

Headless browsers and friends

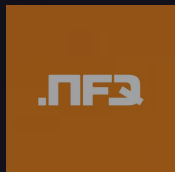
Aurelijus Banelis



Aurelijus Banelis

Software developer
aurelijus.banelis.lt
aurelijus@banelis.lt

PGP public key rsa2048/ **539B6203**
Key fingerprint = 130D C446 1F1A 2E50 D6E3
 3DA8 3202 05E7 539B 6203



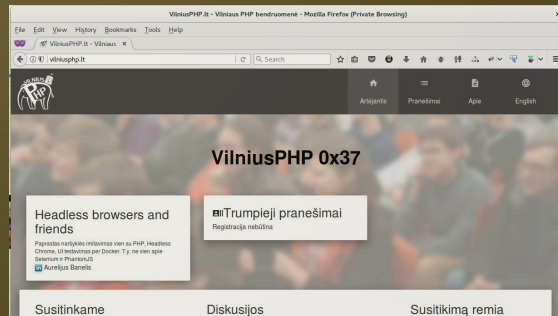
Headless browsers and similar technologies

```
VilniusPHP.lt - Vilniaus PHP bendruomenė (p1 of 5)
(BUTTON) VilniusPHP.LT
* Artejantis
* Pranešimai
* Apie
* English

VilniusPHP 0x37

Headless browsers and friends

Paprastas naršyklės imitavimas vien su PHP, Headless Chrome, UI
testavimas per Docker. T.y. ne vien apie Selenium ir PhantomJS
LinkedIn: Aurelijus Banelis Aurelijus Banelis
Trumpieji pranešimai
Registracija nebūtina
(NORMAL LINK) Use right-arrow or <return> to activate.
```



Headless browsers and similar technologies

Headless browsers and similar technologies

To automate manual tasks

PHP

**Simple way to simulate
browser**

CHROME

**Headless Chrome for
JavaScript**

DOCKER

Edge cases for UI testing

PHP

**Simple way to simulate
browser**

CHROME

**Headless Chrome for
JavaScript**

DOCKER

Edge cases for UI testing

One line to fetch

```
<?php
```

```
$html =  
    file_get_contents(  
        "http://vilniusphp.lt"  
    );
```


Checking status... 2017-05-26T20:14:02+00:00



<http://aurelijus.banelis.lt>: Took: 2.63078: Size: 0



<http://aurelijus.banelis.lt/pranesimai/>: Took: 0.10146: Size: 0



<https://aurelijus.banelis.lt>: Took: 0.10031: Size: 0



<https://aurelijus.banelis.lt/pranesimai/>: Took: 2.06: Size: 38.53

Simple status pages



<https://auginte.com/>: Took: 0.43537: Size: 9.96



<http://demo.auginte.com/>: Took: 0.27119: Size: 1.74



<https://demo.auginte.com/>: Took: 0.42991: Size: 1.74



<https://blog.auginte.com/>: Took: 1.00034: Size: 48.55

Finished checking... 2017-05-26T20:14:12+00:00



[< Browse Apps](#)



Add Configuration

Help and support >

Privacy policy >

Incoming WebHooks

Incoming Webhooks are a simple way to post messages from external sources into Slack. They make use of normal HTTP requests with a JSON payload, which includes the message and a few other optional details described later.

[Message Attachments](#) can also be used in Incoming Webhooks to display richly-formatted messages that stand out from regular chat messages.

Webhook URL

```
https://hooks.slack.com/services/AAAAA/BBBBBBB/CCCCCCCCCCCCCCCCCCCC
```

```
<?php
```

```
$data = file_get_contents(  
    "http://vilniusphp.lt");
```

```
$success = preg_match("/.+Susitinkame.+/", $data)  
    && !preg_match("/.+Warning.+/", $data);
```

```
$emoji = $success ? ':white_check_mark:' : ':volcano:';  
$status = $success ? 'OK' : 'Failed';
```

```
$webHookUrl = 'https://hooks.slack.com/services/AAAAAAAAA/BBBBBBBBB/CCCCCCCCCCCCCCCCCCCC';  
$json = json_encode(['text' => "$emoji $status", 'username' => 'php-status-bot']);  
$opts = ['http' => ['method' => 'POST', 'header' => "Content-type: application/json", 'content' => $json]];
```

```
$context = stream_context_create($opts);
```

```
$response = file_get_contents(  
    $webHookUrl, false, $context);
```

```
echo "WebSite: $status Sent to slack: $response";
```

**If not
enough...**

cURL

Guzzle

spatie/crawler

....

If not
enough...

cURL

Guzzle

spatie/crawler

....

JavaScript

PHP

**File_get_contents for
simple status page**

CHROME

**Headless Chrome for
JavaScript**

DOCKER

Edge cases for UI testing

PHP

**File_get_contents for
simple status page**

CHROME

**Headless Chrome for
JavaScript**

DOCKER

Edge cases for UI testing

Headless mode

Misc 

Headless mode allows running Chromium in a headless/server environment. Expected use cases include loading web pages, extracting metadata (e.g., the DOM) and generating bitmaps from page contents -- using all the modern web platform features provided by Chromium and Blink. To use headless, start Chrome with a command line flag: `$ chrome --headless --remote-debugging-port=9222 https://chromium.org`

Chromium status

Behind a flag

 Chrome desktop 59

 Opera desktop 46

  Opera for Android 46

Launch bug [#705916](#)

Owner(s) skyostil@chromium.org

Consensus & standardization



No public standards discussion



CLI to enable headless

```
linux-shell:~$
```

```
chromium-browser --headless  
--remote-debugging-port=9222  
http://vilniusphp.lt
```

CLI to enable headless

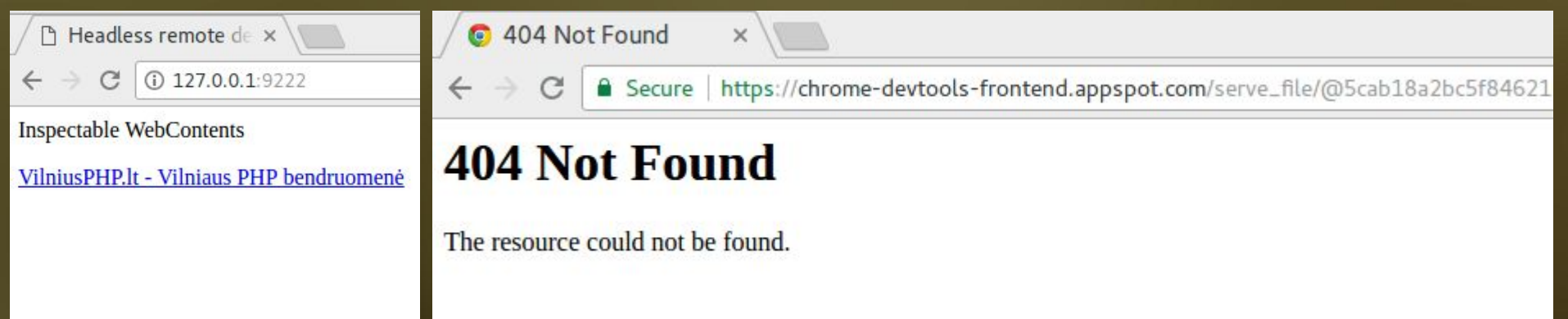
linux-shell:~\$

```
chromium-browser --headless  
--remote-debugging-port=9222  
http://vilniusphp.lt
```

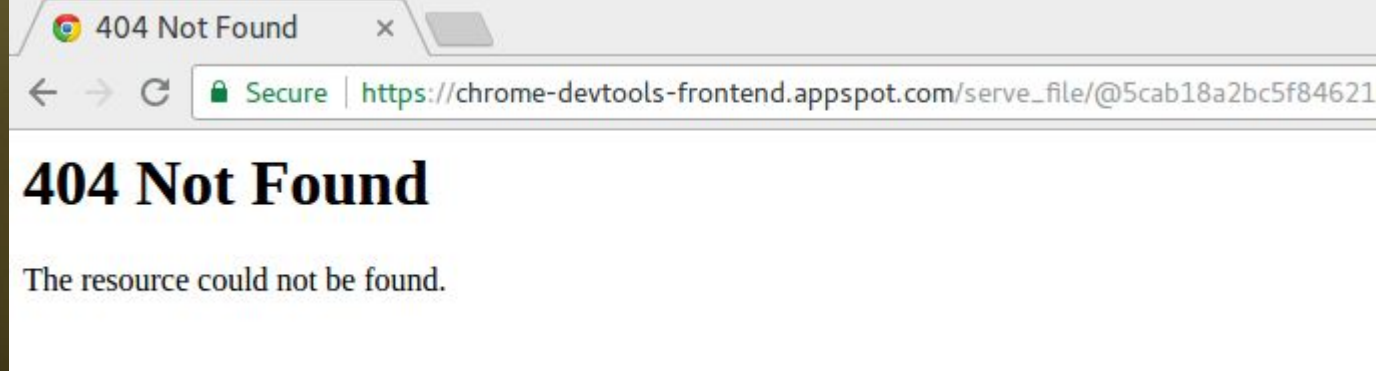
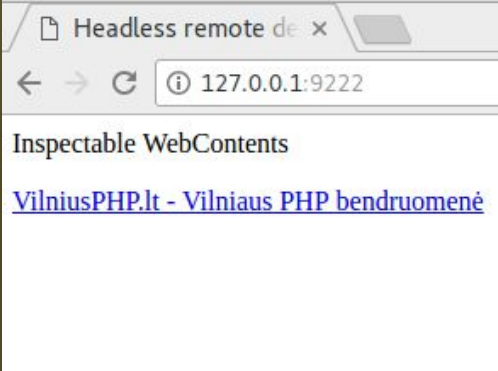
chromium-browser --headless --**dump-dom** http://vilniusphp.lt

chromium-browser --headless --**screenshot** http://vilniusphp.lt

google-chrome-**beta** --headless --disable-gpu --**print-to-pdf** http://vilniusphp.lt

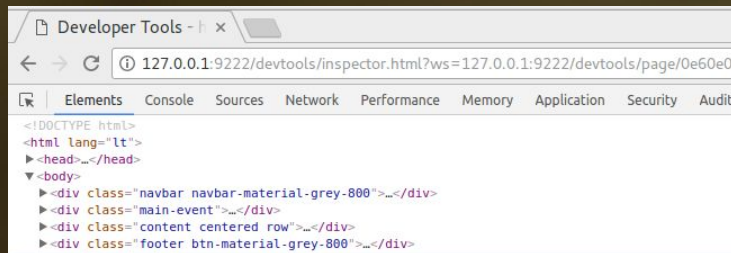


Known issues



Known issues

```
aurelijus@aurelijus-a:~$ curl 127.0.0.1:9222/json  
[ {  
  "description": "",  
  "devtoolsFrontendUrl": "/devtools/inspector.html?ws=127.0.0.1:9222/devtools/page/0e60e091-d22b-4d69-9fd8-c816dc84818f",  
}
```



Headless remote de x

404 Not Found

127.0.0.1:9222

Secure | https://chrome-devtools-frontend.appspot.com/serve_file/@5cab18a2bc5f84621

Inspectable WebContents

[VilniusPHP.lt](https://viliusphp.lt) - Vilniaus PHP bendruomenė

404 Not Found

The resource could not be found.

Known issues

```
aurelijus@aurelijus-a:~$ curl 127.0.0.1:9222/json
[ {
  "description": "",
  "devtoolsFrontendUrl": "/devtools/inspector.html?ws=127.0.0.1:9222/devtools/page/0e60e091-d22b-4d69-9fd8-c816dc84818f",

```

Developer Tools

127.0.0.1:9222/devtools/inspector.html?ws=127.0.0.1:9222/devtools/page/0e60e091-d22b-4d69-9fd8-c816dc84818f

Elements Console Sources Network Performance Memory Application Security Audit

```
<!DOCTYPE html>
<html lang="lt">
<head>
</head>
<body>
  <div class="navbar navbar-material-grey-800">
  <div class="main-event">
  <div class="content centered row">
  <div class="footer btn-material-grey-800">
```

Developer Tools

https://chrome-devtools-frontend.appspot.com/serve_file/@2a03c99a5f45c6d507af8eb2345ad68a565d1518/inspector.html?ws=127.0.0.1:9222/devtools/page/0ad7259a-5550-43b9-a6d5-1884710c1dfe&remoteFrontend=true

viliusphp.lt

VilniusPHP 0x37

Headless browsers and friends

Paprašas mitykiles imitavimas vien su PHP, Headless Chrome, UI testavimas per Docker. Ty. ne vien apie Selenium ir PhantomJS

Aurelijus Banelis

Elements Console Sources Network Performance Memory Application Security Audits

```
<!DOCTYPE html>
<html lang="lt">
<head>
  <script src="scripts/bootstrap.min.js">
  <script src="scripts/material.min.js">
  <script src="scripts/ripples.min.js">
  <script>
    $.material.init();
  </script>
</head>
<body>
  <div class="navbar navbar-material-grey-800">
  <div class="main-event">
  <div class="content centered row">
  <script>
    $.material.init();
  </script>
</body>
```

Dockerization tips

- Alpine (for small image size) – needs hacking

```
/ # chromium-browser --headless --remote-debugging-port=9222 http://vilniusphp.lt
[0601/104802.651101:WARNING:dns_config_service_posix.cc(326)] Failed to read DnsConfig.
[0601/104802.657307:ERROR:devtools_http_handler.cc(759)] Error writing DevTools active port to file
[0601/104802.658353:FATAL:udev_loader.cc(38)] Check failed: false.
[0601/104802.684887:ERROR:gl_implementation.cc(223)] Failed to load /usr/lib/chromium/libosmesa.so: Error loading shared library /usr/lib/chromium/libosmesa.so: No such file or directory
[0601/104802.789632:ERROR:broker_posix.cc(41)] Invalid node channel message
```

- Ubuntu – works, but needs proxy for debug port outside docker

FROM ubuntu:latest

RUN apt-get update \

&& apt-get install -y chromium-browser haproxy \

&& rm -rf /var/lib/apt/lists/*

EXPOSE 9222

"HTTP_USER_AGENT": "Mozilla/5.0 (X11; Linux x86_64)
AppleWebKit/537.36 (KHTML, like Gecko) Ubuntu Chromium/58.0.3029.110
Chrome/58.0.3029.110 Safari/537.36"

Simulating real user

"HTTP_USER_AGENT": "Mozilla/5.0 (X11; Linux x86_64)
AppleWebKit/537.36 (KHTML, like Gecko) Ubuntu Chromium/58.0.3029.110
HeadlessChrome/58.0.3029.110 Safari/537.36"

```
Network.setUserAgentOverride({'userAgent': 'Mozilla/5.0 ... Chrome ...'});
```



Simulating real user

```
"HTTP_X_DEVTOOLS_REQUEST_ID": "15158.2",  
"HTTP_USER_AGENT": "Mozilla/5.0 (X11; Linux x86_64)  
AppleWebKit/537.36 (KHTML, like Gecko) Ubuntu Chromium/58.0.3029.110  
Chrome/58.0.3029.110 Safari/537.36",
```



```
linux-shell:~$ curl 127.0.0.1:9222/json
```

```
[[
```

```
....
```

```
  "websocketDebuggerUrl": "ws://127.0.0.1:9222/devtools/...."
```

```
]]
```

From PHP - WebSockets

```
$conn->send(json_encode(['id' => 1, 'method' => 'Network.enable']));  
$conn->send(json_encode(['id' => 2, 'method' => 'Page.navigate', 'params'  
=> ['url' => 'http://vilniusphp.lt']]));
```

Great potential

- Automate creation of Facebook events
- Diagram on e-mail: Many JS libraries, need PNG
- Embed Youtube audio/video: Open new tab for Dashboard
- Screencasting – Visual manipulation for AdBock
- ...

PHP

File_get_contents for
simple status page

CHROME

Great potential, but needs
hacking

DOCKER

Edge cases for UI testing

PHP

File_get_contents for
simple status page

CHROME

Great potential, but needs
hacking

DOCKER

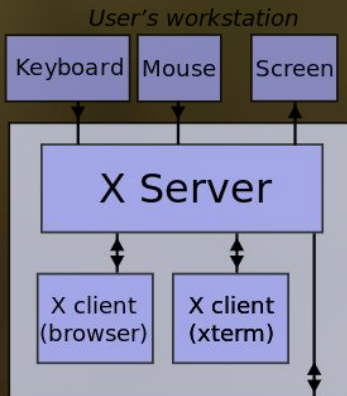
Edge cases for UI testing

```
> docker run -it ubuntu bash
root@660ddd5cc806:/# apt-get update
root@660ddd5cc806:/# apt-get install xvfb
root@660ddd5cc806:/# Xvfb :1 -screen 0 1024x768x16 &> xvfb.log &
```

Not to forget:

```
export DISPLAY=:1.0
```

xvfb to simulate monitor



And then usual workflow:
Firefox + Selenium...

VNC client

Also known as Remote
desktop viewer

E.g. Vinagre

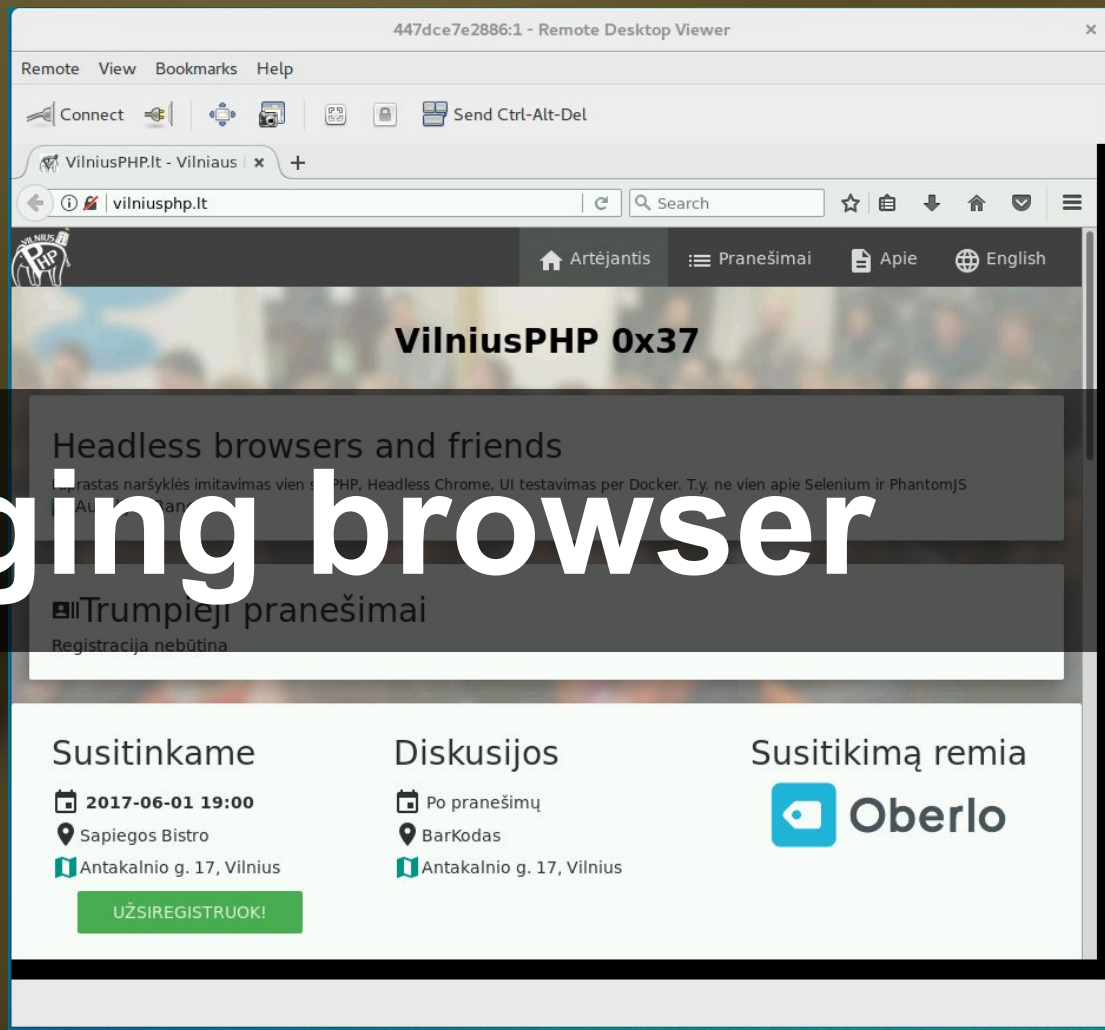
Debugging browser

ports:

- 127.0.0.1:5900:5900

x11vnc -display :1

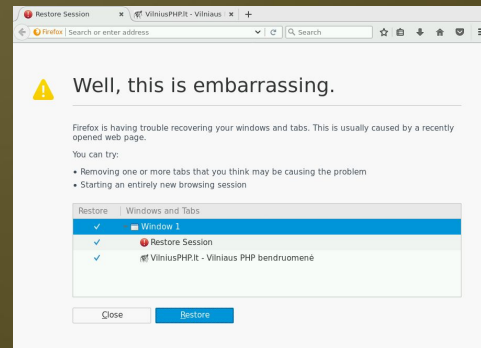
Docker + VNC



xdotool mousemove 30 20 click 1

xdotool mousemove 300 20 click 1

xdotool mousemove 300 50 click 1

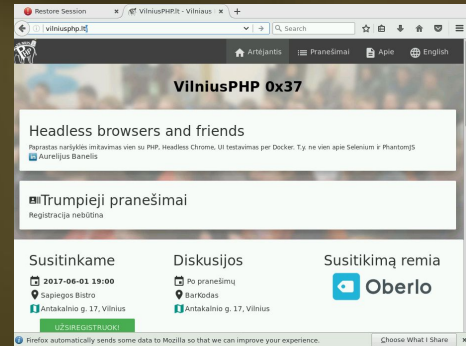


Advanced user input

xdotool key --repeat=20 BackSpace

xdotool key v i l n i u s p h p period l t

xdotool key Return



Benefit of docker

Isolated environment: run it in parallel

```
docker run -e "testCase=a" myDockerContainer & \  
docker run -e "testCase=b" myDockerContainer & \  
docker run -e "testCase=c" myDockerContainer & \  
docker run -e "testCase=d" myDockerContainer & \  
docker run -e "testCase=e" myDockerContainer & \  
docker run -e "testCase=f" myDockerContainer & \  
docker run -e "testCase=g" myDockerContainer & \  
...
```


PHP

File_get_contents for
simple status page

CHROME

Great potential, but needs
hacking

DOCKER

Selenium + xvfb +
xdotool + x11vnc

**Headless browsers
are redefining,
what is and what can
frontend / backend**

**Do not
kill content supplier**

Crawl with morality

PHP

File_get_contents for
simple status page

CHROME

Questions?

Great potential, but needs
hacking

DOCKER

Selenium + xvfb

References

- <https://www.000webhost.com>
- <https://api.slack.com/incoming-webhooks>
- <https://api.slack.com/docs/message-formatting>
- <https://aws.amazon.com/message/41926/>
- <https://www.chromestatus.com/features/5678767817097216>
- <https://github.com/cyrus-and/chrome-remote-interface>
- <https://chromedevtools.github.io/devtools-protocol/tot/Network/#method-setUserAgentOverride>
- <https://www.igvita.com/2012/04/09/driving-google-chrome-via-websocket-api/>
- <https://umaar.com/dev-tips/74-screencasting/>
- <http://interestingengineering.com/programmer-automates-job-6-years-boss-fires-finds/>
- <https://medium.com/@marco.luethy/running-headless-chrome-on-aws-lambda-fa82ad33a9eb>
- [https://stackoverflow.com/questions/40208051/selenium-using-python-geckodriver-executable-need s-to-be-in-path](https://stackoverflow.com/questions/40208051/selenium-using-python-geckodriver-executable-need-s-to-be-in-path)
- <https://unix.stackexchange.com/questions/259294/use-xvfb-to-automate-x-program>
- <http://cloak-and-dagger.org/>
- https://github.com/vi/virtual_touchscreen