mais de 20% das ocorrências estão nessa faixa.

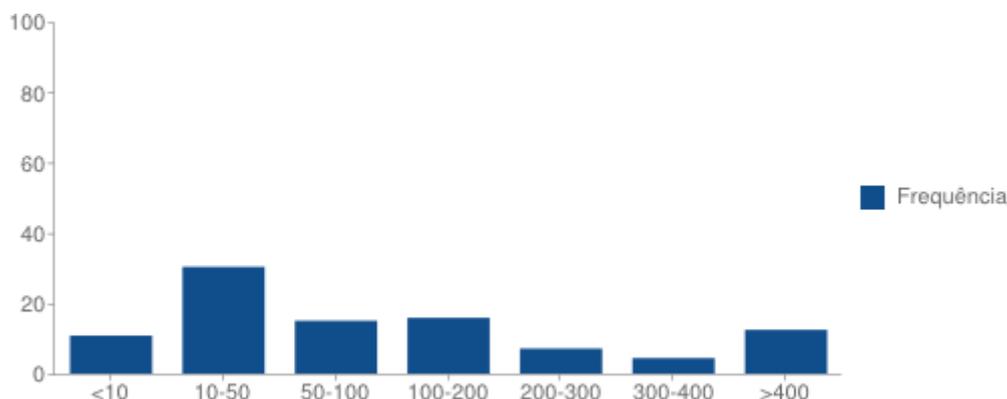


Figura I2-1: Frequências das classes de tempo de resposta dos servidores da Web governamental brasileira (.gov.br)

Tabela I2-3: Quantidades e porcentagens de sítios da Web governamental brasileira (.gov.br) por classe de tempo de resposta de seus servidores segundo as unidades federativas (UF).

UF	Nº sítios	≤10 (%)	>10 e ≤50 (%)	>50 e ≤100 (%)	>100 e ≤200 (%)	>200 e ≤300 (%)	>300 e ≤400 (%)	>400 e ≤500 (%)	>500 e ≤600 (%)	>600 e ≤700 (%)	>700 e ≤800 (%)	>800 e ≤900 (%)	>900 e ≤1000 (%)	>1000 (%)
.GOV.BR	2.233	8,02	54,59	12,63	7,39	4,93	2,19	1,52	1,07	1,66	0,63	0,45	0,36	4,57
.outros	2.099	12,24	17,53	7,24	22,34	15,15	6,43	3,43	2,14	1,43	2,1	1,38	0,95	7,62
AC	55	1,82	5,45	49,09	12,73	14,55	3,64	1,82	1,82	0	1,82	1,82	1,82	3,64
AL	177	2,26	11,3	37,29	12,99	3,39	3,95	2,26	5,65	2,82	1,69	2,26	0,57	13,56
AM	195	1,54	28,21	28,21	6,67	4,1	6,15	6,67	6,67	5,64	2,05	1,03	0	3,08
AP	64	0	1,56	1,56	1,56	3,13	0	0	17,19	39,06	6,25	3,13	3,13	23,44
BA	568	3,35	9,68	29,93	12,85	11,97	3,52	3,35	3,7	2,11	2,46	0,7	1,41	14,96
CE	480	0,83	31,67	35,42	14,17	3,13	4,38	1,25	1,04	0,21	0,83	0,42	0,21	6,46
DF	178	0	10,11	14,61	47,19	10,11	5,62	1,12	1,12	1,12	0,56	0,56	1,69	6,18
ES	365	6,85	20,55	14,79	11,23	7,12	3,56	2,74	3,56	2,47	1,64	0,55	1,37	23,56
GO	254	6,3	37,4	11,81	9,84	9,45	6,69	1,97	1,57	1,97	1,18	1,18	0,79	9,84
MA	161	1,24	9,32	16,77	36,65	7,45	4,97	7,45	3,73	1,24	1,86	0,62	1,24	7,45
MG	1.050	8,38	34,95	10,76	17,52	5,81	3,71	2,95	1,9	1,43	1,9	0,76	1,14	8,76
MS	246	4,47	8,54	45,94	13,41	4,47	8,13	2,03	2,85	2,44	1,63	0	0,41	5,69
MT	201	7,46	17,91	4,98	41,79	15,92	1,49	1	1,49	1	0,5	1,49	1	3,98
PA	230	2,61	16,09	20	25,65	5,22	9,13	6,52	3,04	0,43	1,3	0,43	0,87	8,7
PB	232	1,72	13,79	16,38	35,78	6,9	3,45	2,59	1,72	0,43	1,29	1,29	0,43	14,22
PE	253	2,77	5,93	57,71	14,23	3,16	5,14	1,19	1,19	0	1,58	0,79	0,4	5,93
PI	126	2,38	14,29	56,35	15,87	0,79	3,97	0,79	1,59	0,79	0	0,79	0	2,38
PR	1.597	23,17	33	3,57	13,65	5,57	6,89	3,76	2,63	1,38	1,19	1,19	0,69	3,32
RJ	684	11,11	41,52	11,11	14,62	5,85	5,12	1,9	1,75	1,02	1,32	0,44	0,44	3,66
RN	226	0,89	17,26	6,19	54,42	8,41	2,65	4,87	0,44	0,44	0,89	0	0,89	2,65
RO	118	5,08	16,1	28,81	11,86	8,47	2,54	3,39	0,85	2,54	3,39	2,54	1,69	11,86
RR	47	2,13	10,64	2,13	0	2,13	0	0	65,96	4,26	4,26	0	0	8,51
RS	759	1,71	54,28	11,46	11,46	3,95	3,16	1,71	2,11	1,05	0,79	0,66	0,66	6,98
SC	923	22,97	30,77	8,99	10,51	5,53	3,9	2,28	3,58	2,71	1,84	0,54	0,22	6,18
SE	220	0	7,27	48,18	8,18	5,91	7,73	4,55	3,18	1,82	1,36	1,36	1,36	9,09
SP	2.098	32,7	24,83	7,01	14,68	5,05	2,96	1,81	1,67	1,38	0,81	0,81	0,62	5,53
TO	173	0	22,54	36,42	13,87	9,25	4,62	0,58	1,16	0,58	0,58	0,58	0,58	9,25

5. Análise

Pode-se considerar que um tempo de resposta inferior a 50 milissegundos é excelente, inferior a 200 milissegundos é razoável, até 1000 milissegundos é regular e acima de 1000 milissegundos já não mais satisfatório. A média geral foi de 360 milissegundos (6 segundos), com um coeficiente de variação de 3,33. A mediana indica que metade dos sítios tem tempo igual ou menor a 103 milissegundos (menos de 2 segundos).

Isso indica que os tempos de resposta medidos apresentam um resultado satisfatório para uma boa parte dos sítios, porém quando os domínios são separados por unidade federativa, nota-se degradação em alguns casos, conforme apresentado nas tabelas. A pesar de um maior tempo de resposta por parte do estado do Amapá ser esperado, devido a sua distância em relação ao estados de São Paulo, de onde os testes foram realizados, o resultado Espírito Santo é surpreendente. Como nesse último caso a rede não se configura num fator tão relevante, o resultado ruim indica uma baixa qualidade dos servidores e/ou serviços providos, o que implica diretamente em seu desempenho na resposta a requisições Web.

O método adotado permite ter uma visão geral do indicador, mas deve ser aperfeiçoado de forma a possibilitar uma análise mais detalhada e real do desempenho em termos de conectividade e qualidade de serviço dos servidores Web. Assim, para futuras análises pode-se pensar no uso de formas mais robustas para coleta desse indicador de desempenho, que considerem, por exemplo, diferentes rotas de acesso ao servidor.

6. Análise Comparativa com estudo anterior e considerações

Ao comparar estes resultados com a pesquisa anterior da Web governamental brasileira (.gov.br) de 2010, nota-se uma degradação do tempo de resposta, dado que a média aumentou de 190,20 milissegundos para 360,28 milissegundos. Por outro lado, comparando-se os valores das medianas, nota-se uma degradação menor, de 71 para 103 milissegundos, o que corresponde a 45%.