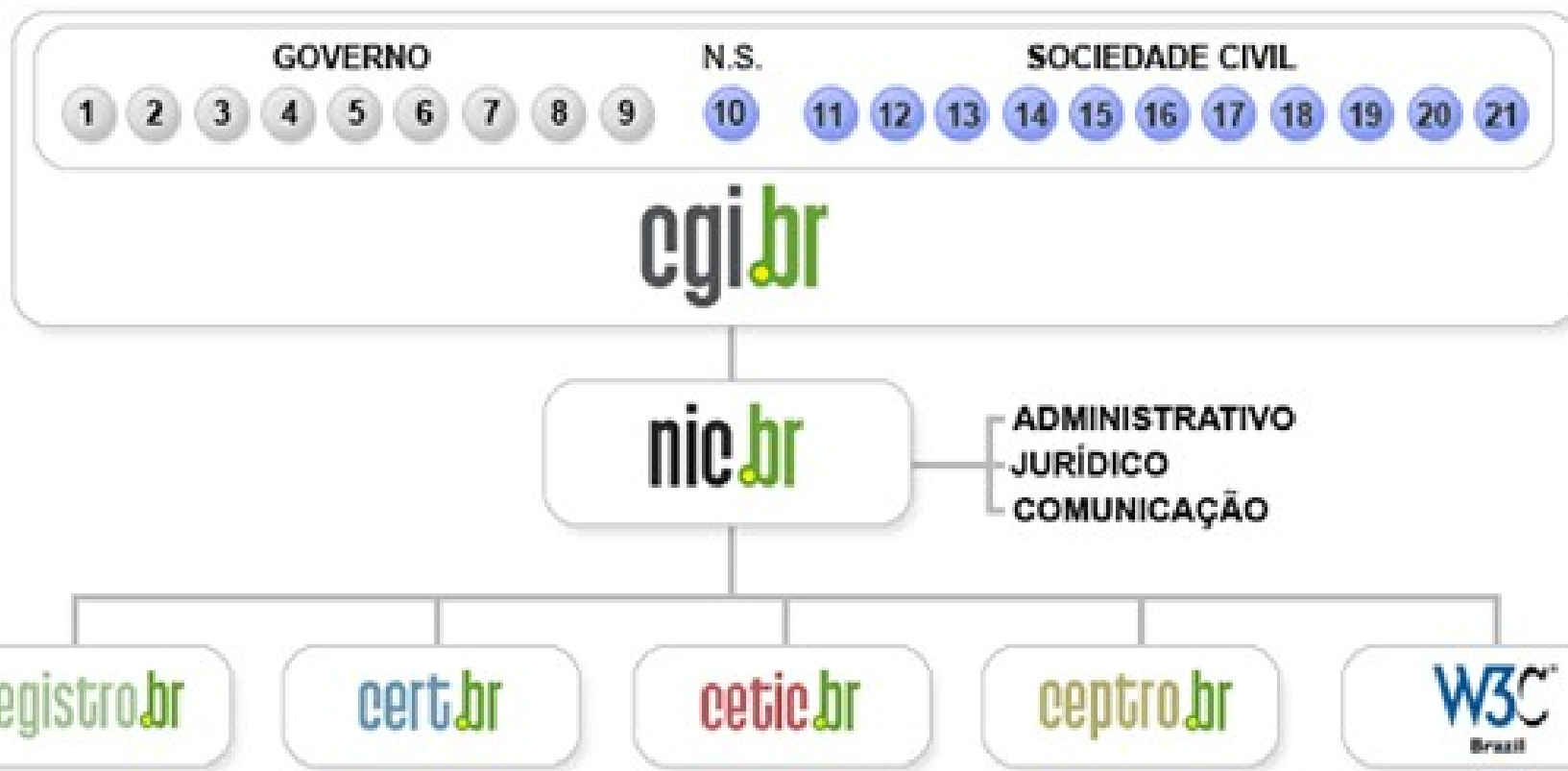




Analizador Automáticos de Sites para o .br



Introdução



O que é o Projeto?

>> *Monitor de sites*



<http://vpereiro.files.wordpress.com/2010/05/dominio.jpg>

*Realiza a verificação **automática e periódica** da qualidade de sites de usuários*

Características – fase 1

- Verificação da **quantidade** de páginas do site
- Verificação de suporte **IPv6**
- Distribuição de **idiomas** utilizados no site
- Proporção de páginas da Web aderentes aos padrões **HTML do W3C**
- Proporção de páginas da Web aderentes aos padrões de **acessibilidade WCAG**

Características – fase 1

- **Idade** das páginas
- **Geolocalização** do servidor
- **Sincronização de tempo** do servidor web
- Acesso aos relatórios na Web

Características – fase 2

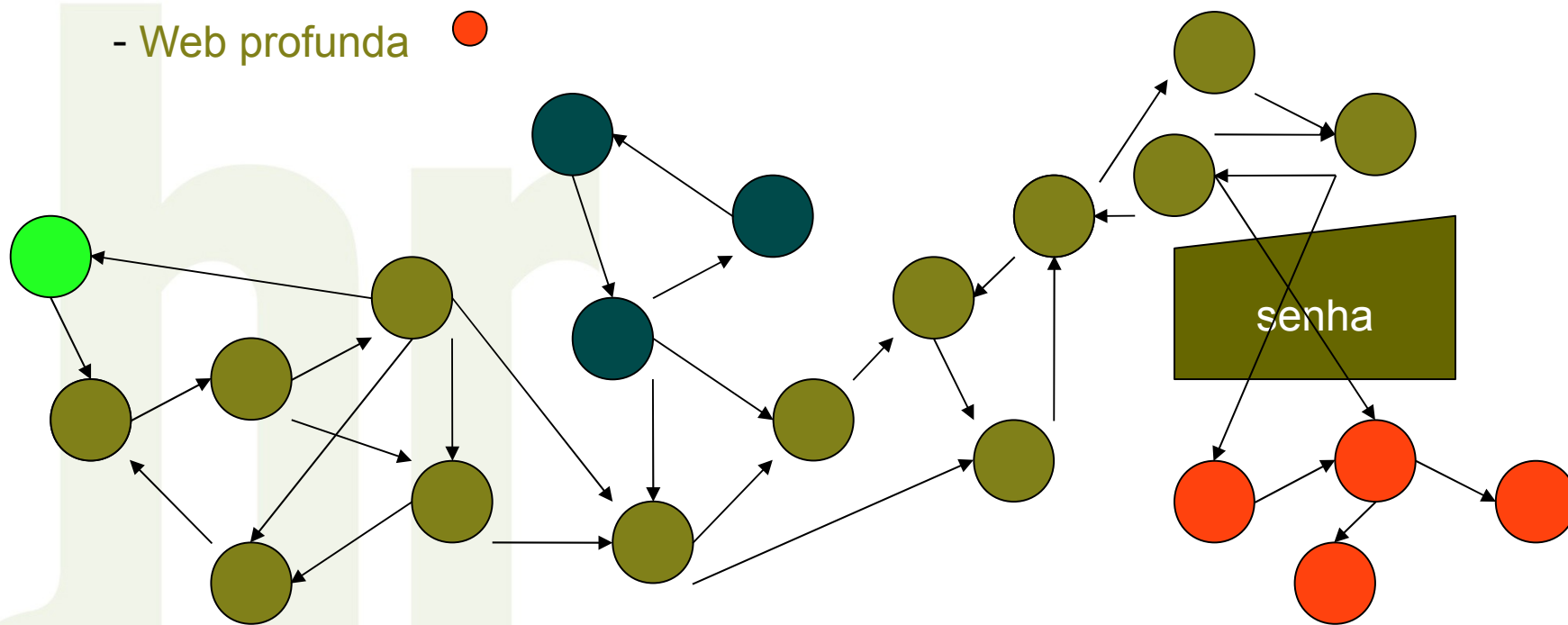
- Validação de padrões de **CSS do W3C**
- Verificação de suporte a **DNSSEC**
- Verificação de características de **segurança**
- Envio de relatórios por e-mail

Desafios técnicos



Limites para o estudo

- A Web é uma rede
- Index ●
- Ilhas ●
- Web profunda ●

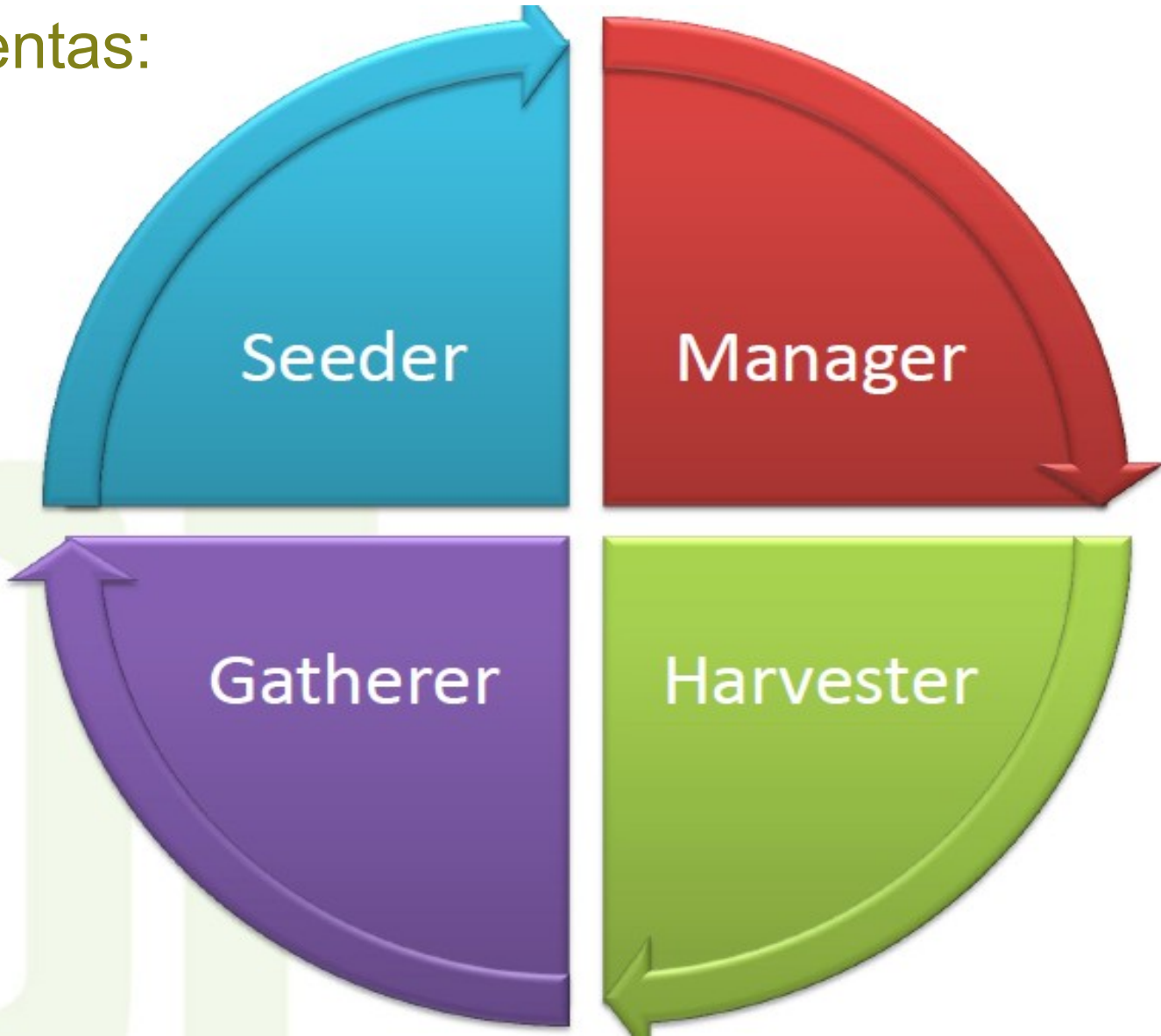


Ferramentas

- **WIRE** = webcrawler
 - **Software livre**
 - **Desenvolvido no Chile para estudos acadêmicos sobre a Web**
 - **Fizemos correções e adaptações no NIC.br**
- **Análise Internet**
 - **Desenvolvido no NIC.br**
 - **Localização Geográfica, IPv6, Sincronização, Aderência a padrões, Tipos de documentos, Tipos de Servidores, Tempo de Resposta.**



Ferramentas: WIRE



Além disso...

- Calcula a **quantidade e tamanho** das páginas
- Informa e **classifica as URLs**
- Informa a **idade das páginas**
- Diferencia páginas **estáticas e dinâmicas**
- Calcula diferentes índices de páginas como **pagerank e siterank**
- Identifica o **idioma** das páginas

Ferramentas: Análise Internet

- Realiza 5 tipos de teste:
 - Carregamento dos dados do WIRE
 - Validação de páginas
 - Testes sobre sítios e servidores
 - Testes sobre links encontrados
 - Download de Arquivos adicionais



Ferramentas: Análise Internet

- **Validação de páginas:**
 - Validador de HTML do W3C
 - Validados de acessibilidade ASES
- Arquitetura Distribuída
 - Aumenta a velocidade de processamento



<http://xenlights.com/images/SoftwareValidation.jpg>

ASES

Aderência à padrões HTML

- Universalidade do acesso:
 - Acesso sem barreiras
 - Compatibilidade
 - Acessibilidade
 - Ganho de desempenho
 - Economia de banda
 - Código mais simples e fácil de atualizar
 - Melhor visibilidade em ferramentas de busca
 - Evita instabilidade e versões de páginas



[http://cssti.files.wordpress.com/2009/06/26032008www.jpg?
w=363&h=362](http://cssti.files.wordpress.com/2009/06/26032008www.jpg?w=363&h=362)

Aderência a padrões de acessibilidade

- Garante acesso universal aos sites Web:
 - O modelo de acessibilidade considerado foi o e-MAG
 - O e-MAG tem como referência as diretrizes de acessibilidade do W3C publicadas no WCAG
 - A ferramenta utilizada para a validação foi o ASES, criado pela SLTI do Ministério do Planejamento



Ferramentas: Análise Internet

- Testes sobre Sítios e Servidores: **Reposta**
 - Realiza uma requisição HEAD
 - Obtêm:
 - Tempo de resposta
 - Tipo de servidor
 - Diferença de tempo
 - Ipv4



IPv6

- Todo computador na Internet necessita de um endereço: o número IP
- A numeração usada atualmente (IP versão 4) está se esgotando. Previsões indicam que o estoque global acabará no 1o. Semestre de 2011.
- Uma nova versão de IP está sendo implantada na Internet: o IPv6
- O IPv6 é necessário para:
 - Garantir o crescimento e desenvolvimento da rede
 - Inclusão digital
 - Internet das coisas

Ferramentas: AnáliseInternet

- Testes sobre Sítios e Servidores: IPv6
 - Não é suficiente verificar se o domínio possui ipv6

ipv6.google.com.br

www.v6.facebook.com

- Utiliza variações do nome do site: [www6](#), [www.ipv6](#), [ipv6](#)
- Realiza ping6 e requisição GET ao endereço
- Verifica se o **NameServer** possui suporte a IPv6



Sincronização com a Hora Legal Brasileira

A Sincronização com a Hora Legal Brasileira (que equivale ao padrão mundial UTC) é recomendada pelo CGI.br:

<http://www.cgi.br/regulamentacao/resolucao2008-009.htm>

Todo servidor deve estar sincronizado e, na medida do possível, também computadores pessoais. Isso é importante para:

- segurança
- funcionamento correto das aplicações

Deve-se utilizar o NTP: <http://ntp.br>. É uma configuração simples de ser realizada, porém pouco conhecida. Os servidores de tempo são oferecidos pelo NIC.br, em conjunto com o Observatório Nacional.

A medida foi realizada obtendo-se a hora dos servidores, via HTTP, e comparando-a com a hora correta.

Testes de Sincronização de Tempo

Resposta do servidor Web

Depende da qualidade da rede e processamento do servidor

Precisão de ~ **seg**

Protocolo NTP

Mais preciso: ~ **μ seg**

Mais difícil de ser obtido



Localização Geográfica dos Servidores

O indicador mostra a proporção de servidores localizados no Brasil e no exterior. Usa dados de uma base especializada (GeoIP MaxMind) que tem cerca de 95% de exatidão.

Servidores fora do Brasil podem ser mais baratos, contudo implicam em maior lentidão no acesso, e no uso de canais de comunicação de internacionais, com alto custo, levando a um aumento nos custos de acesso no Brasil.

É importante que os servidores que hospedam sítios destinados aos internautas do Brasil estejam hospedados no Brasil.

Ferramentas: Análise Internet

- Testes sobre Sítios e Servidores: **Geolocalização**
 - Geo Localização de servidores por IP
 - Integra API do **GeoIP®**



Status

- Geração de relatórios web
 - Teste de resposta
 - Teste de NTP
 - Teste de geolocalização
 - Validação de páginas
 - Teste de IPv6
 - Contagens idiomas, páginas, idade

SITE

- Criar novo Site
- Listar todos Sites

RELATÓRIO DO SITE

- Listar todos Relatórios do Site

USUÁRIO

- Efetue aqui seu Logout

Mostrar Relatórios do Site

Site : www.ipv6.br lkajsdikfj 1
 Data do Relatório 15/01/11
 :

Tipo de Servidor : Apache

Ipv4 : 200.160.4.22

Ipv6 : 2001:12ff:0:4:0:0:0:22

Número total de 268 páginas :

Tamanho médio 21517.8027 da página :

Geolocalização : BR +

Offset de tempo 133 via Http : ✖

Delay via Http : 707 ⚠

Tem Ntp : true +

Localização geográfica do servidor. Esse é obtido apartir da base de dados da ferramenta Geolp da MaxMind. Caso o seu servidor estaja fora do pals em que o site é mais acessado, podem correr reclamações de lentidão no acesso ao seu site.

Obrigado !

- Contatos:
 - **web@ceptro.br**
 - **moreiras@nic.br**
 - **heitor@nic.br**
 - **phadek@nic.br**
- Links úteis
 - **<http://ceptro.br/asa>**
 - **<http://w3c.br/>**
 - **<http://validator.w3.org/>**
 - **<http://www.governoeletronico.gov.br/acoes-e-projetos/e-MAG/ases-avaliador-e-simulador-de-acessibilidade-sitios>**
 - **<http://www.cwr.cl/projects/WIRE/>**
- Perguntas???