Masaryk University
Faculty of Informatics

Visual testing something catchy

Diploma thesis

Juraj Húska

Brno, 2015

Declaration

Hereby I declare, that this paper is my original authorial work, which I have worked out by my own. All sources, references and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Juraj Húska

Advisor: Mgr. Marek Grác, Ph.D.

Acknowledgement

Some people helped me a lot and some not at all. Nevertheless, I would like to thank all.

Abstract

This thesis is very important!

Keywords

key word1, and so on

Table of contents

1 Introduction

There is a big demand for this thesis. Need and cost of manual testing, space for improvement.

2 Visual testing of software

Testing of software in general is any activity aimed at evaluating an attribute or capability of a program and determining that it meets its required results [1]. It can be done either manually by actual using of an application, or automatically by executing testing scripts.

If the application under test has also a graphical user interface (GUI), then one has to verify whether it is not broken. Visual testing of an application is an effort to find out its non-functional errors, which expose themselves by changing a graphical state of the application under test.

Typical example can be a web application, which GUI is programmed usually with combination of HyperText Markup Language (HTML) and Cascading Style Sheets (CSS). HTML is often used to define a content of the web application (such as page contains table, pictures, etc.), while CSS defines a structure and appearance of the web application (such as color of the font, absolute positioning of web page elements, and so on).

The resulting web application is a set of rules (CSS and HTML) applied to a static content (e.g. pictures, videos, text). The combination of rules is crucial, and a minor change can completely change the visual state of the web application. Such changes are very difficult, sometimes even not possible to find out by functional tests of the application. It is because functional tests verify a desired functionality of the web application, and do not consider web page characteristics such as red color of heading, space between two paragraphs, and similar.

That is why a visual testing has to take a place. Again, it is done either manually, when a tester by working with an application, is going through all of its use cases, and verifies, that the application has not broken visually. Or automatically, by executing scripts which assert a visual state of an application.

In this thesis we are going to focus on the visual testing of web applications only. As we mentioned above, the way how web page looks like is mainly determined by CSS script. There are two ways of automated testing used:

1. asserting the CSS script
2. or comparing screen captures (also known as screenshots) of new and older versions of the application.

2.1 Visual testing in release testing process

Nowadays software is often released for a general availability in repetitive cycles, which are defined according to a particular software development process. Such as Waterfall [2], or Scrum [3].

Testing of software has an immense role in this release process. While automated tests are often executed continuously, as they are quicker to run than manual tests, which are carried out at a specific stage of the release process.

For example in RichFaces[1] Quality Engineering team[2] visual testing was done manually, before releasing the particular version of RichFaces library to a community. In practice it involves building all example applications with new RichFaces libraries, and to go through its use cases with a particular set of web browsers.

To be more specific, consider please a web page with a chart elements showing a sector composition of gross domestic product in the USA (as figure 2.1 demonstrates). To verify its visual state is not broken, would involve e.g.:

1. Checking the size, overflowing and transparency of all elements in charts.
2. Checking colors, margins between bars.
3. Putting a mouse cursor over a specific places in the chart, and verifying whether a popup with more detailed info is rendered in a correct place.
4. Repeat this for all major browsers[3], and with all supported application containers[4].

2.2 Need for automation

The chapter 2.1 tried to outline how tedious and error prone might manual visual testing be. From our experience in the RichFaces QE team, any activity which needs to be repeated, and does not challenge tester's intellect enough, become a mundane activity. The more one repeats the mundane activity, the more likely an mistake is introduced: one forgets to try some use cases of an application, overlooks some minor errors, etc.

Automated visual testing addresses this shortcomings, as it would unburden human resources from mundane activities such as manual testing, and would allow spending their time on intellectually more demanding problems. However, it introduces another kind of challenges, and needs to be implemented wisely. Following are minimal requirements for a successful deployment of an automated visual testing.

2.3 Requirements for automation

An overall cost of the automation has to be taken into consideration. It is necessary to take into account higher initial cost of automation, and consequences it brings: such as increased time to process relatively huge results of testing, cost of test suite maintenance.

1. RichFaces is a component based library for Java Server Faces, owned and developed by Red Hat
2. Quality Engineering team is among the other things responsible for assuring a quality of a product
3. Major browsers in the time of writing of this thesis are according to the [4]: Google Chrome, Mozilla Firefox, Internet Explorer, Safari, Opera
4. Application containers are special programs dedicated to provide a runtime environment for complex enterprise web applications, e.g. JBoss AS, Wildfly, Apache Tomcat

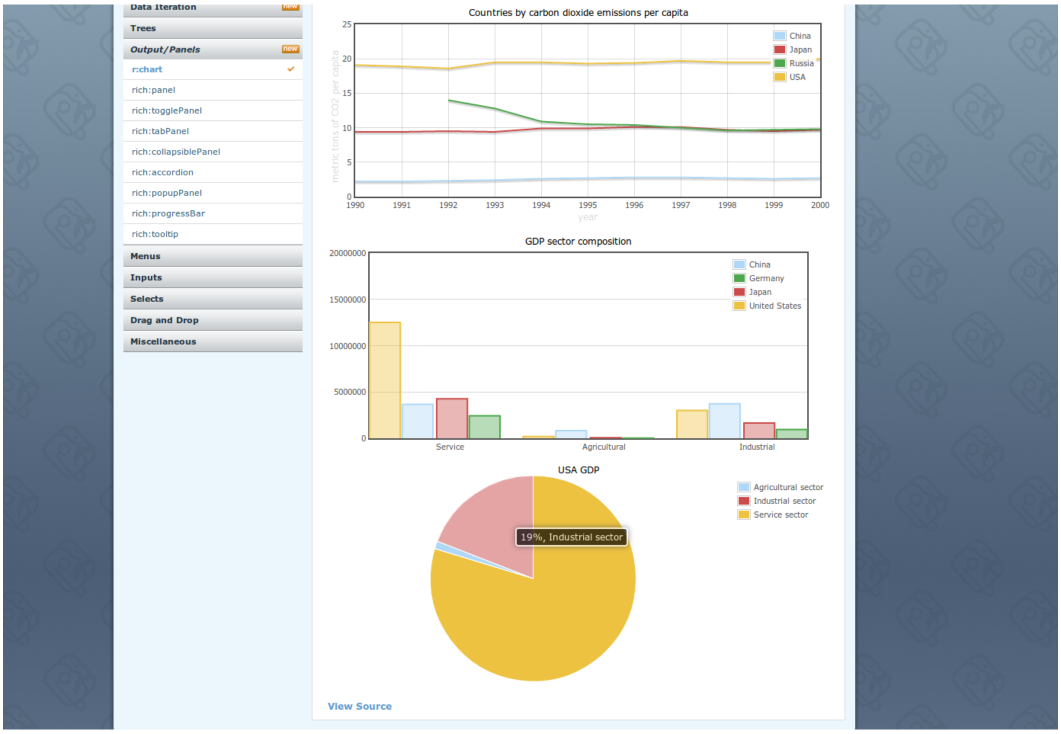

Figure 2.1: RichFaces chart component shown in Showcase application

Therefore, to foster an effectiveness in quality assurance teams, while keeping the cost of automation reasonable low, automated visual testing would require:

1. A low cost of a test suite maintenance;
2. a low percentage of false negative or false positive tests results;
3. a reasonable time to execute the test suite;
4. a concise yet useful test suite output;
5. a support for Continuous Integration systems[5].

2.3.1 Low cost of test suite maintenance

A test suite needs to reflect a development of an application under test. Therefore, with each change in the application, it is usual that the test suite has to be changed as well. Making a

5. Continuous Integration (CI) system is software to facilitate a practice of CI, which in short is about merging all developer copies with a shared mainline several times a day [5].

change in the test suite can often introduce a bug, and cause false negative or false positive tests results.

To keep this cost as low as possible, the test suite script has to be readable and meaningful, so when the change is about to be introduced, it is clear where and how it should be done.

A test framework in which the test suite is developed should provide high enough abstraction. That would enable better re-usability for various parts of the test suite, while lowering the overall cost of maintenance.

Specifically for visual testing, when done by comparing screen captures, it is very important how well a repository of screen captures is maintainable. Secondly, how reference (those correct ones, other screen captures will be compared with) screen captures are made.

2.3.2 Low percentage of false negative or false positive results

False negative test results incorrectly indicate a bug in an application under test, while it is a bug in the test suite itself. They are unwanted phenomenon as they take time to process and assess correctly.

False positive tests results hide actual bugs in an application. They provide an incorrect feedback by showing the tests as passing, even when there is a bug in the application.

Specifically for visual testing, when it is done by comparison of screen captures, it is very easily to be broken by small changes on a page. For example if the page outputs a current date, then it would break with every different date. There has to exist techniques, which would prevent these situations.

2.3.3 Reasonable time to execute a test suite

Reasonable time is quite subjective matter, but in general, it depends on how many times e.g. per day one needs to run whole test suite. Nowadays trend is a Continuous Integration, when a developer or a tester commits changes of an application several times per day to a shared source code mainline. Each such commit should trigger the test suite, which verifies whether the change did not introduced an error to the application.

According to creators of Continuous Integration practice, the whole point of CI is to provide a rapid feedback. A reasonable time for them is 10 minutes. If the build takes more time, every minute less is a huge improvement (considering a developer/tester runs test suite several times a day).

2.3.4 Concise yet useful test suite output

One of drawbacks of automated testing is its ability to produce huge amount of logs, test results etc. The output therefore needs to be as concise as possible, while still providing an useful information. A tester needs to be able to quickly recognize where the issue might be.

The best situation would be if the tester does not need to run the test again in order to spot the issue. The output should give him a clear message where the issue is.

For visual testing specifically, this can be achieved by providing a tester with screen captures together with difference of old and new version.

2.3.5 Support for Continuous Integration systems

This is quite easily to be achieved, but still, a developer of a tool for visual testing should has this in mind beforehand. Todays CI systems support variety of build systems, for various platforms, and languages. For example Jenkins supports build systems like Maven or Gradle, but it can run also shell scripts.

3 Analysis of existing solutions

How the process of testing with these tools looks like, its advantages and disadvantages.

3.1 Mogo

3.2 BBC Wraith

3.3 PhantomCSS

3.4 Facebook Huxley

3.5 Rusheye

3.6 Drawbacks

Conclusion of drawbacks, and why we try to propose another approach

4 New approach

4.1 Hypothesis

Simply: reuse of functional tests of the application for visual testing

4.2 Process

How one would use my tool and where in testing stack such visual testing has its place, written in business process notation

4.3 Analysis of useful tool output

Requirements for useful output of such a tool based on questionnaire for RichFaces team, or maybe I will ask all JBoss employees

5 Implemented tool

An answer to the new process, requirements: CI viable, reusing what can be reused, extensible, cloud ready, multiple users

5.1 Client part

5.1.1 Arquillian

Integration testing, starting containers, event based machine

5.1.2 Arquillian Graphene

Functional testing of Web UI, screenshooter

5.1.3 Rusheye

Screenshots comparison, rewritten to Arquillian core

5.1.4 Graphene visual testing

An adaptor between Rusheye and Arquillian Graphene

5.2 Server part

5.2.1 Web application to view results

Its architecture, reasoning for chosen solutions, screenshots of app, key functionality

5.2.2 Storage of patterns

Description of solution, reasoning

6 Deployment of tool and process

6.1 Deployment on production application

Deployment on stable app

6.2 Deployment on development application

Deployment sooner on application which is in Alpha phase, my hypothesis is that it will not be worth to deploy it on such a app, due to too many changes

6.3 Usage with CI

Jenkins job and its cooperation with the tool, more particullary tool ability to handle multiple jobs, apps, versions, etc.

6.4 Cloud ready

The app can be easily deployed on Openshift

6.5 Results

The percentage of improvement of QA effectiveness

7 Conclusion

What I developed, What I improved, What can be better, Possible ways of extensions: Openshift cartridge

Bibliography