QtWebKit layout tests
DRT and Tools

András Bécsi
(bbendix)
<abecsi@inf.u-szeged.hu>
Outline

- Update to Qt 4.6 on the BuildBot
  - Font metric issues
  - DumpRenderTree issues
  - Progress of the port measured by passing layout tests
  - manual-tester

- Monitoring the buildbot
  - BuildBot logs
    » Flakey tests
    » Broken tests
  - BuildBotMiner
Qt 4.6 issues

- Issues with fonts
  - After the update to Qt 4.6 we experienced pixel differences in many tests (~ 500)
  - After the update of the expected files however we got feedback that these metrics do only work on our BuildBot
  » More than 300 tests have incorrect layout on developers machines but are correct on the bot
Qt 4.6 font metric issues

- We tested other major distributions, but found only minor differences (5-10 tests)
  - We tried the default configuration of Qt 4.6, the same we use on the buildbot
    
    ```bash
    echo yes | ./configure --opensource --no.webkit
    Default install to /usr/local
    ```

- On my laptop I managed to reproduce the 306 failing tests
  - Suspicious package differences between the buildbot and my machine:
    - fontconfig 2.6.0 → 2.8.0
    - libfreetype 2.3.7 → 2.3.11
    - libxext 1.0.4 → 1.1.1
Qt 4.6

- Why these differences?
  - Was something changed in the 4.6 release that could cause these environment-dependent problems?
    - ~1000 now skipped tests which seem to have only metric differences could be enabled if this issue is solved

- How should we fix this
  - Change the packages to older versions?
  - Find the cause in Qt and fix that?
Proposal for new tests

- Using render-tree dumps only if really needed
  - E.g. if font metrics are important for some reason in the test
  - In this case font rendering differences between platforms may need different results

- Using `getComputedStyle()` with `dumpAsString()` if possible to validate the test

- Lobbying to change current tests to follow this proposal
DRT crashes after change to Qt 4.6

- Some tests seemed to corrupt follow-ups
  - run-webkit-tests reported DRT crash
  - No stderr output
  - Further investigation showed no crashes
    » Instead there was a non-blocking I/O EAGAIN/EWOULDBLOCK issue
- DRT didn't clean up windows opened by javascript correctly.
- Fix landed, which revealed that there are more tests involved in DRT's notifyDone() problem.
Other problems with DRT

- Phonon related crashes
  - Two tests are involved and are skipped because they cause the “crashes”
    - fast/runin/nonblock-runin.html
    - http/tests/security/local-video-source-from-remote.html

- Similar as before
  - No real crash
  - No stderr output
  - Only run-webkit-tests reports crash
Missing DRT features

- DRT was in a really bad shape
  - All components of DumpRenderTree were in jsobjects.h and jsobjects.cpp
    » After refactoring these to separate files the development of missing features progresses
    » We would need some restructuring to match the implementations of other ports

- There are still many missing features
  - Especially in
    » LayoutTestController
    » EventSender
    » TextInputController
Progress of the QtWebKit port

- Tests altogether:
  - 11720

- Tests we pass:
  - More than 6300, slowly increasing
  - 5400 tests show layout problems or are new
    - As already mentioned there are approx. 1000 tests which only have font related pixel differences, or are new and correct tests.
  - A tool for making updates a bit easier:
manual-tester

A not-so-userfriendly Qt 4 application to visually check new or pixel-different tests

git://github.com/bbandix/manual-tester.git

$ run-webkit-tests -o ~/testResultsDir
$ ./generateTestFile.sh ~/testResultsDir \
   ~/path_to_WebKit/LayoutTests
$.manual-tester pathToQtLauncher ~/testResultsDir
The style declarations which apply to the text below are:

```css
.one { color: green; }
.1 { color: red; }
.a1 { color: green; }
P.two { color: purple; }
```

This sentence should be green.
This sentence should be black, not red--class selectors cannot begin with digits in CSS1.
This sentence should be green.
This sentence should be purple.
This sentence should NOT be purple.

- This sentence should NOT be purple.

### TABLE Testing Section

<table>
<thead>
<tr>
<th>This sentence should be green.</th>
</tr>
</thead>
<tbody>
<tr>
<td>This sentence should be black, not red--class selectors cannot begin with digits in CSS1.</td>
</tr>
<tr>
<td>This sentence should be green.</td>
</tr>
</tbody>
</table>
The time-zone issue

- Most of the work on the Mac port is done in US time-zone
- Some developers didn't keep an eye on our bot
- Keeping the BuildBot green without skipping tests and implementing missing features at the same time is hard
- The logs of the buildbot are short, searching is hard
Monitoring the buildbot

- We started to monitor our buildbot one month ago to have more information
  - Now we have raw data on more than 1000 revisions
- We started to design a web-capable information system
  - BuildBotMiner

MSG=SUCCESS
SLAVE=Qt_Linux_Release
BUILD=3890
REVISION=51044
REVIEWER="Gustavo Noronha Silva"
AUTHOR="Kenneth Rohde Christiansen"
FILES_CHANGED=11
PATCH_LLOC=744
DATE="2009-11-16"
JSCREGRUN=1119
JSCREGFAIL=0
JSCREGFIX=0
RUN=6227
OK=6158
NEW=26
FAIL=19
CRASH=24
TIMEOUT=0
BuildBotMiner

- Script gathers the raw info and log files
- Python / Django based web application
  - Parsing the information and logfiles
  - Loading the information to a database
- We plan to fully monitor our new bots and make the information publicly available
  » Additional information to our benchmark site
- Sophisticated queries and forms for presentin the data need to be done yet
- What is it good for?
BuildBotMiner

- Connect the buildbot and the svn trac
- Planned features
  - Monitoring and pointing out flakey tests
  - Pointing out the cause of over-the-night failing tests
  - Logging build issues
  - Gathering information about
    - Development process and progress
    - Layout test debugging progress
  - Creating fancy charts and graphs
BuildBotMiner

- More soon on our blog:

  http://webkit.sed.hu
Thanks for your patience!