Technological study of Brazilian government websites

Newton Calegari / Reinaldo Ferraz – CEWEB.br

nic.br

Newton Calegari

Reinaldo Ferraz

ceweb.br nic.br cgi.br

Introduction

- Colect webpages under .gov.br Top Level Domain

- Perform validation

Technical enviroment

- WIRE-Nic is a fork of the open souce project WIRE (Web Information Retrieval Environment) that was developed by Center for Web Research from University of Chile. This fork brings some bug fixes and modifications to the original system.

  - A web crawler.

  - Tools for extracting statistics from the collection.

  - Tools for generating reports about the collection.

https://sourceforge.net/projects/wire-nic/

Technical enviroment

- ConNeCTOR - Convenient Network Characteristics Testing

  Organized Routines

  - Created to support the carrying out of tests related to geo-

    localization of the servers and adherence to HTML and

    accessibility (eMAG / WCAG) standards.

https://sourceforge.net/projects/connector-nic/

Collecting Webpages

- Some conditions the website should satisfy:

  - owning a domain with identified TLD .gov.br;

  - domain identified with the letters –uf.gov.br, that belong to different states of the Brazilian federation.

http://www.prefeitura.sp.gov.br/
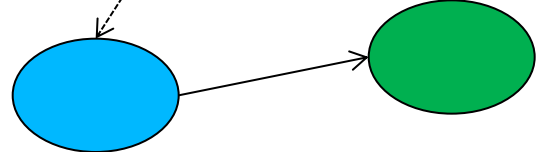
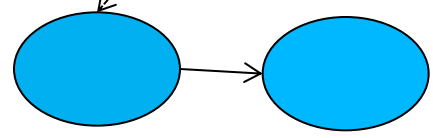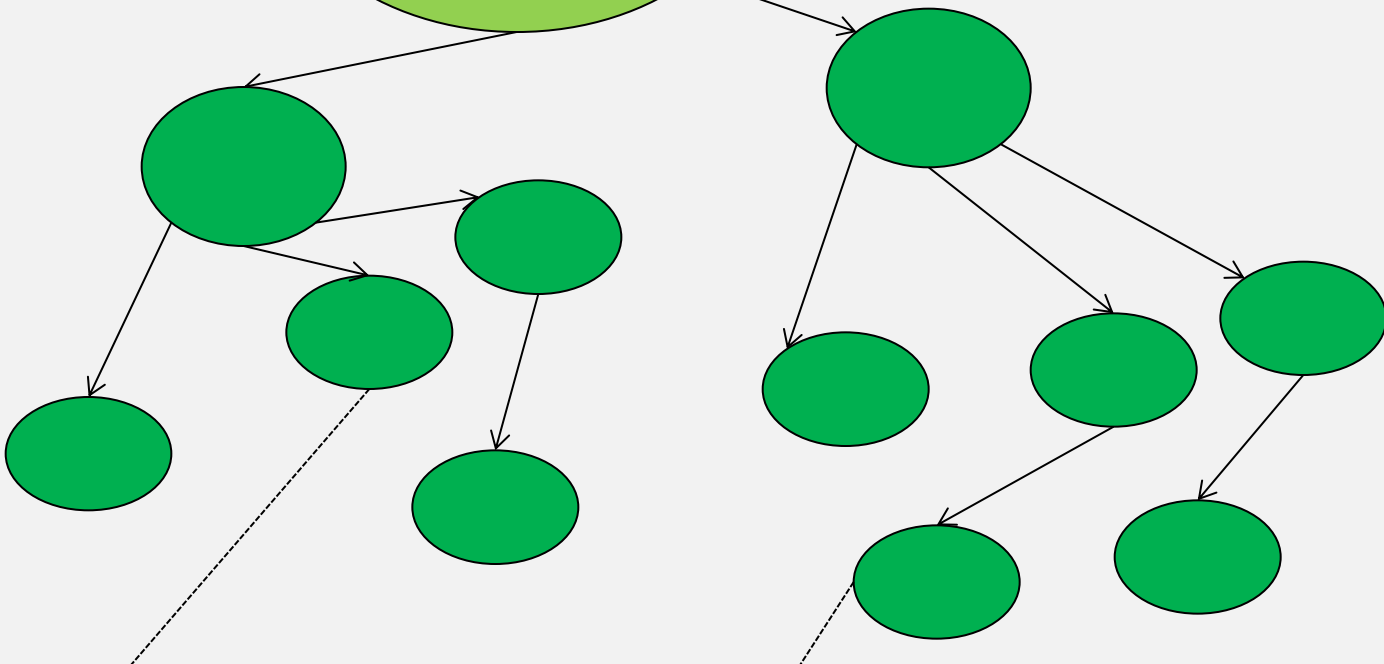http://www.nic.br/

http://www.google.com.br/

Collecting Webpages

- The information in the domains is supplied by the authority for the register of domains in Brazil, Registro.br, with the authorization of the minister for planning, responsible for the use of the domains under .gov.br, and by the companies linked to the respective state governments.

.gov.br websites

List of websites under .gov.br (seed)

Validation tools

- W3C Validator: Check W3C Standards compliance

  - http://validator.w3.org

- ASES - Avaliador e Simulador de Acessibilidade em Sítios

  - http://asesweb.governoeletronico.gov.br/ases/

Data analyzed

- Size of the Brazilian web

- Language of Brazilian web pages

- Geographical location of servers

- W3C standards compliance

- e-Mag standards compliance

Results

- Size of the Brazilian web – in Gigabites

2011
169 GB

2015
286 GB

Results

- Size of the Brazilian web – number of websites

2011
12.891

2015
11.080

Results

- Size of the Brazilian web – number of websites

The average size of websites in 2015: 25 MB

Biggest website in 2015:   4.6 GB

Results

- Size of the Brazilian web – number of webpages

2011

6.334.054

2015

8.323.478

Results

- Webpage language – Webpages in portuguese

2011
97%

2015
98.25%

Results

- Number of servers located in Brazil

2011

93%

2015

90%

Results

- Webpages which adhere to W3C standards

2011
5,02%

2015
4,56%

Results

- Webpages which adhere to W3C standards

2015:
Pages with less than 10 erros: 21%
Pages with more than 100 error: 10%

Results

- Webpages which adhere to e-Mag standards

  (web accessibility)

2011
2%

2015
5,7%

Future work

- Update the validator ASES to reach eMag 3 and WCAG
  2.0
- Check other HTML5 features

Thank you
www.ceweb.br

@ reinaldo@nic.br   Ⓣ @reinaldoferraz

June, 11th, 2015

nic.br cgi.br

www.nic.br | www.cgi.br