

Análise de dados da primeira coleta da TIC Web Acessibilidade

Autores: Reinaldo Ferraz e Vagner Diniz

No dia 3 de dezembro de 2021 o Centro de Estudos sobre Tecnologias Web (Ceweb.br) lançou a plataforma [TIC Web Acessibilidade](#) com o objetivo de apresentar de forma simples o nível de conformidade com padrões de acessibilidade nas páginas Web do governo brasileiro.

Histórico da TIC Web

A TIC Web Acessibilidade é uma evolução da pesquisa [TIC Web](#), também conhecida como “Censo da Web”, publicada entre 2010 e 2017, que apresentou diversos indicadores sobre páginas Web governamentais, como tamanho das páginas, tipo de arquivos linkados, localização de servidores e uso de IPv6, e conformidade com padrões de marcação e acessibilidade.

O indicador de acessibilidade sempre foi muito procurado. Conforme as ferramentas de análise evoluíram, decidimos ajustar o escopo da coleta somente para a conformidade com os padrões de acessibilidade das páginas. Para isso, foi desenvolvida uma plataforma específica que pudesse coletar e analisar, com base no software desenvolvido pelo governo federal, o nível de conformidade com padrões de acessibilidade na Web governamental brasileira.

Como funciona a plataforma

A plataforma tem uma arquitetura robusta para coletar e avaliar páginas Web. Foi desenvolvido um *WEB crawler*¹ que vasculha uma lista de sites (que chamamos de “semente”). A partir da página inicial do site, a plataforma coleta os links (URLs) para outras páginas existentes nela, acessa essas outras páginas, coleta novamente novos links e assim sucessivamente até pelo menos três níveis de cada link, o que pode gerar o acesso a milhares de páginas.

¹ *WEB CRAWLER* é um programa de computador que navega pela rede mundial de uma forma metódica e automatizada e faz coleta sucessivas das páginas Web. São utilizados principalmente para criar uma cópia de todas as páginas visitadas para um pós-processamento.

O próximo passo é acessar cada uma dessas URLs e fazer o download do código HTML de cada página. São coletadas somente páginas que retornam uma resposta de código de status 200 do servidor. Esse código significa que o acesso foi bem sucedido, como se uma pessoa abrisse um site pelo navegador. Caso a resposta seja outro código, como 404 (página não encontrada) ou 403 (proibido), por exemplo, o sistema ignora essa semente e não coleta esse site. Além desse filtro, também retiramos arquivos que não são avaliados, como imagens, arquivos para download etc.

A partir daí, essas páginas são enviadas, por meio de uma API (Interface de Programação de Aplicações, ou *Application Programming Interface* em inglês), para a ferramenta [ASES Web](#) instalada em nossos servidores. Ela faz a verificação da conformidade com os padrões do [eMAG](#), que é o modelo de acessibilidade digital do governo brasileiro, definido pela [Portaria nº 3, de 7 de maio de 2007](#) como o padrão de acessibilidade para sites do governo federal. O eMAG foi desenvolvido com a participação de especialistas e com base no documento internacional de acessibilidade na Web [WCAG 2.0](#).

A verificação feita pela ferramenta AsesWeb não é simples e [calcula o peso de cada erro](#) para gerar a nota do site. Por exemplo, erros de marcação no HTML são muito mais críticos do que erros encontrados nos arquivos CSS.

Com as páginas verificadas, conseguimos extrair informações como quantidade de erros e avisos e listar esses sites com base em alguns filtros para facilitar o consumo dos dados e a compreensão dos resultados.

Um desses filtros é o [portal www.gov.br](#). Desde abril de 2019 os sites de órgãos e entidades públicas vem saindo do seu domínio próprio para integrar o portal único. Apesar de alguns sites ainda responderem pelos dois endereços (por exemplo, o portal <https://portal.mec.gov.br> ainda funciona, mas já com anúncio da migração para <https://www.gov.br/mec/pt-br>). Sendo assim, ainda é possível encontrar resultados de sites na plataforma como <https://www.agricultura.gov.br> e <http://www.gov.br/agricultura> que eram sites distintos quando foram coletados pela primeira vez.

Também criamos um filtro para selecionar os [10 melhores sites](#) com base na média de notas das páginas. Como são muito distintos e com quantidade de páginas variada, criamos também um filtro para agrupar, por exemplo, os melhores sites entre 10 e 50 páginas, ou entre 51 e 100 páginas e assim por diante.

Outra categoria importante é a [separação por unidade federativa](#). Nesse link é possível chegar ao nível de detalhe de acessibilidade por Estados da Federação. Conforme esse indicador crescer será possível também chegar a detalhes no nível municipal.

A área de [erros mais comuns](#) nos permite compreender os tipos de erros que acontecem nas páginas governamentais. Com isso é possível identificar as principais barreiras e aplicar boas práticas que eliminam essas barreiras e criar políticas para evitar esse tipo de erro.

Nosso objetivo é fazer com que essa plataforma tenha um ciclo constante de atualizações. Que novas sementes possam ser adicionadas e que os sites avaliados possam ser revisitados com base na coleta dinâmica, que será incorporada nas próximas coletas. A coleta dinâmica vai permitir coletar as páginas que foram atualizadas e avaliá-las novamente pela plataforma.

A primeira atualização dos dados aconteceu em abril de 2022 e já está disponível na plataforma. Em breve faremos um novo artigo sobre os resultados. A seguir, apresentamos o resultado de uma análise detalhada da coleta e avaliação executadas em 2021.

Resultados 2021

A [primeira coleta](#) avaliou 418 websites governamentais. Desses sites foram coletadas 267.090 páginas. A tabela a seguir mostra um breve resumo do resultado inicial listando o número de sites e páginas relacionados ao percentual de conformidade com o AsesWeb:

PERCENTUAL DE CONFORMIDADE	SITES AVALIADOS	PÁGINAS AVALIADAS
95 - 100%	0	2.407
85 - 94.99%	60	74.704
70 - 84.99%	282	160.397
0 - 69.99%	76	29.582

A partir dos dados extraídos da plataforma em 21/02/2022, faremos algumas análises do seu conteúdo.

Vale lembrar que um site com nota máxima na plataforma não atesta que ele seja totalmente acessível. Para garantir a acessibilidade, além da verificação automática, é necessário uma verificação humana de itens que não podem ser detectados por uma ferramenta como essa (por exemplo, se um texto alternativo descreve corretamente a imagem exibida e não aparece apenas como “foto” ou “imagem”). O próprio eMAG [aponta a verificação manual](#) como uma etapa essencial na avaliação de acessibilidade de um site.

Erros mais comuns

Em todos os 418 sites coletados e em mais de 267 mil páginas analisadas, temos pouco mais de 24 milhões de erros. É um número significativo, considerando que apenas 2.400 páginas atingiram o nível de conformidade entre 95% e 100%, menos de 1% são páginas com poucos ou nenhum erro. Ao levantarmos a quantidade de páginas [sem nenhum erro](#), chegamos a apenas uma.

Esse número grande de erros assusta, mas vale lembrar que esse é o total e nem todos os sites têm uma quantidade tão grande de erros. Por exemplo, pouco mais de 21% dos sites têm em média menos de 20 erros por página, enquanto pouco mais de 15% estão na lista de sites com mais de 100 erros por página. Porém existem sites com muitos erros, elevando a média de erros.

O erro mais comum está relacionado a conformidade com os padrões Web. Está presente em mais de 99% das páginas avaliadas pela plataforma. Esse erro pode ser identificado utilizando o verificador de marcação do W3C disponível em <http://validator.w3.org>.

Não respeitar padrões Web impacta diretamente na acessibilidade da página, já que uma imagem sem texto alternativo (erro comum e que é detectado pelo verificador de marcação) pode impedir o acesso de uma pessoa com deficiência ao conteúdo.

Sites por UF

A separação por sites em unidade federativa permite identificar os que têm a melhor nota na ferramenta. A melhor unidade é a do Ceará, com pontuação pouco acima de 82%. Considerando que as demais estão abaixo desse valor, não temos nenhuma unidade federativa na primeira faixa de conformidade, entre 95% e 100%.

Fazer uma comparação entre elas não traz informação relevante por serem muito distintas entre si, pois algumas unidades têm poucos sites e páginas e outras tem dezenas de sites agrupados.

[10 melhores sites](#)

O site com a melhor nota da plataforma é o <http://educacaoconectada.mec.gov.br>, com pouco mais de 91% de conformidade. São 142 páginas avaliadas, sendo que a melhor tem mais de 98% de conformidade e a pior pouco mais de 88%.

Para permitir alguma comparação, a plataforma permite selecionar filtros pela quantidade de páginas no site. Quando selecionamos o primeiro filtro, com páginas entre 10 e 50 páginas, encontramos sites com mais de 90% de conformidade e alguns com pouco mais de 87%. A diferença pode parecer pequena, mas considerando que estamos falando dos melhores sites, seria ideal que estivessem ao menos na linha de conformidade entre 95% e 100%.

O portal www.gov.br, que vem centralizando serviços do governo, aparece na lista dos 10 melhores sites. Isso aponta que centralizar a padronização dos sites no www.gov.br pode estar surtindo efeito para a acessibilidade, afinal, o site tem nota de 92% com 13.429 páginas verificadas.

O site da Câmara de Itaquaquecetuba não aparece entre os 10 melhores, mas tem a maioria das suas 349 páginas com poucos erros. Esse é o único site com uma página sem erros, mas também conta com 69 páginas com somente um erro. Talvez você esteja se perguntando: porque um site com poucos erros tem nota tão baixa? Isso acontece por causa do tipo do erro. Falta de texto alternativo em imagens tem um peso grande na construção da nota final.

[Portal www.gov.br](#)

Na atual coleta, temos 88 sites que fazem parte do portal gov.br. Existem 8 sites que possuem pontuação maior que 95% e o com a melhor pontuação, apesar de ter apenas duas páginas, conta com nível de conformidade acima de 99%. Vale ressaltar que esses sites não são contabilizados dentro do filtro de “10 melhores sites” e sim como parte do site principal www.gov.br.

Mesmo avaliando com cuidado as páginas do site mais bem colocado no gov.br, podemos notar que ainda há erros relacionados a padrões Web. São 55 e 57 erros respectivamente nas duas páginas.

Conclusão

O Ceweb.br apresentou neste texto os resultados e análise da primeira coleta de dados da plataforma TIC Web Acessibilidade. Foram coletados dados de 418 sites do domínio “.gov.br”, correspondendo a pouco mais de 267 mil páginas.

O trabalho realizado teve como objetivo mostrar o nível de conformidade com padrões de acessibilidade nas páginas Web do governo brasileiro. Não foi feita uma avaliação da acessibilidade dessas páginas, pois isso demandaria uma análise humana complementar. O nível de conformidade com padrões de acessibilidade mostra boa parte das causas das barreiras de acesso conhecidas.

Os resultados mostraram que apenas uma página Web sem erro de conformidade foi encontrada no “.gov.br”. A grande maioria das páginas analisadas apresentou erros de marcação HTML.

Segundo critérios da ferramenta ASES de verificação de acessibilidade, sites e páginas Web são classificados em faixas. A primeira, que é de 95% ou mais, representa um site/página com poucos erros. Na faixa entre 85% e 94,99%, o resultado requer atenção. Entre 70% e 84,99%, há um número considerável de erros. Abaixo de 70%, o site/página apresenta muitos erros.

Análise global dos números mostram que o nível de conformidade aos padrões Web de acessibilidade pelos governos federal e estaduais brasileiros requer muita atenção pois a maioria dos sites e páginas estão na faixa entre 70 e 84,9%, isto é, possuem um número considerável de erros que oferecem barreiras de acesso para pessoas com deficiência.

Pouco menos de 1% de páginas avaliadas apresentou pouco ou nenhum erro. Quando verificamos os indicadores anteriores da pesquisa TIC Web para o quesito acessibilidade na Web, desde 2010, percebemos que, mesmo considerando que foram utilizadas metodologias diferentes em ambos os levantamentos, o nível de conformidade histórico nos sites e páginas no domínio “.gov.br” é muito preocupante, não vem melhorando e mantém um processo de exclusão digital da pessoa com deficiência.

Estima-se que a Web governamental tenha pouco mais de 18.800 sites e milhões de páginas. A ferramenta ainda conta com um número limitado de sites avaliados. Mas, com a expansão da base de coleta de dados, a plataforma poderá cada vez mais retratar um quadro amplo da Web no Brasil, com diferentes recortes. Se você administra um site com final “.gov.br”, envie para a equipe do Ceweb.br o link para que seja adicionado às próximas coletas e verificações.

A TIC Web Acessibilidade é uma aliada dos administradores de páginas públicas para a verificação e manutenção da acessibilidade. Ela também é uma grande ferramenta para deixar mais transparente para a população brasileira como anda a conformidade com os padrões de acessibilidade exigidos pela lei.