

NIXZ and WEB Scraping

Speaker: Robert Minea



Getting ready for the workshop

www.eonics.nl/download



WINDOWS

Download the Windows script and place it on Desktop

1. Right click > Run as Administrator on the executable and follow the instructions

MacOS

Download the MAC script and place it on Desktop:

1. Press CMD + SPACE
2. Search for "Terminal"
3. Execute 'cd Desktop'
4. Execute 'chmod 0755 macInIt.sh'
5. Execute './macInIt.sh'

Then insert your password and press ENTER when required

Linux

Download the Linux script and place it on Desktop

1. Open a terminal
2. Navigate to Desktop
3. Execute 'chmod 0755 linuxInIt.sh'
4. Execute './linuxInIt.sh'

Then insert your password and press ENTER when required



What is NIXZ?

NIXZ does Robotic Process Automation

- Automates repetitive tasks
- Automatically extracts information from the web
- Connects information between web environment
- Basically keeps you up-to-date automatically ;)



- How?
- Scripts, a lot of scripts, for web scraping



Protractor

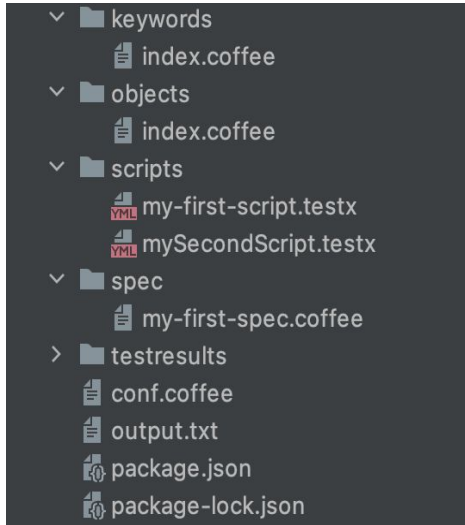
end to end testing for Angular

- Protractor makes life easier for controlling browser behaviour
- Actually great tool for GUI Testing
- We will use it in combination with TestX and CoffeeScript

TestX makes protractor accessible

```
1  - go to:  
2    url: /  
3  - wait:  
4    - googleBigLogo  
5  - refresh page  
6  - wait:  
7    - googleBigLogo  
8  - go back  
9  - go forward  
10 - wait:  
11  - googleBigLogo
```

Project structure that you will work with



- Objects: locators for the elements that you will find on your webPage
- Keywords: contains 'functions' that will be run by calling them in the testx files
- Scripts: contains the actual testx files with the instructions
- Conf.coffee contains the configuration needed by protractor to run

Showcase time

How to extract elements from the page using
xpath and run the script

www.eonics.nl/download

googleScrape.zip file

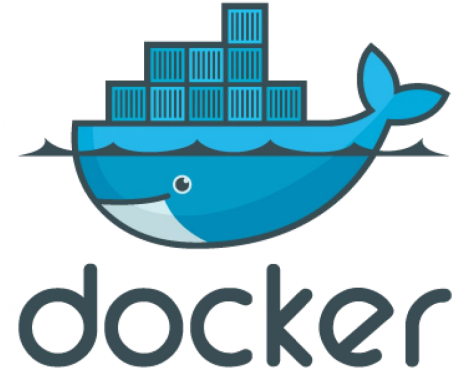


Workshop task

Go to <https://www.pathe.nl/films>, and extract to the output.txt file the title of the movies and their length (e.g. 130 minutes) or some information about them, by clicking on the first 5 movies that appear in the list.

This will run the script:
protractor conf.coffee

What happens in production?



- The scripts get dockerized
- The central dashboard monitors all bots and schedules the scripts
- The servers that have the docker containers with the scripts will run them



Thank you for your
attention!

Time for questions now

