The Archival Acid Test

This page was setup to test the capabilities and shortcomingsof archival web crawlers. More information about the rationale and individual tests can be found below. For questions/comments contact Mat Kelly.

The Tests

The Basics (6 tests)

Javascript (8 tests)

Advanced Features Tests (4 tests)

More Information

The Motivation

The purpose of this web page is to test the capability of web crawlers intended for archiving (e.g., Heritrix) and potentially their corresponding replay systems (e.g., Wayback).

Tests' Rationales

Tell an archival crawler to capture this page. Replay the capture in an archival replay system. Any non-blue squares means that some aesthetic or functionality capability of the page on the live web is not being preserved into the archive.

The Basics

- 1a - Local image, relative to the test
- 1b - Local image, absolute URI
- 1c - Remote image, absolute
- 1d - Inline content, encoded image
- 1e - Scheme-less resource
- 1f - Recursively included CSS

JavaScript

- 2a - Script, local
- 2b - Script, remote
- 2c - Script inline, DOM manipulation
- 2d - Ajax image replacement of content that should be in archive
- 2e - Ajax requests with content that should be included in the archive, test for false positive (e.g., same origin policy)
- 2f - Code that manipulates DOM after a certain delay (test the synchronicity of the tools)
- 2g - Code that loads content only after user interaction (tests for interaction-reliant loading of a resource)
- 2h - Code that dynamically adds stylesheets

HTML5 Features

- 3a - HTML5 Canvas Drawing
- 3b - LocalStorage
- 3c - External Webpage
- 3d - Embedded Objects (HTML5 video)

Metadata

Reported User-Agent string: Mozilla/5.0 (Linux; Android 7.1.1; Nexus 6 Build/N6F26U) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/73.0.3679.0 Mobile Safari/537.36
Detected User-Agent:
Access Time: 2019-03-25 08:39:18