ISTANBUL TECHNICAL UNIVERSITY
FACULTY OF COMPUTER AND INFORMATICS ENGINEERING

Project Description Form for Computer Project II Course

PROJECT 3A

Mining Chrome Repository

1) To investigate Chromium repository in GitHub, firstly We clone Chromium/web-page-replay repository into the local machine via Github for Desktop application.

For "git status" command output is;

```
On branch master
Your branch is up-to-date with 'origin/master'.
Untracked files:
  (use "git add <file>..." to include in what will be committed)

      AUTHORS

nothing added to commit but untracked files present (use "git add" to
track)
```

a)
```
$git log --name-status --since="2015-03-15" (terminal output for this
script);
```

We enter terminal prompt $ sh ./git-status.sh

—Every commit outputs include time, author and edited files

```
commit d23e5a88d82d1e1f6ce759f701bb67d7821df8a1
Author: Nikolaos Papaspyrou <nikolaos@chromium.org>
Date:   Fri May 13 10:15:37 2016 -0700

    Add mechanism to trim HTTP response headers

    The main motivation for this is to drop headers like
    "content-security-policy", which are known to cause problems
    with script injections in sites with strict policies.

    R=nednguyen@google.com
    BUG=#72

    Review URL: https://codereview.appspot.com/297250043 .

    Patch from Nikolaos Papaspyrou <nikolaos@chromium.org>.

M       httparchive.py
M       httparchive_test.py
commit 3cd3a3f6f06a1b87b14b9162c7eb16d23d141241
Author: Ned Nguyen <nednguyen@google.com>
Date:   Mon May 9 18:51:51 2016 -0700

    Remove jsmin/__main__.py & jsmin/test.py since these two files are not needed

    Motivation: these two files have no license headers and do not pass
    chromium license check (https://github.com/catapult-project/catapult/issues/2326)

    Review URL: https://codereview.appspot.com/298090043 .

M       third_party/jsmin/README.web-page-replay
D       third_party/jsmin/__main__.py
D       third_party/jsmin/test.py
commit fc1085dd6c499091d3266f49d3a4f56884496c4c
Author: Nikolaos Papaspyrou <nikolaos@chromium.org>
Date:   Mon May 9 09:20:44 2016 -0700
```

Fix minification of injected JS scripts

This patch uses the `jsmin` JavaScript minifier as a third-party tool:
https://github.com/tikitu/jsmin

R=nednguyen@google.com
BUG=#69

Review URL: https://codereview.appspot.com/292690043 .

Patch from Nikolaos Papaspyrou <nikolaos@chromium.org>.

```
M       script_injector.py
A       third_party/jsmin/LICENSE.txt
A       third_party/jsmin/MANIFEST.in
A       third_party/jsmin/PKG-INFO
A       third_party/jsmin/README.rst
A       third_party/jsmin/README.web-page-replay
A       third_party/jsmin/__init__.py
A       third_party/jsmin/__main__.py
A       third_party/jsmin/test.py
```

commit 706aad41fd4a7fd26c902be55c6458538d9cc90a
Author: Guilaume Abadie <gabadie@chromium.org>
Date: Wed Apr 20 10:31:09 2016 -0700

    Make certutils.generate_cert() compatible with Chrome for Linux

    This CL also implements certutils.install_cert_in_nssdb() as a
    replacement of AndroidCertInstaller.install_cert() for Linux.

    R=nednguyen@google.com

    Review URL: https://codereview.appspot.com/292500043 .

    Patch from Guilaume Abadie <gabadie@chromium.org>.

```
M       certutils.py
```

commit e2fd72f4b3c1c4339ac2995ba985d67eb46e2bd5
Author: Guilaume Abadie <gabadie@chromium.org>
Date: Mon Apr 4 07:59:19 2016 -0700

    Fix AndroidCertInstaller._is_cert_installed()

    BUG=#68
    R=nednguyen@google.com

    Review URL: https://codereview.appspot.com/296840043 .

    Patch from Guilaume Abadie <gabadie@chromium.org>.

```
M       adb_install_cert.py
```

commit 6d675bf397a29d409631b76e0bf9a15c53f77b49
Author: =Guilaume Abadie <gabadie@chromium.org>
Date: Fri Apr 1 09:29:39 2016 -0700

    Let AndroidCertInstaller._adb_shell() check the returncode of the command line.

    The 'adb shell' don't return the return code of the command line ran on the
    device. This commit let AndroidCertInstaller._adb_shell() taking care of this.

    BUG=#62
    R=nednguyen@google.com

    Review URL: https://codereview.appspot.com/294890043 .

    Patch from =Guilaume Abadie <gabadie@chromium.org>.

```
M       adb_install_cert.py
```

commit b03f84da6e90951d4275b5675533d4512a178398
Author: Ned Nguyen <nednguyen@google.com>
Date: Mon Mar 21 20:57:55 2016 -0700

    Use ssl._create_unverified_context() only if available

    It's only available Python 2.7.9 or later.

    BUG=#66

```
        R=nednguyen@google.com

        Review URL: https://codereview.appspot.com/289570043 .

M       httpclient.py
```

commit 41d60dd479b90f3be04978182b61e8826997de35
Author: Ned Nguyen <nednguyen@google.com>
Date: Tue Mar 15 14:00:50 2016 -0700

```
        Force removing certificate with -f

        TBR=perezju@chromium.org
        BUG=https://github.com/chromium/web-page-replay/issues/65

        Review URL: https://codereview.appspot.com/289530043 .

M       adb_install_cert.py
```

commit 7e572527906e28e819c643c97d13c1725536a719
Author: Ned Nguyen <nednguyen@google.com>
Date: Mon Mar 14 21:40:33 2016 -0700

```
        Ignore certificate error when fetching https

        Root cause:
        Since python 2.7.9, httplib.HTTPSConnection check for certificate error which
        cause recording failure on Mac. This patch turn off certificate verification.

        BUG=https://github.com/chromium/web-page-replay/issues/64
        TBR=perezju@chromium.org, petrcermak@chromium.org

        Review URL: https://codereview.appspot.com/287550043 .

M       httpclient.py
M       httpclient_test.py
A       test_utils.py
```

commit c8418ec8be9fb802e26758c194c71ed9913a67dd
Author: Ned Nguyen <nednguyen@google.com>
Date: Mon Mar 14 07:56:05 2016 -0700

```
        Refactor adb shell command to use adb_shell method

        BUG=https://github.com/chromium/web-page-replay/issues/62
        R=perezju@chromium.org

        Review URL: https://codereview.appspot.com/288600043 .

M       adb_install_cert.py
```

etc.
2)
For commit counts each developers we use "git-summary.sh" (for Top developers.)
Each line in the output represent commit count for that developers, developers
username, and ratio of him/her commits in the project.

```
$ sh ./git-summary.sh


  project  : web-page-replay
  repo age : 6 years
  active   : 306 days
  commits  : 697
  files    : 162
  authors  :
```
//top developers

```
196    tonyg@google.com        28.1%
137    mbelshe@chromium.org    19.7%
 71    slamm@google.com        10.2%
 66    Mary Ruthven             9.5%
 31    steve.lamm@gmail.com     4.4%
 30    simonjam@chromium.org    4.3%
 24    1              3.4%
```

//top developers

```
21    Steve Lamm              3.0%
17    stevelamm               2.4%
16    nednguyen               2.3%
13    Ned Nguyen              1.9%
10    nednguyen@google.com    1.4%
 8    towlarn@google.com      1.1%
 8    rmcloughlin@google.com  1.1%
 8    Erik Chen               1.1%
 3    steve.lamm              0.4%
 3    Yoav Weiss              0.4%
 3    lzheng@chromium.org     0.4%
 2    Guilaume Abadie         0.3%
 2    Hubert Wu               0.3%
 2    Hugo Holgersson         0.3%
 2    Megha Mohabey           0.3%
 2    Nikolaos Papaspyrou     0.3%
 2    ziga@chromium.org       0.3%
 2    Tony Gentilcore         0.3%
 1    zwr                     0.1%
 1    =Guilaume Abadie        0.1%
 1    Balazs Kelemen          0.1%
 1    Chase Phillips          0.1%
 1    Cheng-Yu Lee            0.1%
 1    Ethan Kuefner           0.1%
 1    Jeremy Moskovich        0.1%
 1    Ting-Yuan Huang         0.1%
 1    Zhen Wang               0.1%
 1    chrisgao@chromium.org   0.1%
 1    colin-scott             0.1%
 1    dayoung.shin            0.1%
 1    erikchen                0.1%
 1    fdeng@chromium.org      0.1%
 1    hubertwu                0.1%
 1    sullivan@chromium.org   0.1%
 1    tonygentilcore          0.1%
 1    (no author)             0.1%
```

Frequency of commits for each author can be found *commit count of author /Total project active time* formula. For top developers, results are above;

-tonyg@google.com = 0.64
-mbelshe@chromium.org = 0.44
-slamm@google.com = 0.23
-Mary Ruthven = 0.21
-steve.lamm@gmail.com = 0.1
-simonjam@chromium.org = 0.09
-rmcloughlin = 0.07

3)

To create relationship matrix between Authors and files, we use script which find edited files by that Author

```
git log --pretty="%H" --author="tonyg@google.com" | while read
commit_hash; do git show --oneline --name-only $commit_hash | tail -n
+2; done | sort | uniq
```

We enter terminal prompt $ sh ./freq.sh

(*In author section is changed according to author username)

After saved that script in a shell file, we execute it command prompt.Here is the results for author = tonyg@google.com

```
customhandlers.py
daemonserver.py
deterministic.js
dnsproxy.py
httparchive.py
httpclient.py
httpproxy.py
httpzlib.py
main.py
net_configs.py
perftracker/app/app.yaml
perftracker/app/index.yaml
perftracker/app/main.py
perftracker/app/models.py
perftracker/app/scripts/util.js
perftracker/app/styles/style.css
perftracker/app/templates/compare_set.html
perftracker/app/templates/index.html
perftracker/app/templates/search.html
```

```
perftracker/app/templates/view_rtt.html
perftracker/app/templates/view_set.html
perftracker/app/templates/view_summary.html
perftracker/extension/background.html
perftracker/extension/manifest.json
perftracker/extension/script.js
perftracker/extension/server.js
perftracker/extension/start.js
perftracker/runner.py
perftracker/runner_cfg.py
persistentmixin.py
platformsettings.py
replay.py
replayspdyserver.py
script_injector.py
servermanager.py
setup.py
third_party/ipfw_win32/README.txt
third_party/ipfw_win32/ipfw.exe
third_party/ipfw_win32/ipfw.sys
third_party/ipfw_win32/netipfw.inf
third_party/ipfw_win32/netipfw_m.inf
third_party/throxy/LICENSE
third_party/throxy/README
third_party/throxy/README.web-page-replay
third_party/throxy/setup.py
third_party/throxy/throxy.py
trafficshaper.py
trafficshaper_test.py
wpr_cert.pem
```

Obviously there are too much edited files in the project. So, to limit files we use
"git effort —above 15" command which returns file that edited 15 times or more.

```
httpproxy.py................................. 114        88
replay.py.................................... 109        80
httparchive.py............................... 90         73
platformsettings.py.......................... 83         63
httpclient.py................................ 63         51
trafficshaper.py............................. 37         34
dnsproxy.py.................................. 35         30
certutils.py................................. 29         21
sslproxy.py.................................. 28         14
adb_install_cert.py.......................... 25         18
sslproxy_test.py............................. 21         12
customhandlers.py............................ 17         12
```

File/Author Martix

	httppro xy.py	repla y.py	httpa rchiv e.py	platf ormse tting	httpc lient .py	traff icsha per.p	dnspr oxy.p y	certu tils. py	sslpr oxy.p y	adb_i nstal l_cer	sslpr oxy_t est.p	custo mhand lers.
tonyg@ google.	1	1	1	1	1	1	1	0	0	0	0	1
mbelsh e@chro	1	1	1	1	0	1	1	0	0	0	0	0
slamm @goog	1	1	1	1	1	1	1	0	0	0	0	1
Mary Ruthve	1	1	0	1	0	0	1	1	1	1	1	0
steve.l amm@	1	1	1	1	1	1	1	0	0	0	0	0
simonj am@ch	1	1	1	1	1	1	1	0	0	0	0	1
rmclou ghlin	1	1	1	1	1	0	0	0	0	0	0	1

After gathering information about edited files and authors, we create a matrix which shows us relationship between files and authors.

As seen in the matrix top developers are all related each other because of the first two columns.

Graph is shown below for first and second developers for simplicity.

1. tonyg@google.com
2. mbelshe@chromium.org