

Headless browsers and friends

Aurelijus Banelis



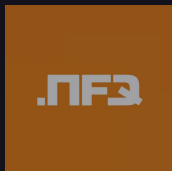
2017-07-04
PHP Night



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

SPACE (p1 of 9)

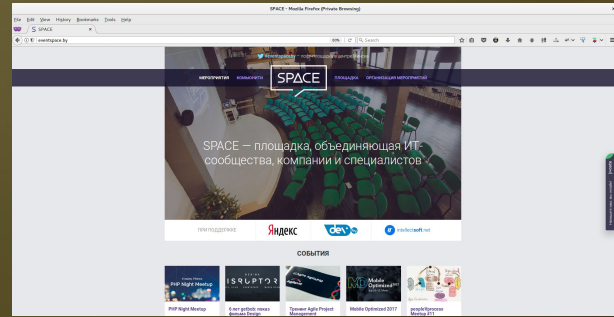
```
#eventspaceby — loft-площадка в центре Минска
(BUTTON) Toggle navigation SPACE
  * Мероприятия(current)
  * Комьюнити

SPACE
  * Площадка
  * Организация мероприятий
```

SPACE — площадка, объединяющая ИТ-сообщества, компании и специалистов

При поддержке

- * Яндекс
- * dev.by
- * IntellectSoft



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://eventspace.by"  
    );
```


Simple status pages

✓ <https://aurelijus.banelis.lt/pranesimai/> Took: 1.1516 | Size: 39.88
🔥 @aurelijus <http://auginte.com/> Took: 0.20588 | Size: 0
🔥 @aurelijus <https://auginte.com/> Took: 0.42304 | Size: 0
🔥 @aurelijus <http://demo.auginte.com/> Took: 0.20239 | Size: 0
🔥 @aurelijus <https://demo.auginte.com/> Took: 0.42335 | Size: 0
✓ <https://blog.auginte.com/> Took: 1.41967 | Size: 48.55

php-status-bot APP 11:30 PM
Checking status... 2017-06-29T20:30:09+00:00

✓ <http://aurelijus.banelis.lt> Took: 4.9014 | Size: 11
✓ <http://aurelijus.banelis.lt/pranesimai/> Took: 0.94799 | Size: 39.84

Real use case: Automate health checks. Especially from external hosting provider

North America

South

Contact Us

Recent Events



No recent events.

Remaining Services



Amazon API Gateway (Frankfurt)

Service is operating normally



Amazon API Gateway (Ireland)

Service is operating normally



Amazon API Gateway (London)

Service is operating normally



Amazon AppStream 2.0 (Ireland)

Service is operating normally



Amazon Athena (Ireland)

Service is operating normally



Amazon Chime



Amazon Cloud Directory



Amazon Cloud Directory



Amazon CloudFront



Amazon CloudSearch (Frankfurt)

Service is operating normally



Sean Spicer

@TrumpPressRoom

Follow



Amazon S3 has never suffered an outage.
Period!

10:59 AM - 28 Feb 2017



Amazon Web Services

@awscloud

Follow



The dashboard not changing color is related
to S3 issue. See the banner at the top of the
dashboard for updates.

11:17 AM - 28 Feb 2017

To monitor the monitoring



[< 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://eventspace.by/");
```

```
// Expected and unexpected phrases
```

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

```
// Formatting output for Slack
```

```
$emoji = $success ? ':white_check_mark:' : ':volcano:';
```

```
$status = $success ? 'OK' : 'Failed';
```

```
// Use value from Slack -> Custom Integrations -> Incoming WebHooks -> Configuration
```

```
$webHookUrl = 'https://hooks.slack.com/services/AAAAA/BBBBBBB/CCCCCCCCCCCCCCCCCCCC';
```

```
$json = json_encode([
    'text' => "$emoji $status",
    'username' => 'php-status-bot',
]);
```

```
// PHP streams magic to simulate POST request
```

```
$opts = ['http' => [
    'method' => 'POST',
    'header' => "Content-type: application/json",
    'content' => $json
]];

```

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

```
// Sending to Slack WebHooks
```

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

```
// Example output: Website: OK Sent to slack: ok
```

```
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://eventspace.by
```

CLI to enable headless

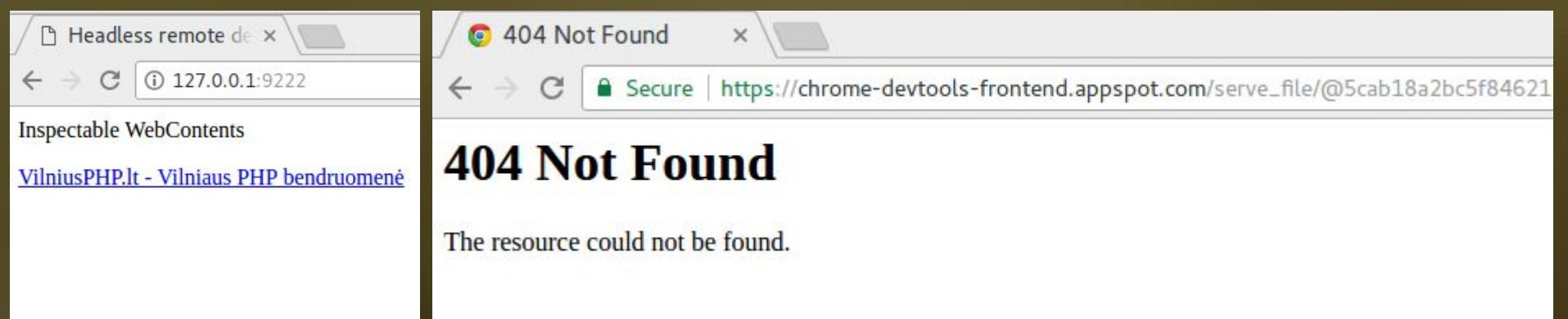
linux-shell:~\$

```
chromium-browser --headless  
--remote-debugging-port=9222  
http://eventspace.by
```

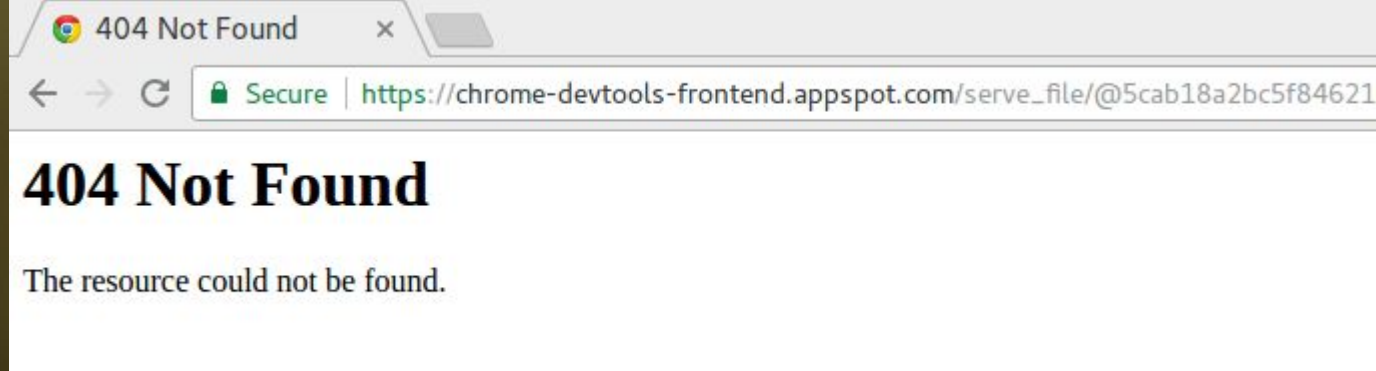
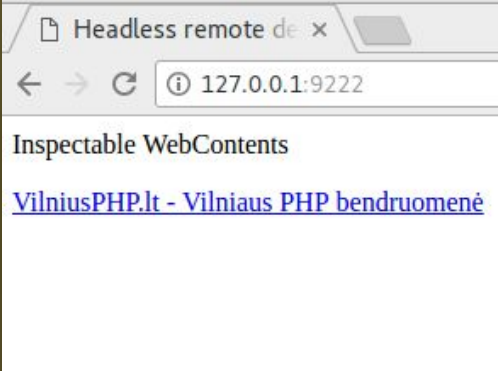
chromium-browser --headless --dump-dom http://eventspace.by

chromium-browser --headless --screenshot http://eventspace.by

google-chrome-beta --headless --disable-gpu --print-to-pdf http://eventspace.by

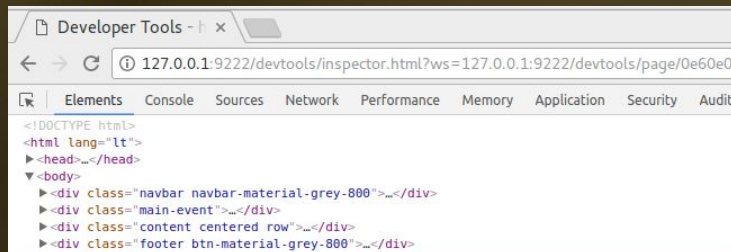


Known issues



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",  
}
```



"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://eventspace.by']]));
```

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

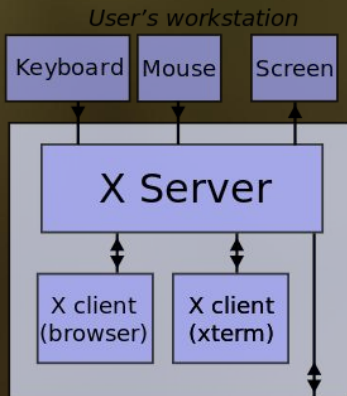
Edge cases for UI testing

```
> 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

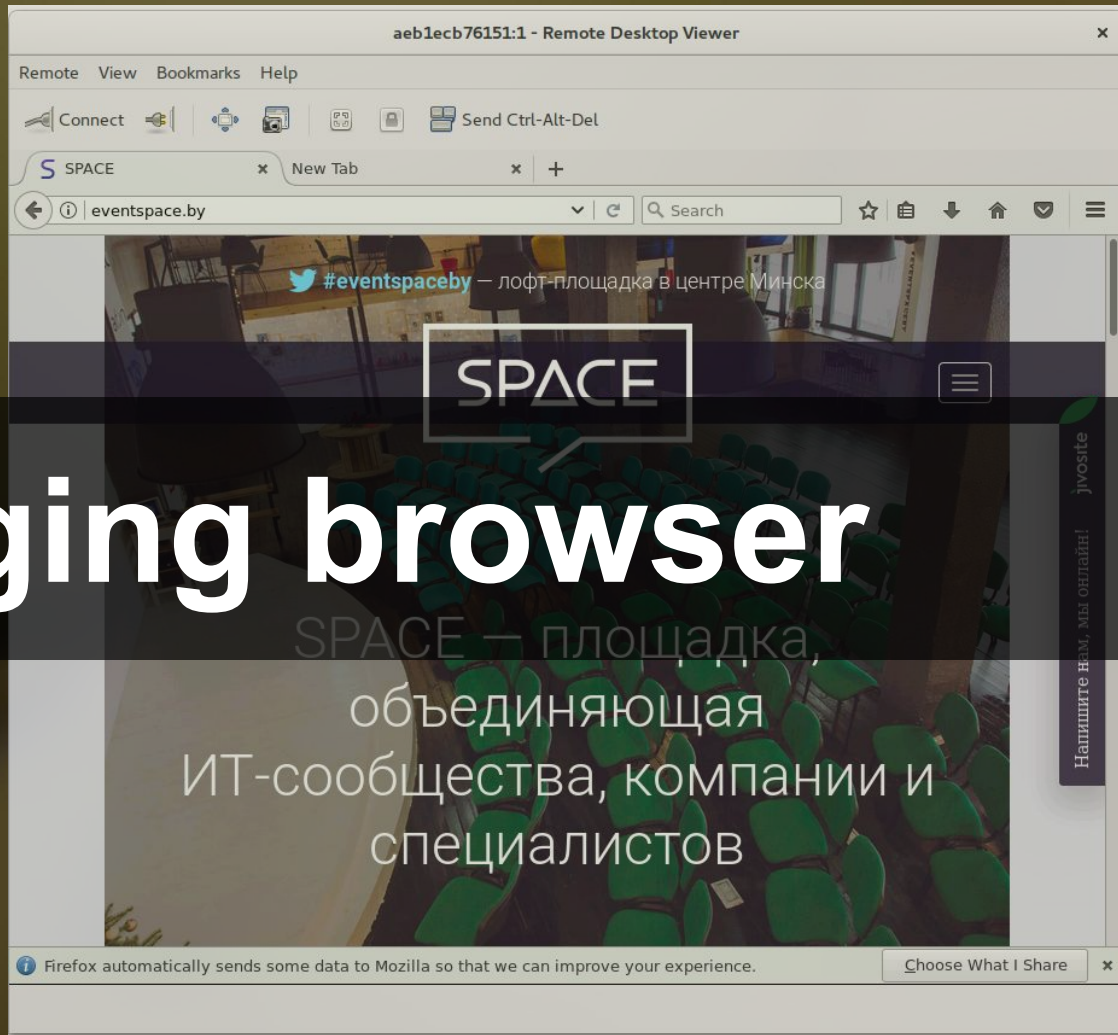
Debugging browser

ports:

- 127.0.0.1:5900:5900

x11vnc -display :1

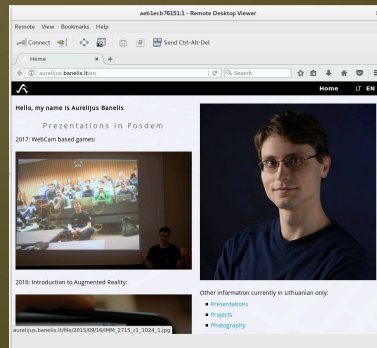
Docker + VNC



xdotool mousemove 240 20 click 1

xdotool mousemove 50 20 click 1

xdotool mousemove 300 50 click 1

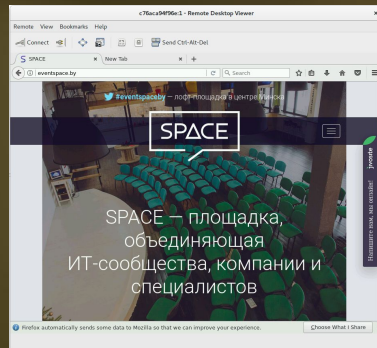


Advanced user input

xdotool key --repeat=50 BackSpace

xdotool key e v e n t s p a c e period b y

xdotool key Return



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**

Automate your tasks

and

Not be victims of other bots

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
- <http://thiemonge.org/getting-started-with-uinput>
- <https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b>