Improving interop with web-platform-tests

Philip Jägenstedt <foolip@chromium.org> at Web Engines Hackfest 2017

Slides: bit.ly/hackfest-wpt
Why care?
Mortality
Pain

Lars Berg
@larsberg

knobby rolls
larsberg.net/sketches/2017/ ...

I made this in chrome & was bummed how different it sounds in FF & safari
#threejs #tonejs #cannonjs

Lambtaco
@Lambtacotweets

I noticed that in Microsoft Edge. Seems to work fine in Chrome and Firefox. Try a different browser?

AlextheTroller
@Alex_theTroller

@Lambtacotweets Yo Alex here, decided to give 2.0 a try and the controls are iffy, enable hold to dash and try to jump left while dashing.

6:51 PM - 20 Sep 2017

Thom McGrath
@tekcor

Left: Chrome. Right: Safari. Two different tax calculations for the same address.

Adrian Roselli
@aardrian

Playback speed for <video> content works only in Windows, only in Firefox and Edge.
rosel.li/092217
Safari test by @scottohara.
What is the problem?
Tragedy of the commons

- We usually think of our browser as the product
- Web developers want a cohesive platform
- Let’s think of the platform as the product
- Neglected, so improving is easy
- Apply basic software engineering
WPT one year ago:

- No one’s primary test suite
- Harder to use than non-shared tests
- No continuous integration
- No dashboard/waterfall
- Out of sync with specs
Progress so far
2-way sync

WPT changes imported into Chromium within 12 hours

Chromium changes exported to WPT within 1 hour

Humans not in the loop, but on rotation if needed
2-way sync (export)
Per-PR testing

Catches unstable tests

Does not require tests to pass
# web-platform-tests dashboard

## Root  About

Search test files, like cors/allow-headers.htm

WPT /

---

Data below are intended for web platform implementers and do not contain useful metrics for evaluation or comparison of web platform features. Also note that tested Edge and Safari are not pre-release versions (#109, #110).

<table>
<thead>
<tr>
<th>Spec</th>
<th>Chrome 62.0 Linux 3.16</th>
<th>Edge 15 Windows 10</th>
<th>Firefox 57.0 Linux 3.16</th>
<th>Safari 10 MacOS 10.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>/2dcontext</td>
<td>1818 / 1912</td>
<td>1469 / 1742</td>
<td>1843 / 1912</td>
<td>1760 / 1908</td>
</tr>
<tr>
<td>/fileapi</td>
<td>491 / 499</td>
<td>374 / 493</td>
<td>494 / 499</td>
<td>445 / 497</td>
</tr>
<tr>
<td>/indexeddb</td>
<td>1578 / 1579</td>
<td>763 / 1312</td>
<td>1530 / 1579</td>
<td>1270 / 1579</td>
</tr>
<tr>
<td>/webcryptoapi</td>
<td>22858 / 35622</td>
<td>959 / 16961</td>
<td>17305 / 37253</td>
<td>119 / 1250</td>
</tr>
<tr>
<td>/webidl</td>
<td>436 / 445</td>
<td>249 / 445</td>
<td>440 / 445</td>
<td>191 / 297</td>
</tr>
<tr>
<td>/xmlhttprequest</td>
<td>1483 / 1563</td>
<td>734 / 1096</td>
<td>1518 / 1564</td>
<td>814 / 1373</td>
</tr>
<tr>
<td>/ambient-light</td>
<td>3 / 33</td>
<td>3 / 33</td>
<td>3 / 33</td>
<td>3 / 33</td>
</tr>
<tr>
<td>/apng</td>
<td>0 / 1</td>
<td>0 / 1</td>
<td>0 / 1</td>
<td>0 / 1</td>
</tr>
<tr>
<td>/assumptions</td>
<td>12 / 17</td>
<td>15 / 17</td>
<td>16 / 17</td>
<td>14 / 17</td>
</tr>
<tr>
<td>/audio-output</td>
<td>12 / 13</td>
<td>3 / 14</td>
<td>2 / 13</td>
<td>2 / 13</td>
</tr>
<tr>
<td>/background-fetch</td>
<td>23 / 179</td>
<td>14 / 72</td>
<td>23 / 179</td>
<td>7 / 16</td>
</tr>
<tr>
<td>/battery-status</td>
<td>34 / 36</td>
<td>4 / 10</td>
<td>4 / 11</td>
<td>4 / 11</td>
</tr>
<tr>
<td>/beacon</td>
<td>51 / 101</td>
<td>55 / 67</td>
<td>95 / 101</td>
<td>21 / 100</td>
</tr>
<tr>
<td>/bluetooth</td>
<td>1 / 2</td>
<td>0 / 1</td>
<td>1 / 2</td>
<td>1 / 2</td>
</tr>
<tr>
<td>/clear-site-data</td>
<td>21 / 21</td>
<td>12 / 14</td>
<td>13 / 19</td>
<td>7 / 17</td>
</tr>
<tr>
<td>/clipboard-apis</td>
<td>12 / 31</td>
<td>3 / 31</td>
<td>3 / 31</td>
<td>2 / 12</td>
</tr>
<tr>
<td>/concert</td>
<td>5 / 10</td>
<td>18 / 19</td>
<td>18 / 19</td>
<td>15 / 10</td>
</tr>
</tbody>
</table>
WPT improvements in Chromium

dom/  79% → 99%
editing/  0% → 90%
html/  74% → 93%
selection/  16% → 99%
Improving interoperability

January 30th, 2017 by Phillip Jägenstedt in WHATWG

The goal of the WHATWG’s Living Standards is to achieve interoperable implementations. With an ever-evolving web platform, we want changes to our standards to reach all implementations quickly and reliably, but from time to time there have been mishaps:

- Two table-related interfaces were ignored by implementers for a decade, but when finally removed from the standard it turned out that WebKit had just added them. Sorry!
- In 2007, the serialization algorithm for `<pre>` was changed to add an extra newline, so that it would roundtrip correctly. It was implemented in Presto (Opera) in 2012 (and the standard was tweaked), then implemented in Gecko in 2016, but they found it was not web compatible, so it was reverted. Oh well.
- We changed how non-empty URL attributes are reflected with support from two implementers, but nobody followed through with the change and it was later reverted. Ahem.

Three months ago, we changed the process for the HTML Standard to encourage writing tests and filing browser bugs for normative changes. (Normative means that implementations are affected.) This was the first step on a path towards improving interoperability and shortening the feedback cycle, and it has thus far exceeded our own expectations:
Standards

80+ specs now have a policy to ask for tests with normative changes

**WHATWG (16)**, **Service Workers**, **Web Performance WG (12)**, **IndexedDB**, **Pointer Lock**, **Payment Request**, **Web Animations**, **WebVTT**, **FX Task Force (6)**, **CSS WG (34)**, **SVG WG (8)**, **Content Security Policy**

100+ remain
Non-WPT bonus

Visualizing API support as seen by traversing JS global object

Time-based metrics

web-confluence.appspot.com
What’s next?
Automation

pointerevents/ has 71 manual tests

Automate input with WebDriver

Also: Permissions, WebRTC
More per-PR testing

Test regressions should be intentional.

Build PASSED with FYI

Started: 2017-09-29 19:26:58
Finished: 2017-09-29 19:31:53

FYI: Test "foo bar" passed on Firefox without the changes but fails with them. Please make sure this is intended.

View more information about this build on:
- WPT PR Status
- TravisCI

Fix unintended text difference in test ref
More wpt.fyi

An “interop health metric” where the incentives always point towards interop.

Per-browser reports without summary numbers, to be used for prioritization.

Also investigating time-based metrics.
More testing in standards

Visualizing the day-to-day work in specs and tests

[link to day-to-day visualization]
Make WPT the default in Chromium

Triage new test failures

Upstream as much of LayoutTests as possible

~70k files, compared to ~60k in WPT now

Challenge: duplication and pre-fork WebKit tests
Can we relax?

No!
(trick question)
You can help!

- Write and use wpt for your work
- Tell us about what’s not working
- Make 2-way sync work for everyone
  - Automatic 2-way sync in Gecko is coming!
  - Edge?
  - WebKit?
We can do it! \o/