

CREATING A TESTING FRAMEWORK AND WORKFLOW FOR
DEVELOPERS NEW TO WEB APPLICATION ENGINEERING

A Thesis

presented to

the Faculty of California Polytechnic State University

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Taggart Ashby

June 2014

© 2014
Taggart Ashby
ALL RIGHTS RESERVED

COMMITTEE MEMBERSHIP

TITLE: Creating a Testing Framework and Workflow for Developers New to Web Application Engineering

AUTHOR: Taggart Ashby

DATE SUBMITTED: June 2014

COMMITTEE CHAIR: Michael Haungs, Ph.D.
Professor of Computer Science

COMMITTEE MEMBER: Franz Kurfess, Ph.D.
Professor of Computer Science

COMMITTEE MEMBER: David Janzen, Ph.D.
Professor of Computer Science

ABSTRACT

Creating a Testing Framework and Workflow for Developers New to Web Application Engineering

Taggart Ashby

Web applications are quickly replacing standalone applications for everyday tasks. These web applications need to be tested to ensure proper functionality and reliability. There have been substantial efforts to create tools that assist with the testing of web applications, but there is no standard set of tools or a recommended workflow to ensure speed of development and strength of application.

We have used and outlined the merits of a number of existing testing tools and brought together the best among them to create what we believe is a fully-featured, easy to use, testing framework and workflow for web application development.

We then took an existing web application, PolyXpress, and augmented its development process to include our workflow suggestions in order to incorporate testing at all levels. PolyXpress is a web application that “allows you to create location-based stories, build eTours, or create restaurant guides. It is the tool that will bring people to locations in order to entertain, educate, or provide amazing deals.”[10] After incorporating our testing procedures, we immediately detected previously unknown bugs in the software. In addition, there is now a workflow in place for future developers to use which will expedite their testing and development.

TABLE OF CONTENTS

| | |
|--|----|
| List of Tables | ix |
| List of Figures | x |
| 1 Introduction | 1 |
| 1.1 What Already Exists | 1 |
| 1.2 Problems Still Faced | 2 |
| 1.3 Contributions of This Thesis | 4 |
| 2 Background | 5 |
| 2.1 Web Application Definition | 5 |
| 2.2 Web Application Strengths | 5 |
| 2.2.1 Versioning | 6 |
| 2.2.2 Availability | 6 |
| 2.2.3 Multi-device | 6 |
| 2.2.4 Rapid Prototyping | 7 |
| 2.3 Web Features | 7 |
| 2.3.1 HTML5 | 7 |
| 2.3.2 Node.JS | 8 |
| 2.4 Client/Server Architecture | 9 |
| 2.4.1 Client-side | 9 |
| 2.4.2 Server-side | 10 |
| 2.4.3 Client and Server Interactions | 10 |
| 2.5 Defining Testing | 11 |
| 2.5.1 Unit Testing | 12 |
| 2.5.2 Integration Testing | 14 |
| 2.5.3 Continuous Integration | 15 |
| 2.5.4 Acceptance Testing | 15 |
| 2.5.5 Testing Statistics | 16 |

| | | |
|-------|---|----|
| 3 | Testing Frameworks Survey | 17 |
| 3.1 | ToDo Application | 17 |
| 3.1.1 | The Average Web Application | 19 |
| 3.1.2 | Tool Requirements | 20 |
| 3.1.3 | Choosing Tools | 20 |
| 3.2 | Unit Testing Frameworks | 21 |
| 3.2.1 | Jasmine | 21 |
| 3.2.2 | Mocha | 24 |
| | Chai.js | 26 |
| 3.2.3 | QUnit | 27 |
| 3.2.4 | Intern | 29 |
| 3.3 | Browser and UI Testing | 32 |
| 3.3.1 | Selenium | 33 |
| 3.3.2 | Nightwatchjs | 34 |
| 3.3.3 | Headless Browsers | 36 |
| | PhantomJS | 36 |
| | ZombieJS | 37 |
| 3.4 | Continuous Integration Frameworks | 38 |
| 3.4.1 | Jenkins | 38 |
| 3.4.2 | Travis | 40 |
| 3.4.3 | Semaphore | 41 |
| 3.4.4 | MagnumCI | 42 |
| 3.4.5 | Drone.io | 43 |
| 3.4.6 | Manual Continuous Integration | 43 |
| 3.5 | Miscellaneous Frameworks | 44 |
| 3.5.1 | Sauce Labs | 44 |
| 3.5.2 | Istanbul | 45 |
| 3.5.3 | Grunt.js | 45 |
| 3.5.4 | Node Inspector | 46 |
| 3.6 | Framework Decision Matrices | 46 |
| 4 | Framework Selection | 48 |

| | | |
|-------|---|----|
| 4.1 | Chosen Tools | 48 |
| 4.1.1 | Mocha with Chai.js | 48 |
| 4.1.2 | MagnumCI | 49 |
| 4.1.3 | Nightwatch | 50 |
| 4.1.4 | Miscellaneous Tools | 51 |
| 5 | Testing and Development Workflow | 52 |
| 5.1 | Implementation Configuration | 52 |
| 5.2 | Tool Setup | 53 |
| 5.2.1 | Setting up Mocha | 53 |
| 5.2.2 | Setting up MagnumCI | 55 |
| 5.2.3 | Setting up Nightwatch | 57 |
| 5.2.4 | Setting up Istanbul | 58 |
| 5.2.5 | Setting up Grunt | 58 |
| 5.3 | Suggested Workflow | 60 |
| 5.3.1 | Template Files | 63 |
| 6 | Evaluation | 68 |
| 6.1 | PolyXpress (PX) | 68 |
| 6.1.1 | PolyXpress Author | 69 |
| 6.1.2 | Writing Unit Tests for Web Applications | 70 |
| 6.2 | Test Metrics | 71 |
| 7 | Related Work | 74 |
| 7.1 | Testing Frameworks | 74 |
| 7.1.1 | “A Framework for Automated Testing of JavaScript Web Applications” | 74 |
| 7.1.2 | “A Multi-Agent Software Environment for Testing Web-based Applications” | 75 |
| 7.1.3 | “A 2-Layer Model for the White-Box Testing of Web Applications” | 75 |
| 7.1.4 | “Invariant-Based Automatic Testing of AJAX User Interfaces” | 76 |
| 7.1.5 | “JSART: JavaScript Assertion-based Regression Testing” | 76 |
| 7.1.6 | “DOM Transactions for Testing JavaScript” | 77 |
| 7.1.7 | “Testing Web Applications in Practice” | 77 |

| | | |
|--------|---|----|
| 7.1.8 | “Contract-Driven Testing of JavaScript Code” | 78 |
| 7.1.9 | “JAWS: A JavaScript API for the Efficient Testing and Integration of Semantic Web Services” | 78 |
| 7.1.10 | “WebMate: A Tool for Testing Web 2.0 Applications” | 79 |
| 7.1.11 | “Continuous Testing with Ruby, Rails, and JavaScript” | 79 |
| 7.2 | Finding Bugs | 80 |
| 7.2.1 | “Finding Bugs In Dynamic Web Applications” | 80 |
| 7.2.2 | “Testing Web Services: A Survey” | 81 |
| 7.2.3 | “A Framework for Testing RESTful Web Services” | 81 |
| 7.3 | Writing Effective Test Cases | 82 |
| 7.3.1 | “Going Faster: Testing The Web Application” | 82 |
| 8 | Conclusion | 84 |
| 9 | Future Work | 85 |
| | Full PX coverage | 85 |
| | “Real-world” Evaluation | 85 |
| | Headless Browsers | 85 |
| | Sauce Labs | 86 |
| | Bibliography | 87 |
| 10 | Appendices | 92 |
| 10.1 | PolyXpress Tutorial | 92 |
| 10.2 | ToDo App Test Code | 92 |
| 10.3 | PolyXpress Player Screenshots | 93 |
| 10.4 | PolyXpress Author Screenshots | 96 |

LIST OF TABLES

| | | |
|-----|---|----|
| 3.1 | Decision Matrix for Unit Testing Frameworks | 46 |
| 3.2 | Decision Matrix for Continuous Integration Frameworks | 47 |
| 6.1 | Lines of Code | 72 |
| 6.2 | Code Coverage | 72 |
| 6.3 | Separate Tests | 73 |
| 6.4 | Asserts | 73 |
| 6.5 | Mock Objects and Functions | 73 |

LIST OF FIGURES

| | | |
|-------|--|----|
| 2.1 | Illustration relating phase bug found in with effort required to fix. . . | 16 |
| 3.1 | ToDo app home page, simple three-button navigation | 18 |
| 3.2 | ToDo app task page | 18 |
| 3.3 | ToDo app completed tasks page | 19 |
| 3.4 | Each individual spec is further broken down into the assertions within it. | 27 |
| 3.5 | Jenkins main screen | 39 |
| 10.1 | Story selection | 93 |
| 10.2 | Chapter pane with approximate distances to each chapter | 93 |
| 10.3 | Each page and element of PolyXpress has instructions, here they are for Chapters | 94 |
| 10.4 | A map of the chapter location in relation to the player | 94 |
| 10.5 | PolyXpress marketplace, to find other published stories | 95 |
| 10.6 | A map centered around the player | 95 |
| 10.7 | PolyXpress Authoring Tool | 96 |
| 10.8 | Form to create an Event | 96 |
| 10.9 | Bottom of form to create an Event | 97 |
| 10.10 | Form for creating Chapters | 97 |
| 10.11 | Events are selected manually for inclusion in Chapters | 98 |
| 10.12 | Story creation asks if the user wants their story published | 98 |

CHAPTER 1

Introduction

Since the advent of the internet, the web has undergone an impressive evolution from plain-text webpages to the highly stylized and functional pages that we see today. In the last four to six years there has been an explosion of new web technologies that make applications more functional: HTML5, CSS3, WebGL, Touch APIs, Geo-location, and a host of others [34]. In that same time period, the number of global internet users has grown to somewhere around two and a half billion [34]. The combination of these new technologies and the ever-increasing usage of the internet has led to a significant growth in the number and complexity of web applications.

Like any piece of software, these web applications require testing in order to determine their correctness and give their users the best experience possible. Unfortunately, web applications are new and there are no standard practices for testing them. Their dynamic and distributed nature require a different testing toolkit than their predecessors. In addition, web applications are being produced in high volume and with quick development times, leading to a need for rapid testing.

We have found that this is not a solved problem by looking at some of the biggest web applications and websites around. A 2011 study showed that YouTube has at least eight errors, Apple.com has at least seven, Microsoft.com at least four, and the list goes on [29]. That may seem like a small amount for such large sites, but every error is an opportunity for malfunction and customer dissatisfaction.

1.1 What Already Exists

Web developers are not clueless, they understand that there is a need for testing of web applications. Unfortunately, web development is still a relatively new domain, and developers have not come together to work on the problem of testing but rather

have come up with their own proprietary solutions in isolation or merely on a per-project basis. Some of these tools have gained traction and become larger, more general, tools.

Many of these tools are targeted at solving one or two aspects of complete testing but do not cover everything. In addition, some of the tools overlap or require tight coupling that prevents them from being used in combination with one another. Having to wade through a sea of similar tools or try to force tools to work together is time-consuming, which can be the death of a cutting-edge application. The more time an application is off the market, the more time competitors have to develop their own versions. Yet another problem is that deciding between similar tools can be discouraging to the developer that just wants to write their application and test it quickly.

Even the tools that already exist suffer from additional problems regarding their stages of development and amount of support. As will be discussed later, we've found tools that are in their beta stages that show incredible promise as well as mature tools that are widely used but leave much to be desired. These problems affect both industry tools as well as independent developers.

There needs to be a framework in place for testing these multi-layer, multi-faceted web applications that is easy to use, allows for quick development of both features and tests, and is thorough in its coverage. This thesis will explore some of the most popular tools with the end goal of gathering the best tools out there and collecting them in one place. We will attempt to gather tools that work well together and cover all aspects of testing for an average developer, particularly a developer who has experience in some field of software but may be new to web applications. Our hope is that this thesis will serve as a starting point and guide to future web application developers.

1.2 Problems Still Faced

Testing web applications is difficult. The web offers an impressive set of features, but with those features come some draw-backs. Two of the most frustrating drawbacks for developers are the number of platforms and the web's asynchronous nature.

In the very beginning, web applications were easy to write and test because the only code was plain HTML with a common set of HTML tags. It did not matter what browser was used to view the content because the viewer was always a home computer rendering plain HTML. As time progressed, individual browsers began offering additional features and web applications became more rich, but harder to test because the developer would need to run their code on multiple browsers to make sure they all behaved similarly. In today's technological world, it is nearly impossible to run one's web application on all of the possible browser versions, configurations, and platforms that are available. Web applications are on home computers, phones, tablets, MP3 players, refrigerators [12], and a variety of other devices of varying sizes and capabilities. As a developer, one wants to target as many of those platforms as possible with a single code-base to save time and effort. A naive developer might just release an application on all platforms and hope for the best, but releasing on all platforms without testing will likely drive consumers away if the application fails on their given platform.

The addition of JavaScript to web applications led to a golden age of feature-rich applications with exciting user interface (UI) features that had not before been seen. This set of features and UI improvements came with the complication of asynchronous execution. Unlike traditional multi-threaded software, however, the developer is not given access to individual threads, they only have access to the main thread. This sort of implementation is essential for dealing with asynchronous events like button clicks, loading data separately so as not to hold up the web page, and a variety of other niceties, but is a cause of major headaches in testing. Without access to individual threads, the developer has to either write synchronous code that will be slow and not user-friendly, or rely on other mechanisms that allow code to be sequenced. Often it is hard to get code properly sequenced and it requires changing straightforward functions to a hideous amalgamation of callbacks.

There is no overarching solution to these problems, but in order to understand the motivations behind this thesis, one must understand the problems facing web application developers.

1.3 Contributions of This Thesis

The goal of this thesis is to bring together tools that are already available for testing and combine them into a single framework that offers novice web application developers a toolkit for their testing needs. There are plenty of strong tools available and we hope to gather a subset of those tools to cover as much of the testing spectrum as possible.

In addition to creating a single framework, we have set out to weigh the options between the most popular frameworks so that a developer can make informed decisions about the needs of their projects and substitute appropriate technologies. Since every project is unique, a single set of tools cannot possibly cover the gamut. In addition to the single set of recommendations, the most popular tools are discussed so that the developer is free to pick and choose as they see fit.

CHAPTER 2

Background

To make the rest of the paper more understandable and to define the terms used in the remainder of the paper, we will outline the various concepts and technologies surrounding the web. If one is familiar with HTML5, JavaScript, and Testing concepts and statistics, it will likely be safe to skip this background section.

2.1 Web Application Definition

The difference between a website and web application is hard to precisely define, but it comes down to the user's and developer's intentions. A website is an online set of documents whose purpose is mainly informational. A web application is an online set of documents whose purpose is functional. Compare a website like CNN with with a web application like Google Drive and the difference is striking. On the other hand, thinking of a web page like Google it's tougher to make the distinction. Although Google is used as an information gathering tool, it has incredible functionality behind its search box and so we consider it to be an application. If a web page, or pages, is either highly functional or requires significant behind the scenes functionality, the web page is considered an application. If the page, or pages, is informational and does not have significant processing power, it is a website.

2.2 Web Application Strengths

Web applications are great for both developers and consumers. They have a number of characteristics that make them more desirable than shrink-wrap software. These attributes include versioning, availability, platform independence, and the ability to rapidly prototype.

2.2.1 Versioning

Web applications can be updated at any time with instant roll-out and feedback. It requires absolutely no user interaction because the web page always serves the latest resources. In comparison to something like Microsoft Office that requires users to manually update, this feature leads to applications that can address problems quickly, roll-out security updates instantly, and incorporate consumer feedback much more often than any other type of software. For example, Google Docs is becoming popular as an online document editing suite. Updates to Google Docs are rolled out frequently and require little to no input from the user. Another perk is that users are not given a choice. This can lead to occasional customer outrage, but more often than not it is a powerful gain for both feature improvement and application security. There is no need to support legacy systems that may contain serious bugs.

2.2.2 Availability

Web applications are available anywhere there's internet¹. There is rarely need to install anything on a given device. Due to the ever-increasing popularity of mobile devices, most applications have both mobile and full versions of their software and so the application is “with” the user at all times. This instant and continuous availability means that users only have to know how to use a web browser in order to access the application.

2.2.3 Multi-device

Very similar to the availability strength, web applications are device independent. Whether the user is on their phone, tablet, laptop, or PC, the application only requires a web browser with certain functionality. Nearly all devices made today come equipped with browsers that support the necessary functionality. This also means that the developer need not, generally, worry about hardware configurations, conflicting software, and other problems that can arise in desktop applications.

¹A recent push in the web application community is toward offline versions of applications that merely cache changes until the user can connect to the internet once again and all changes are pushed online. This increases the availability of web applications even further.

The developer is not free from all worry as they must still consider differing web browsers. This problem, however, is almost fully remedied by a number of online APIs, such as jQuery [7], that will rework underlying code based on browser versions and types.

2.2.4 Rapid Prototyping

Web applications are fantastic for rapid prototyping because of all of the strengths discussed above. An idea can be prototyped and posted on a website quickly and rapid iteration can occur without the user even necessarily noticing. There are a number of web tools that will record where users are clicking, what features they're using, how long they're spending on a given page or step of a process, and a number of other metrics that can make iterations more directed and effective.

2.3 Web Features

Here we'll discuss some of the prevalent technologies that recently appeared in web applications and development. By no means is this an exhaustive list, just a sampling of the more prominent ones, particularly those that are integral to our evaluation application.

2.3.1 HTML5

The latest revision of HTML is HTML5 which first debuted on Firefox in 2009 [34]. HTML5 has not been officially recommended by the W3C (World Wide Web Consortium) who is the official keeper and recommender of web standards, however, there is a plan for final recommendation this year, 2014, called Plan 2014 [28]. What this means is that in 2014 the W3C will release a stable HTML5 Recommendation which will make HTML5 an official standard and help universalize all the separate implementations.

The HTML5 specification brings with it a number of new features including: audio and video tags, the canvas element, drag-and-drop functionality, web storage, offline page support, and a number of other modern web features. A lot of HTML5's fea-

tures revolve around the inclusion of multimedia in web pages and stronger graphical processing and programming. HTML5 is a collection of new HTML elements and markup as well as libraries written in JavaScript.

As the web is being increasingly used on mobile devices, HTML5 also introduces geolocation functionality. This functionality requires user approval, but once that approval is gained, the web application has access to a rough area in which the user is located. When used on devices that have GPS, such as phones and tablets, this area is more defined and the user's position can be obtained more reliably and accurately. This geolocation functionality has opened the door for brand new types of web applications that can more closely tie in with what the user is seeing or the area that they are in. For example, advertisements can be more targeted or an application can display distances to nearby points of interest.

JavaScript, though not the only scripting language with web use, is the primary language used to complement HTML5 when making web applications. It is an interpreted language that is most often associated with client-side functionality, but in recent times has extended to the server-side. JavaScript is a prototype based object oriented scripting language with dynamic typing and first-class functions. Much of JavaScript's popularity stems from this dynamic nature that allows for quick, and often dirty, implementation of features in an iterative environment.

2.3.2 Node.JS

As a quick aside, it will be useful to have information about Node.JS as it is used in the sample application and the evaluation application. Node.JS, often called simply Node, is a JavaScript API for creating and running servers. "Node's goal is to provide an easy way to build scalable network programs." [36] Node was created in 2009 and is sponsored by Joyent. Node does not use thread-based networking, instead opting for a single-threaded event loop and non-blocking I/O. Node allows developers to create and control web servers without the need for external software, such as Apache, commonly found in other sites.

Node is found in a number of popular sites including: PayPal, The New York Times, Yahoo!, Uber, LinkedIn, Microsoft Azure, and a number of others [36].

One of the biggest draws of Node.js is the ability to write all of one's code, client and server, in a single programming language, JavaScript. This leads to better understanding of the codebase as well as the opportunity for code reuse between client and server.

Node.js does have some drawbacks though. The biggest problem with Node.js is that it does not handle heavy computation well. Any CPU-intensive code will cause an application to hang where other server frameworks can offload tasks much more judiciously. This is due to Node.js' single threaded event loop. Another drawback is that Node.js is still immature in relation to other server frameworks such as Apache. It is not yet regarded as an enterprise framework.

2.4 Client/Server Architecture

By their nature, web applications are distributed; they have a client-side and a server-side.

Here we'll discuss what the client-side is and does, what the server-side is and does, and how they communicate.

2.4.1 Client-side

First, we'll address what a client is. The simple answer is: The User. The User's phone, computer, tablet, or other device communicating with a given website or application is the client. More specifically, the browser or service being used to communicate with the web is the client.

The client-side portion of an application interacts with the browser to request assets from the server and then displays them in a human-readable way. The client-side also deals with the user interface and interactivity of a web page, e.g., the client-side displays a web form, the client fills in the information and clicks "submit" and the client-side packages up the data and sends it to the server.

Client-side programming is done in a language that the browser understands, most frequently JavaScript. Though not technically programming languages, client-side programming is also done through HTML and CSS which control the look and

feel of web pages.

2.4.2 Server-side

A server is the entity, be it machine, person, or cloud service, that stores the information about a given web page or application. The server stores all assets: HTML pages, images, videos, databases, etc, and provides them upon request to the client. The server also handles all of the processing for interactivity, accounts, and general number-crunching.

2.4.3 Client and Server Interactions

The client has limited access to resources on any given web application. This access is restricted by the server. Without such restrictions in place, someone could access user databases, credit card information, order reports, etc. In addition, there would be an overwhelming amount of content for a client to parse through, even if they have no malicious intent.

Clients communicate with the server through a request-response architecture. Clients send a request to the server, often with information attached, such as a form submittal or login, and the server returns a response with either a success message, a failure message, or additional data.

One example of a client-server architecture that is popular for the web is the Representational State Transfer, also known as REST, architecture. If a client-server architecture meets all of the REST criteria, it is said to be RESTful. As this paper is not about Client Server interactions, we will not outline the constraints but instead outline how RESTful APIs look in web applications. The RESTful web APIs use: 1) a base uniform resource identifier (URI), 2) an internet media type, and 3) the standard HTTP methods.

A base URI is simple; in the context of a web application, it is just a URL such as “<http://myrestapi.com/entities>”.

An internet media type is a way to represent and store data that may need to be sent with a request going either direction, from client to server or server to client.

Most web APIs use JavaScript Object Notation, known as JSON, to store data.

The standard HTTP methods are GET, PUT, POST, and DELETE. These methods define what the request is intending to do. For example, a GET request sent to “http://myrestapi.com/entities” would indicate to the server that the client wants to get information regarding all entities. A PUT request would indicate that the client wishes to replace the entire entity collection with another collection that would be sent along with the request. A POST request would ask the server to create a new entity in the collection. Finally, a DELETE request would tell the server to delete the entities. The four methods take on slightly different, but very similar meanings when used with specific element URIs, such as “http://myrestapi.com/entities/entity1”. A GET request asks the server for a displayable version of the entity and the client then displays it. A PUT request asks the server to replace the existing entity or create an entire new entity if one does not exist. A POST request is generally not used as that implies that this single entity is itself a collection. A DELETE request asks the server to remove this entity.

An important part of being a RESTful API is that any change to a web page or element is done through a hypertext link and all actions on an element share a single URI. For example, all updates to “entity1” are done through “http://myrestapi.com/entities/entity1” using the various HTTP methods that can be called using hypertext links on the page. Compare this to a different way to handle that functionality such as “http://myrestapi.com/entities/UpdateEntity” where the HTTP request contains information regarding the entity to update. This method handles all updates, regardless of which entity is changing, through a single URI.

2.5 Defining Testing

IEEE defines software testing as “the dynamic verification of the behavior of a program on a finite set of test cases, suitably selected from the usually infinite executions domain, against the expected behavior.”[20] Put a bit more simply, software testing is the process of making sure the program does what it is supposed to.

There are a number of different types of tests and testing concepts. We will focus on Unit Testing, Integration Testing, Continuous Integration, and Acceptance

Testing.

2.5.1 Unit Testing

Unit Testing is the testing of specific functions and functionality, or “test units.” According to IEEE, test units are “[a] set of one or more computer program modules together with associated control data, [...] usage procedures, and operating procedures that satisfy the following conditions: 1) All modules are from a single computer program. 2) At least one of the new or changed modules in the set has not completed the unit test. 3) The set of modules together with its associated data and procedures are the sole object of a testing process.”[19] Generally this means the testing of individual functions but can extend to a set of functions that produce a single module of functionality. What sets Unit Testing apart from other types of testing is that tests focus on one portion of the code or system rather than interactions between modules or the entire system.

Here’s a brief example:

```
function Word(startingWord) {  
    this.makePlural = function () { ... }  
    this.pastTense = function () { ... }  
    this.partOfSpeech = function () { ... }  
}  
  
function Sentence(listOfWords) {  
    this.listOfWords = listOfWords;  
    this.isGrammaticallySound = function () { ... }  
    this.wordCount = function () {  
        return this.listOfWords.length();  
    }  
    this.makePastTense = function () { ... }  
    this.pluralize = function () { ... }
```

```

}

// Word Tests

function testMakePluralWord() { // Unit Test

    var myWord = new Word("cheese");

    myWord.makePlural();

    assert(myWord == "cheeses");

}

function testPastTense () { ... } // Unit Test

function testPartOfSpeech () { ... } // Unit Test

function testMakePluralCheckPartOfSpeech () { ... } // Not a Unit Test,
    testing multiple units

// Sentence Tests

function testWordCount () { // Unit Test, no Word-specific functionality
    called

    var myWords = [new Word("hello"), new Word("world"), new Word("cheese")];

    var mySentence = new Sentence(myWords);

    assert(mySentence.wordCount == 3);

}

function testPluralize () { ... } // Integration Test, not Unit Test,
    would need to use Word.makePlural

function testIsGrammaticallySound () { ... } // Another Integration Test,
    would need to call Word.partOfSpeech

```

In the above example there are two modules, **Word** and **Sentence**. Unit Testing the **Word** is relatively simple because it is self-contained and only runs functions on the string given to it. Unit Testing the **Sentence** module is tougher because of its reliance on **Words**. The goal of a good set of Unit Tests is to test as much of a module as possible in isolation from all other modules. This includes interactions between modules. In our example, testing the **wordCount** function is a Unit Test because none of the **Word** functionality is actually required. The other methods of **Sentence** pose a problem because of their need to call **Word** functions. There are ways to Unit Test these more complicated modules that have to do with creating “mock objects.” Mock objects are very simple objects whose behaviors and methods are entirely predictable. In our example, we would want to create a **MockWord** which would always return the same word or part of speech regardless of its arguments. In this way, we are not testing the **MockWord** functions in addition to the **Sentence** functions because we know exactly what to expect from the **MockWord** object.

Unit Tests are important because they are done in isolation from one another. The purpose of a Unit Test is to ensure that once individual modules are put together, the developer need only focus on the interaction between them instead of their individual functionality. Without the assurances of prior Unit Tests, integration would be a more stressful and mysterious proposition.

2.5.2 Integration Testing

After writing Unit Tests, a developer writes Integration Tests. Integration Tests are used to determine if modules work together to produce correct results. In our previous example, this would mean testing most of the **Sentence** functions to ensure that they properly use the **Words** that the **Sentence** contains. Integration Tests cover a wider variety of test situations from single methods to entire modules.

Integration Testing is important because it gives the developer a sense of what real features are working, as opposed to which individual modules are producing acceptable results. This distinction is critical to make. Unit Tests ensure that individual modules function. Integration Tests are usually written in a way that is close to how they will actually be used in the final product, giving the developer a sense of which

real features are working and which are not as opposed to just knowing that a small part of the total result is right.

2.5.3 Continuous Integration

The original definition of Continuous Integration (CI) involves the idea of developers committing from their branches to a master branch daily so that changes are continuously integrated in an effort to make sure all branches can, in fact, be integrated. It has since evolved into a more general idea of a tool that runs commands after each commit to a repository. These commands generally encompass running Unit Tests, Integration Tests, user interface tests, and also often publishing and disseminating notifications to interested parties. This is invaluable feedback because the developer can see if they're making progress in the right direction as they write code, as opposed to a situation in which the developer finishes coding and finds out they've made a bad assumption or decision between this commit and the last.

One of the most important aspects of a good piece of CI software is that it is automatic. There should be no need for a developer to do anything besides, perhaps, "turn it on" and start coding. As soon as a tool, CI or otherwise, requires more than one or two interactions to run, it becomes more of a nuisance and less of an asset.

2.5.4 Acceptance Testing

Acceptance Testing is testing to confirm that the product meets a certain specification. Often this is making sure that the software meets an SRS (Software Requirements Specification) or a customer's outlined expectations. Often this testing is performed by the user of the software or customer that ordered the software, rather than the developers.

Some developers and non-technical project members create "User Stories" which are short descriptions of desired functionality. The traditional User Story template is: "As a <role>, I want <goal/desire> so that <benefit>." For example: "As a supervisor, I want to find my employees by their employee ID." These user stories are then broken down by developers to specific steps that can be tested automatically.

2.5.5 Testing Statistics

It is widely accepted that the earlier a developer detects a defect, the easier it is to fix that defect. A famous book, “Code Complete” by Steve McConnell [40], includes the following figure, showing this statistic on a rough graph.

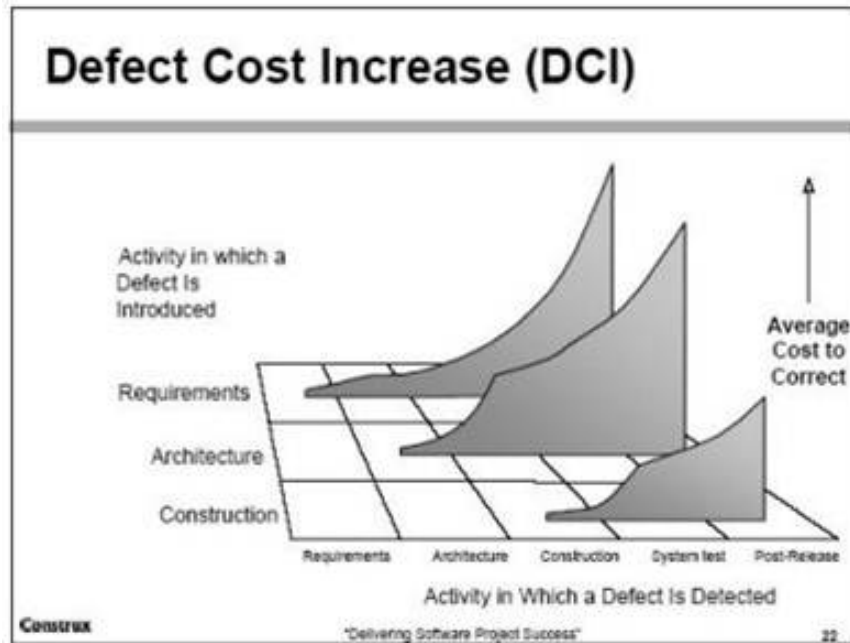


Figure 2.1: Illustration relating phase bug found in with effort required to fix.

What we see here is that the earlier a defect is introduced the more it costs to fix it later. Of particular note is that the costs are roughly exponential, not linear. It’s relatively easy to fix a bug that was introduced in the same stage that it is found, e.g. introducing a coding bug during “Construction” and fixing it while still writing code, prior to the “System Test” phase. The longer a bug lies undetected, the increasingly harder it becomes to fix that bug.

We are setting out to give developers a framework for running tests from the start of development. Our hope is that developers will take on a Test-Driven Development style in which they write tests first and then write the functionality so that bugs are detected soon after introduction. With the combination of Unit Testing and Continuous Integration, our workflow will give quick feedback so that the developer knows exactly what state their project is in at all times.

CHAPTER 3

Testing Frameworks Survey

In the following chapter we will discuss the testing frameworks that we looked at and the application we used to evaluate them. This chapter will conclude with a set of decision matrices outlining the strengths of each individual framework and where they overlap.

3.1 ToDo Application

In order to evaluate the various frameworks that we found online, we needed a small application with some functionality, but without the inherent complexity that was going to come with our final evaluation application. We chose a ToDo application made with Express.js, Node.js, and MongoDB [39]. This application was an open-source application available online. Our choice of this application comes from its limited functionality and the ease of understanding the interactions that take place within the application, i.e. adding tasks to a ToDo list, marking tasks as completed, and viewing completed tasks. Here is a brief preview of the application's UI:

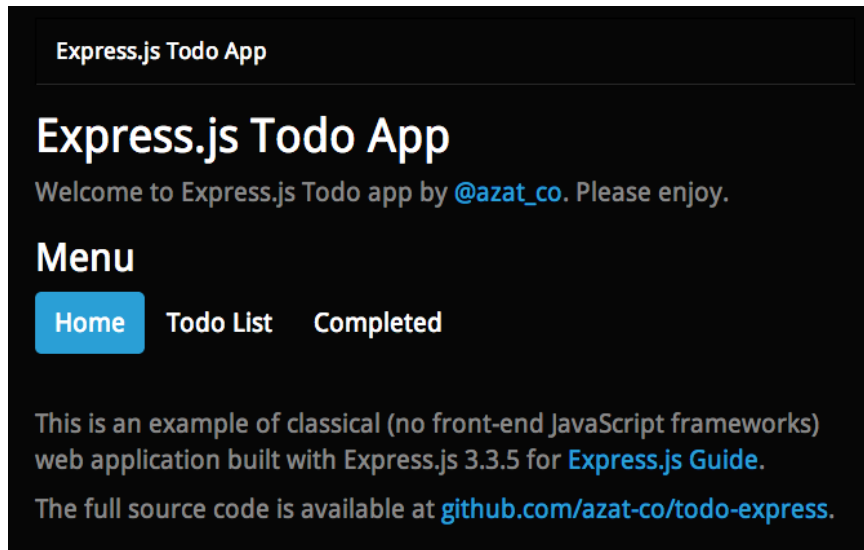


Figure 3.1: ToDo app home page, simple three-button navigation

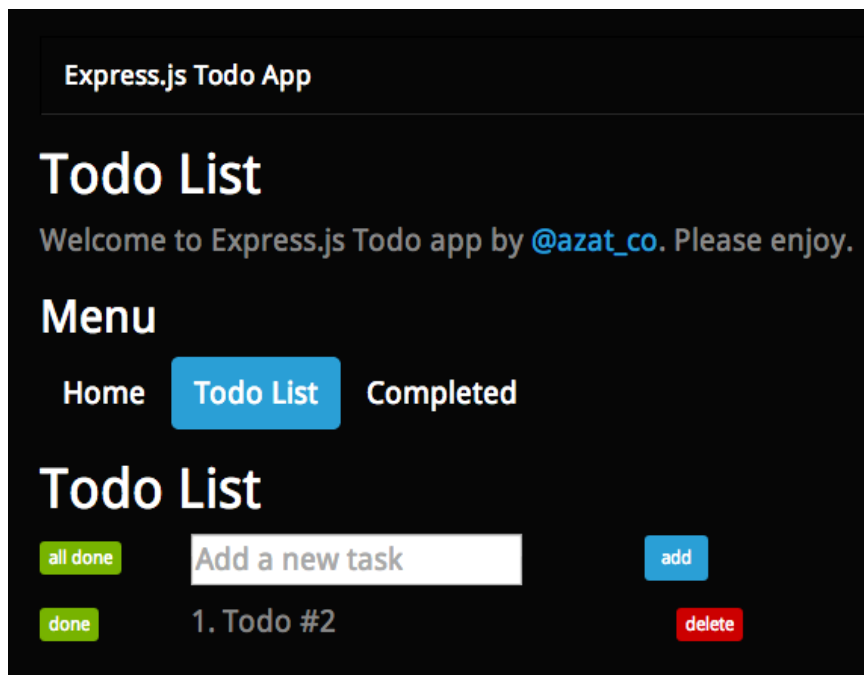


Figure 3.2: ToDo app task page

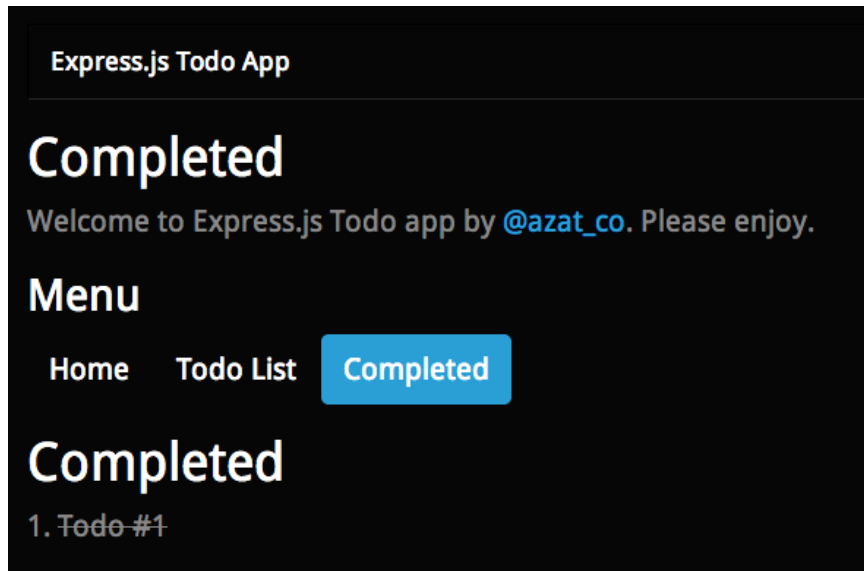


Figure 3.3: ToDo app completed tasks page

In addition to the simplicity and ease of understanding, this application was selected because it included Node.js and MongoDB as we wanted to see if the tools would be able to handle our final evaluation application which also uses those two.

One of the challenges in testing that we quickly came across was the asynchronous nature of the MongoDB function calls. As many web applications feature asynchronous calls, we will be discussing how the following frameworks deal with asynchronous test support.

3.1.1 The Average Web Application

In an effort to create a detailed and in-depth workflow and collection of tools, we have chosen to focus on web applications that use Node.js, JavaScript, and a backing database. Although this will not cover all web application configurations, we wanted to narrow our scope in order to provide a specific, instead of general, set of guidelines and workflow suggestions.

Node.js is in wide use and because it is written in JavaScript we were able to test our client and server code using the same frameworks. Node.js also offers an incredible package manager, NPM, that allows for a smooth and efficient workflow.

JavaScript is the language of HTML5 and Node.js. It is used in over 87% of all

websites [17] and as such allows us to create a workflow to cover a majority of web applications.

Many websites use databases to store user records, items in stock, and a number of other pieces of information. We wanted to make sure that the testing frameworks could support database functionality as it is often pivotal in running web applications.

3.1.2 Tool Requirements

Our requirements for the selection of testing tools was relatively simple. In addition to meeting the requirements of our “average web application”, the only other requirement was that the tool was free. What “free” means is that the developer would have access to at least the minimum functionality needed to use the tool on *one* application that is *open-source*. What we quickly found is that web testing tools support open-source projects for free but are far less willing to provide free functionality to private or closed-source projects.

Occasionally the cost of tools will be discussed but in general we did not review tools that only offered paid options.

3.1.3 Choosing Tools

In order to find suitable tools, we wanted to take several approaches. First, we looked at existing research paper in the domain of web application testing and selected tools that were mentioned that still had active projects. We looked at the related works sections of those papers and chose tools from related papers as well. We then looked at major web applications like jQuery, LinkedIn, Twitter, and Facebook to what tools they were using and found a few tools that way. We looked at a number of projects that were internal and external to Google that were discussed at Google I/O 2013. Finally, we searched online for the latest open-source tools that may be in beta or new to the testing scene.

3.2 Unit Testing Frameworks

The following frameworks are reviewed in chronological order of our evaluation. There was no determined order in which we evaluated them. As such, comparisons will generally only be drawn between a framework and its predecessors in paper order. After all of the frameworks are outlined and evaluated, a section regarding the final decision will be made where comparisons will be drawn between all of them.

Web applications can be broken down into two general parts. There is functionality behind the scenes that drives the web page and there is the user interface of the web page that conveys the information to the user. Unit Testing frameworks are aimed at the functionality that drives the web page, not what the user sees. This means that the Unit Testing frameworks were used to run both Unit Tests and Integration Tests for our ToDo application.

3.2.1 Jasmine

The first of the Unit Testing frameworks that we looked at was Jasmine [5], a Behavior Driven Development (BDD) testing framework. BDD is based on Test Driven Development and adds some simplifications and patterns that attempt to bring both developers and businessmen into the software testing process [24]. The idea behind Behavior Driven Development is that input from non-technical stakeholders as well as technical stakeholders can come together so that everyone understands what the project should do [24].

This relates to Jasmine in the way that the developer writes Jasmine tests. The idea behind Jasmine tests is that they read, as much as possible, like English sentences. As an example, the following is a valid Jasmine test:

```
describe("Arithmetic Test", function () {  
  it("should compute the addition of two numbers", function () {  
    expect(1 + 2).toEqual(3);  
  });  
});
```

```
it("should be able to subtract numbers", function () {  
    var theAnswer = 10;  
    expect(12 - 2).toEqual(theAnswer);  
});  
});
```

The `describe` statement creates a test suite to organize related tests. Each `it` statement is generally one test with some number of `expect` statements. The above code is quite readable even to someone with no coding experience. Jasmine's Expect assertions come with 12 default assertions that can all be negated, meaning 24 assertions in all. The Expect API can be extended with custom assertions if needed.

The default Jasmine did not support Node.js but there was a package available through NPM called `jasmine-node` [31] that gave Jasmine access to the node package and functions.

Asynchronous testing using Jasmine is a bit of a chore because it requires a series of functions that are not straightforward. The asynchronous support comes in the form of a three-set function progression. The first function, `runs(function () { })`, runs the asynchronous code. The second function, `waitsFor(function () { ... })`, waits for a true value from a flag or other system. Finally, the third function, another `runs(...)`, houses the assertions to determine if the code ran properly or not. These functions will be outlined in an example below. This method of asynchronous testing requires a fairly substantial amount of typing as well as complicated function chains when more than one asynchronous call is required.

```
describe("Asynchronous specs", function() {  
    var value, flag;  
  
    it("should support async execution", function() {  
        runs(function() {  
            flag = false;  
            value = 0;
```



```
// Create an asynchronous function to run in half a second
setTimeout(function() {
    flag = true;
}, 500);

});

waitsFor(function() {
    value++;
    return flag;
}, "The Value should be incremented", 750);

runs(function() {
    expect(value).toBeGreaterThan(0);
});

});

});
```

The above test merely sets the `flag` to true after half of a second, thus fulfilling the `waitsFor` function and finally the assertions are run in the final `runs` function.

Jasmine's BDD testing is something that we'd not see before. Having come from a more structured and developer-oriented testing background, it was interesting to see assertions and tests written in this different manner. The idea behind writing tests in this style is that requirements and customer feedback can be more easily translated into tests or even written by customers or other non-technical stakeholders. Jasmine does have asynchronous support but the three-function system is a bit tedious.

3.2.2 Mocha

Mocha [9] is another BDD testing framework that is almost identical to Jasmine. Unlike Jasmine, however, Mocha is not completely self-contained and requires an additional assertion library of one's choosing. We will outline our choice of Chai.js below.

Mocha, by default, also uses the `describe -> it("should...", ...)` pattern to create tests. Mocha offers a variety of other test interfaces for developers to write their test code. A few of the testing interfaces are outlined here:

```
// BDD (default)

describe('BDD Suite', function () {
  it('should be test 1', function () {
    ...
  });
});
```

```
// TDD

suite('TDD Suite', function () {
  test('test 1', function () {
    ...
  });
});
```

```
// Exports

module.exports = {
  'Exports Suite': {
    'test 1': function () {
      ...
    }
  }
}
```

```

    }
};

// QUnit
suite('QUnit Suite');
test('test 1', function () {
    ...
});

// Require, allows the developer to structure and call tests what they want
var testCase = require('mocha').describe

testCase('Require Suite', function(){
    testCase('test 1', function(){
        ...
    });
});

```

The developer doesn't gain anything by using one interface over another except the ability to write tests in a way that makes sense to them. There is no standard for writing assertions, as each assertion library is likely to be slightly different.

Mocha's approach to asynchronous testing is quite a bit more straightforward than Jasmine's. In the `it(...)` anonymous function, there is an additional parameter called `done` which is then called when the asynchronous call has finished. This required us to edit the code we were testing to make all the asynchronous calls accept an additional callback, but it was a very small price to pay for extended testability. An example asynchronous test would look like this:

```

describe("Asynch Mocha Test", function () {
    it("looks like this", function (done) {

```

```

myAsyncCall(param1, param2, function () {
    // This callback expected to run when myAsyncCall finishes
    // Test ...
    // Test ...
    // ...
    done();
});
});
});

```

Overall Mocha offered more flexibility and better asynchronous support at the cost of additional setup. This increased flexibility came in the way of being able to select a testing interface and assertion library so that tests could be written in the way a developer wants. The cost of the flexibility was the need to find an assertion library among the dozen that are out there. We prefer Mocha's method of asynchronous testing because it required considerably less typing on our part and was more easy to follow and read than the Jasmine chain of method calls.

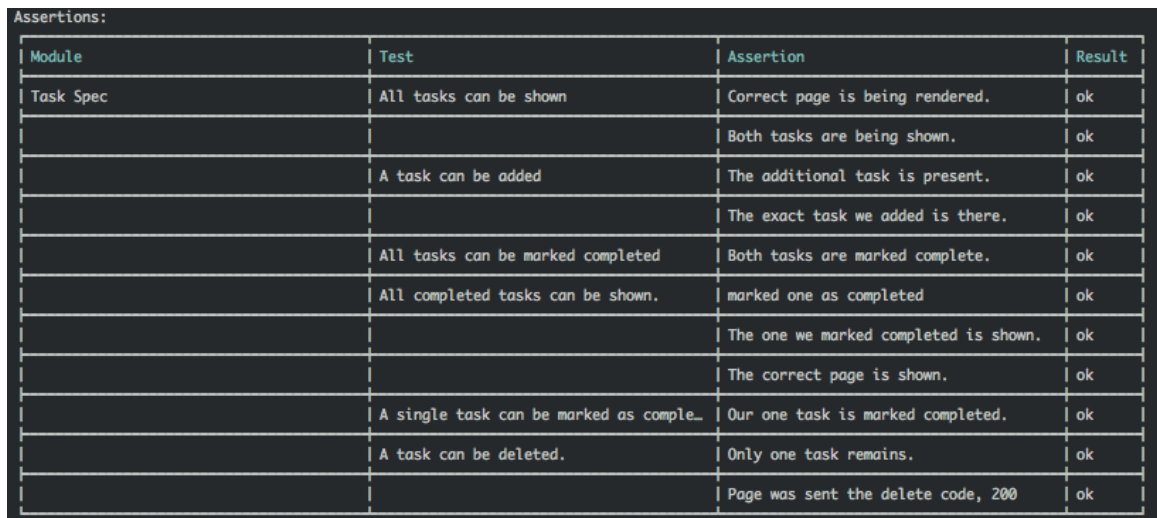
Chai.js

For the assertion library we chose Chai.js [2]. It contains three different assertion APIs: an Expect API, a Should API, and a more general Assert API. The Expect API provides assertions of the form `expect(myVar).to.equal(10)`; or `expect(myArr).to.have.length(10)`. The Should API has assertions of the form `myVar.should.be.a('number')`; or `myArr.should.have.length(10)`. Finally, the Assert API has the more tried and true `assert.equal(myVar, 10)`; and `assert.lengthOf(myArr, 10)`. The various APIs combined with Mocha's flexibility led to an experience that can appeal to a variety of different developers.

Chai.js' Assert library has 50 different assertions to cover a wide range of tests. The Expect and Should APIs have around 33 standard assertions but because of the various ways to chain functions together, there are many more than just 33.

3.2.3 QUnit

QUnit [11] is an Assertion style framework developed by the jQuery Foundation, the makers of jQuery, a popular framework for Document Object Model (DOM) manipulation and access [7]. The DOM is a convention for interacting with objects in HTML, XHTML, and XML documents. QUnit is a bit more developer-centered and not as “feel-good” as the other frameworks but has, so far, been the most easy to use and has the most thorough test reporting of any framework out of the box. Mocha and Jasmine show how many “specs”, the `it` functions, have passed, which is great information, but QUnit takes it a step further and breaks down each spec into the asserts contained within:

A screenshot of a web browser displaying QUnit test results. At the top, it says "Assertions:". Below this is a table with four columns: "Module", "Test", "Assertion", and "Result". The table contains 14 rows of data. The first row shows a test "All tasks can be shown" with two assertions: "Correct page is being rendered." and "Both tasks are being shown.", both with a result of "ok". The second row shows a test "A task can be added" with two assertions: "The additional task is present." and "The exact task we added is there.", both with a result of "ok". The third row shows a test "All tasks can be marked completed" with one assertion: "Both tasks are marked complete.", with a result of "ok". The fourth row shows a test "All completed tasks can be shown." with two assertions: "marked one as completed" and "The one we marked completed is shown.", both with a result of "ok". The fifth row shows a test "A single task can be marked as comple..." with one assertion: "Our one task is marked completed.", with a result of "ok". The sixth row shows a test "A task can be deleted." with one assertion: "Only one task remains.", with a result of "ok". The seventh row shows a test "Page was sent the delete code, 200" with one assertion: "Page was sent the delete code, 200", with a result of "ok".

| Module | Test | Assertion | Result |
|-----------|--|---------------------------------------|--------|
| Task Spec | All tasks can be shown | Correct page is being rendered. | ok |
| | | Both tasks are being shown. | ok |
| | A task can be added | The additional task is present. | ok |
| | | The exact task we added is there. | ok |
| | All tasks can be marked completed | Both tasks are marked complete. | ok |
| | All completed tasks can be shown. | marked one as completed | ok |
| | | The one we marked completed is shown. | ok |
| | | The correct page is shown. | ok |
| | A single task can be marked as comple... | Our one task is marked completed. | ok |
| | A task can be deleted. | Only one task remains. | ok |
| | | Page was sent the delete code, 200 | ok |

Figure 3.4: Each individual spec is further broken down into the assertions within it.

An interesting feature that QUnit has is the ability to specify how many assertions are expected to run in the test. This is useful for callbacks. By setting an expected number of assertions, tests can no longer fail silently just because an assertion was never run. Silent failure can happen if the developer forgets to specify that a test is asynchronous or if there are synchronous callbacks. If a function in the test has a callback that runs the assertions, the test framework may finish executing the test and move on to the next test before ever running the assertions in the callback. If the QUnit Test finishes and the assertion count is not what was specified, it will fail the test and let the developer know fewer assertions were called than expected.

QUnit Tests are not grouped into suites by default and the syntax for grouping them leaves much to be desired. In order to group tests into “modules”, as QUnit calls them, the developer writes code like the following:

```
QUnit.module("Module A"); // All code following is grouped

test("Addition", function () {
    equal(2 + 2, 4, "Two + Two = Four");
    equal(3 + 5, 8, "Three + Five = Eight");
});

test("Subtraction", function () {
    equal(4 - 2, 2, "Four - Two = Two");
    equal(2 - 2, 0, "Two - Two = Zero");
});

// More tests....

QUnit.module("Module B");
// Tests now grouped under Module B
// ... Etc
```

It can be difficult to visually determine where one module starts and ends, resulting in additional time spent finding the right place to add a test.

QUnit’s handling of asynchronous tests is similar to Mocha’s except instead of adding a parameter to the `it` function, the developer calls the `asyncTest` function instead. Finally, the developer calls `start()` when the asynchronous test is finished. Example:

```
asyncTest("Async Test", function () {
    myAsyncCall(param1, param2, function () {
        // This callback expected to run when myAsyncCall finishes
        // Test ...
    });
});
```

```
// Test ...  
  
// ...  
  
start();  
  
});  
  
});
```

QUnit provides another style of writing tests and grouping modules. QUnit's handling of asynchronous tests is nice because they are more explicitly stated for the casual read-through instead of an extra parameter or set of functions within the test itself. QUnit's built-in assertions only offer 10 different types, but their generic nature allows them to be flexible.

3.2.4 Intern

Intern [4] is a hybrid Unit Testing and browser testing framework that aims to be an all-in-one testing framework. Intern has Unit Testing, browser testing, and code coverage information all built into the same tool. It is our intention to build a testing framework with all of these components, so finding a framework with all of them built-in was promising. Intern runs on its own server which gives it access to the internals of the project, but caused problems when attempting to interact with a separate server running the Node.js application. We were unable to get the browser testing portion of Intern to work because of our inability to set up cross-server communication due to inexperience.

Intern uses Asynchronous Module Definition (AMD) form. This is merely a change in how modules are included in a piece of code, instead of separately requiring modules, the entire program is wrapped in a `define([...])` block that passes modules as parameters to an overarching function. That will make more sense after an example:

```
// Intern  
  
define([ // Here's the AMD  
  
    'intern!object',  
  
    'intern/chai!assert',
```

```

    'app/hello'
  ], function (registerSuite, assert, hello) {
    // Test code here
  });

// CommonJS, what Node.js uses
var registerSuite = require('intern/object');
var assert = require('intern/chai/assert');
var hello = require('app/hello');

// Test code here

// Another way to write CommonJS so that it's not global scope, used
// frequently
(function () {
  var registerSuite = require('intern/object');
  var assert = require('intern/chai/assert');
  var hello = require('app/hello');

  // Test code here
})();

```

There are no assertions in the example because the focus is on how the modules are being loaded. Intern offers a variety of testing interfaces and so showing assertions without detailing that will lead to confusion.

Using either style you get access to the modules outlined. AMD form offers a tighter, more straightforward way to look at module loading in addition to being

better supported and having fewer gotchas [18]. Intern offers support for CommonJS modules but strongly recommends switching to AMD styles for your project. We chose to keep our CommonJS format because that is what Node.JS uses, which is used by our evaluation applications.

Intern uses Chai.js for assertion methods but has it built-in as opposed to Mocha in which the developer had to get it themselves. This gives the developer access to Expect, Assert, and Should testing methods. In addition to the various assertion styles, Intern offers three different testing interfaces: TDD, BDD, and Object. These testing interfaces affect how the general layout of test code looks. Here's a very quick example of the different layouts:

```
// TDD

suite('TDD Suite', function () {
  test('test 1', function () {
    ...
  });
})

// BDD

describe('BDD Suite', function () {
  it('should be test 1', function () {
    ...
  });
})

// Object

registerSuite({
  name: 'Object Suite',
  'test 1' : function () {
    ...
  }
})
```

```
},  
'test 2' : ...  
)
```

The ability to mix assertion styles with testing interfaces provides a diverse way of writing tests. Some testing interfaces seem to fit better with certain assertion styles, but there are no restrictions on which can be used with which.

Intern uses the WebDriver API to communicate with Selenium to drive the automated browser testing portion of Intern. Selenium and the WebDriver API will be discussed in more detail in the next section “Browser and UI Testing”.

Intern’s unit-test versus browser test separation is termed, by Intern, nonfunctional versus functional testing. Nonfunctional tests are those that run without a browser while functional tests are WebDriver tests.

Intern includes configurable code coverage metrics on tests by default. Running Unit Tests gives the number of lines covered (run) by the tests. Code coverage is provided by another JavaScript library, Istanbul, created by Krishnan Ananteswaran [22].

Intern offers a fully-featured framework for all testing needs but was difficult to configure for a Node.js server setup. If a developer had a more traditional web application setup with a separate server that served normal HTML pages, Intern would be a go-to framework for all testing needs.

3.3 Browser and UI Testing

Web applications, for the most part, require user input in order to be useful. The way the user interacts is through the outward facing web pages. Testing of a web application would be incomplete without testing the interactions and elements of these web pages. It’s important that the inner workings are running the right code and producing the right answer, but it’s almost just as important that the parameters get passed correctly and from the right places on the web pages. Also, it is important that navigation work as expected or a user will become frustrated and lose interest

in the application.

What's important in browser testing is repeatability and being able to test on multiple browsers. Without the ability to automatically repeat tests, developers are left manually clicking through web pages every time anything changes. In addition, because there are so many different browsers and browser versions in the wild, it's important that the web pages run correctly (and predictably) on as many of them as possible.

3.3.1 Selenium

Selenium [14] is an automated browser driver. It is one of the only automated browser drivers; in all of our searching, we did not find a single other browser automation tool, all other frameworks were built on top of Selenium. A developer can write code that will cause Selenium to drive a browser. On its own Selenium has very minimal testing usage. It lacks a test reporting mechanism and only offers output in two forms: 1) No output at all, a browser window will open, input will be generated, and the window will close, or 2) the browser will exit with a cryptic error printed to the console. It is up to the developer to include print statements as to what is happening at any given time, what test is being run, and what the results are.

Generally, all browser testing tools will be using Selenium behind the scenes and doing the housekeeping automatically. This is the case for the web testing tool that will follow this description.

Selenium has support for running a number of different browsers. Out of the box it comes with support for Firefox but getting Chrome, Safari, and Internet Explorer running takes minimal effort.

Having the ability to automatically run browser tests at any point takes a huge burden off of the web developer and allows for UI testing in addition to behind the scenes functionality testing. Whenever code is changed that may affect the UI flow or functionality, Selenium tests can be run without the need for developers to manually click all of the HTML elements and input information in the relevant places.

Selenium was written with maximum functionality and language flexibility in mind

and therefore can be verbose.

The Selenium WebDriver has APIs for a variety of programming languages including: JavaScript, C, Java, and Python.

An example of one wrapper around Selenium is Nightwatch, which is outlined in the next section with example code included.

3.3.2 Nightwatchjs

Nightwatch [49] is a JavaScript testing library built on top of Selenium in an effort to make browser testing more concise and straightforward. Nightwatch uses Node.js' assertion library paired with Selenium's WebDriver to both run browser automation and also report test results.

Tests are written as a long chain of function calls that produce an easy-to-read series of browser events. For example, the following code segment opens a browser, goes to Google.com, makes some assertions, writes “nightwatch” in the search box, clicks the search button, and then makes a few more assertions:

```
module.exports = {  
  "Demo test Google" : function (client) {  
    client  
      .url("http://www.google.com")  
      .waitForElementVisible("body", 1000)  
      .assert.title("Google")  
      .assert.visible("input[type=text]")  
      .setValue("input[type=text]", "nightwatch")  
      .waitForElementVisible("button[name=btnG]", 1000)  
      .click("button[name=btnG]")  
      .pause(1000)  
      .assert.containsText("#main", "The Night Watch")  
      .end();  
  }  
}
```

```
};
```

As one can see, just reading the code line by line paints a picture of exactly what is happening and what is being checked for.

The same code in normal Selenium is quite a bit more verbose:

```
var assert = require('chai').assert,
    test = require('selenium-webdriver/testing'),
    webdriver = require('selenium-webdriver');

test.describe('Google Search', function() {
  test.it('should work', function() {
    var driver = new webdriver.Builder()
      .withCapabilities(webdriver.Capabilities.chrome())
      .build();

    driver.get('http://www.google.com');
    driver.findElement(webdriver.By.name('q')).sendKeys('nightwatch');
    driver.findElement(webdriver.By.name('btnG')).click();

    driver.sleep(1000);
    driver.findElement(webdriver.By.id('main')).getText().then(function
      (text) {
        assert.isTrue(text.indexOf("The Night Watch") !== -1);
        driver.quit();
      });
  });
});
```

In addition to being more verbose, we've had to add an assertion library and this

code has to be run using Node.js and the Mocha module instead of using the Selenium command. By default, Selenium does not come with its own assertion library. We chose Mocha with Chai.js because we were most familiar with it.

Nightwatch is a Node.js module, and as such can be easily installed and run via NPM and the commandline.

Though a small feature, Nightwatch has the ability to either assert or verify tests. The difference is that assertions will stop execution while verifications will merely log the failure and continue on. Verification can be useful if the developer wants to get an update on how many things are currently failing as opposed to having the entire test stop executing immediately after a single failure.

Nightwatch, as of this writing, has the ability to open Firefox, Chrome, and Internet Explorer. As of February 2014, Chrome is used by 46.69% of internet users, Internet Explorer is used by 24.39% and Firefox is used by 20.78%[1]. Together that accounts for 91.86% of all internet users.

3.3.3 Headless Browsers

In an attempt to speed up UI testing, which is notoriously slow, there have been efforts to create “headless” browsers which behave like normal browsers except that there’s no display element. They emulate the DOM and skip the display phase. We opted to stick with visual browsers due to the added complexity of running headless browsers.

Nonetheless, we still looked at two options that we will outline briefly below for other developers’ reference. The two options outlined below by no means cover all of the headless browser options, they are simply the two that we saw most often discussed by the other testing tools.

PhantomJS

PhantomJS [32] came up repeatedly in our searches for testing frameworks as an option for headless browsing; it is the reason we had the idea to look into headless browsers. PhantomJS is a headless browser written using WebKit, which is what

most web browsers are based on. This means that running PhantomJS is nearly indistinguishable from a normal browser to any testing code.

There are, however, a couple of problems with PhantomJS. It is generally fairly verbose when it comes to writing the scripts that run your commands. It is also not a testing tool, it is a headless WebKit. The distinction here is that PhantomJS does not have any assert functionality built-in and thus requires the developer to find and use their own assertion library.

The difference between PhantomJS and ZombieJS, which I talk about in the next section, is akin to the difference between Selenium and Nightwatch. PhantomJS is heavier duty and requires an additional set of testing tools whereas ZombieJS is lightweight and has assertions built-in.

ZombieJS

ZombieJS [23] is the other headless browser that we saw when searching for browser testing tools. ZombieJS is not a WebKit implementation but rather a simple DOM emulator. This means that certain tests may not be as accurate as they could be with PhantomJS.

Where ZombieJS shines is in its simplicity and brevity. Much like Nightwatch, calls are chained together to create a flow of calls that generally end with assertions.

The following is an excerpt from ZombieJS's examples:

```
// Zombie.js website's example code
var Browser = require("zombie");
var assert = require("assert");

// Load the page from localhost
browser = new Browser()
browser.visit("http://localhost:3000/", function () {

    // Fill email, password and submit form
```

```
browser.  
  
  fill("email", "zombie@underworld.dead").  
  fill("password", "eat-the-living").  
  pressButton("Sign Me Up!", function() {  
  
    // Form submitted, new page loaded.  
  
    assert.ok(browser.success);  
  
    assert.equal(browser.text("title"), "Welcome To Brains Depot");  
  });  
});
```

Developers like ZombieJS syntax so much that there exists a wrapper that uses ZombieJS's API around PhantomJS [53].

ZombieJS offers a lightweight solution to headless browser testing, but does not have the same depth of features as PhantomJS or the same assurances that your tests do exactly what they would do in a browser with a head.

3.4 Continuous Integration Frameworks

Continuous Integration frameworks attempt to make development easier by running specified actions, usually tests, after a change to the project. This change can be simply saving a modified file, or, more usually, an action like pushing a change to a repository.

Continuous Integration frameworks, therefore, need to be flexible in the type of project they support, customizable in their chosen environments to support differing projects, and have a simple way to connect to repositories or file systems.

3.4.1 Jenkins

Jenkins [37] is an offline Continuous Integration tool, the only offline tool that we found with active development and support. It was created by developers of a

previous tool called Hudson when a dispute with Oracle led them to fork the Hudson project and rename it Jenkins. As of February 2014, the Jenkins organization has over 1,100 public repositories and 570 members [6]. Jenkins has, according to their website, two main focuses: (1) building and testing software continuously and (2) monitoring executions of externally-run jobs. The first focus is the general goal for any CI framework. The second focus is of note and will be discussed after the initial evaluation.

Jenkins is a series of web pages that are set up to run on a given server. This is what Jenkins looks like:

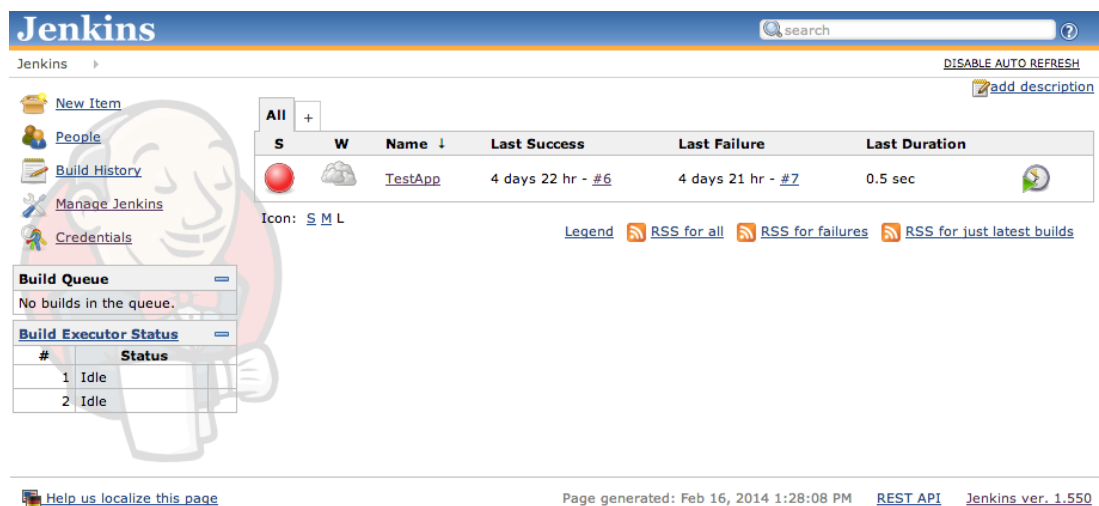


Figure 3.5: Jenkins main screen

Most companies give Jenkins a sub-domain on their web pages.

Jenkins comes without many features but is highly customizable and has a wide array of plugins. The plugins integrate Jenkins with various services, integrate Jenkins with source control systems, and even just change the look of the Jenkins page or how it behaves. As an example, Jenkins can integrate with GitHub to run builds and tests when changes are pushed to a branch.

Because Jenkins comes bare and runs on the system of the developer's choosing, it can run any project that is set up. The downside is that there is no support out-of-the-box unless the system already has the necessary compilers, interpreters, applications, and/or libraries installed.

One of the exceptional strengths of Jenkins is that the developer is only limited by the hardware of the machine running Jenkins. Jenkins can run an “unlimited” number of builds, tests, and processes with the only limiting factor the hardware that it is running on.

The other use for a tool like Jenkins is the ability to start services like cron jobs or similarly long-running tools and status monitors and then collect the output for them and allow the developer to check on them at any time.

Jenkins is an exceptionally strong tool for CI, but requires immense overhead in setting up the environment to run a project in. However, once the server has the necessary applications and libraries, the developer can run builds and tests. Jenkins is 100% free and can connect with public and private repositories alike.

3.4.2 Travis

Travis [25] is an online continuous integration framework with 29 active developers working to improve it and 64 repositories that support it at the time of this writing. Unlike Jenkins, Travis is all online and includes a dashboard from which the developer can monitor builds and processes. Travis has a free version which is fully-featured but requires the project to be open-source and a paid version which gives the developer more processing power and access to private repositories.

Travis comes with internal support for eleven different languages including Node.JS, Ruby, Java, and Python. Travis has thorough documentation and a wealth of built-in services to work with projects. These services include databases, notifications, deployment options, and GUI/Headless browsers. Travis supports all of the most popular databases: MySQL, PostgreSQL, MongoDB, CouchDB, and a myriad of others. It offers a number of notification options ranging from Email to IRC to HipChat (a chat service from Atlassian). Travis also offers deployment options, from Heroku to AWS S3 to RubyGems and also provides the ability to create custom web hooks to deploy to a provider not supported natively.

If at any point the project requires something that Travis does not support natively, Travis gives the developer permissions necessary to install any sort of libraries or services that can be installed with apt-get on an Ubuntu machine.

Travis requires that projects be hosted on GitHub, but by doing so the setup is easy. The developer signs into Travis using GitHub credentials and Travis pulls up all of the developer’s repositories automatically.

Getting a project running with Travis is fairly easy; just create a YAML file in the repository with certain lines pertaining to the language, needed services, commands to run, etc. We were able to set up everything we needed in six simple lines.

In order to get access to private repositories, Travis offers several pricing plans, the cheapest of which is \$129 per month for their “Startup” plan [26]. That’s a bit steep, but it does offer unlimited private repositories, unlimited collaborators, and two concurrent jobs.

From what we gathered, Travis appears to be targeted at companies, not individual developers. Travis requires developers to host their projects on GitHub and only supports one process running at a time. However, Travis is easy to set up, does not require its own server, comes with impressive built-in features, and works wonderfully with open-source repositories on GitHub.

3.4.3 Semaphore

Semaphore [15] is an online CI framework that connects to GitHub to bring in projects. Semaphore is made by RenderedText, a software consulting team. Semaphore connects with public and private repositories but requires a subscription after 30 days no matter which repositories are used.

Semaphore was originally designed for only Ruby applications but in January of 2014 they announced official support for Node.js and JavaScript and as of February 26th they now offer official support for Clojure projects [16]. These types of projects require no setup or files in the repository, Semaphore automatically detects project type and auto-fills build steps. Their “supported stack” page also lists C/C++, Java, Perl, and Python, but those must be configured manually.

Semaphore supports most databases and has a number of testing frameworks built-in. Once again any missing libraries or dependencies can be installed using apt-get. Deployment support is only available for Ruby projects as of this writing.

Semaphore’s pricing scheme is different than Travis’ as it focuses on an individual, not a company. Their lowest cost subscription is one project (public or private) for \$14 per month. They offer more subscriptions with the ability to have more projects and concurrent processes in higher tiers.

Semaphore is still ruby-centric and requires a subscription. As our goal is to find a framework that is free for open-source projects, Semaphore was out automatically. It is included here because it is well-known and fairly established, especially in the Ruby community.

3.4.4 MagnumCI

MagnumCI [8] is another online CI framework. MagnumCI has every feature available for free, however, it’s still in public beta and although prices haven’t been released, the team making MagnumCI has already stated that there will be price tiers. It is not unreasonable, though, to believe that MagnumCI will be free for open-source projects.

MagnumCI offers support for Ruby, PHP, Node.JS, and Go. MagnumCI, much like Semaphore, will attempt to auto-detect commands that it should run based on the project structure and language. All settings are able to be overridden in a file.

MagnumCI has all the popular databases built-in, many of which run automatically on startup. MagnumCI also has two headless browser frameworks built in for running any browser testing that needs to be done. The developer can, again, use apt-get to install missing packages. MagnumCI does not, at current, provide automatic deployment options given a successful build.

MagnumCI offers support for pulling from GitHub, BitBucket, GitLab, and a few others. It does require a bit more setup to get repositories connected, but what it comes down to is a couple lines of copied and pasted text within the repositories’ settings.

Right now MagnumCI offers impressive features, private repositories, and extensive repository sources, all for free. It is hard to say that MagnumCI is the best because those features may end up costing money soon after this paper is completed.

3.4.5 Drone.io

Drone.io [50] is an impressive online CI framework that offers substantial language support, is free for open-source projects, and offers the most reasonable pricing scheme for private repositories.

Drone.io supports eleven languages including: C++, Dart, Go, Haskell, Java, Node.JS, Ruby, and others. Setup requires the developer to specify language, as opposed to the automatic tools provided by the previous two CI frameworks, but will auto-generate build commands.

Drone.io offers support for an impressive number of databases, has a variety of deployment options, and is entirely configurable from the online dashboard. It also offers the most browsers and headless browsers for use with web page testing tools such as Selenium and PhantomJS.

Drone.io is able to connect to GitHub, BitBucket, and Google Code repositories to get projects automatically. Setup is simple, select the source of the repository (GitHub, BitBucket, or Google Code), select the repository from the list it gives once it connects, select the language of the repository, edit the default build script, and the project is ready to go.

Drone.io offers unlimited free open-source repositories. The next step up offers 5 private repositories for \$25 per month and for \$50 per month, unlimited private repositories. They also offer higher tier plans that allow for concurrently running builds.

Drone.io is a strong contender because of its language support, deployment options, and ease of setup. Where it lacks is in its notification options. The only notification option at present is email while other CI frameworks offer a number of other notification options for various messaging services.

3.4.6 Manual Continuous Integration

Most of what these CI frameworks accomplish can be done by hand with enough script writing. Continuous Integration tools are notified of updates by scripts that run after a commit is pushed. The Continuous Integration tools above write these

“post-commit” scripts without the knowledge of the developer. There is, however, no reason that a developer cannot write those scripts themselves.

Manual CI is similar to Jenkins in that its only limitations are the libraries and hardware that the scripts are run on.

There are a few differences between Jenkins and Manual CI, the biggest being that Jenkins has a GUI. This GUI is used to configure the project while manual CI requires script and text editing.

Another big difference between Manual CI and all of the other CI frameworks is that the non-manual frameworks store information regarding the status of builds, the success and failure of builds, which files changed in each commit, and a lot of other information that can be quickly viewed. It is possible to store this information manually, but it would require considerable effort to make it as easy to use as the other CI frameworks.

Finally, Manual CI uses a developer’s system, rendering it unusable to the developer, particularly during the UI testing phase. The developer must accept a productivity hit every time they commit to their repository.

3.5 Miscellaneous Frameworks

In addition to the Unit Testing, Continuous Integration, and Browser Testing frameworks outlined above, we came across a number of helpful frameworks that did not quite fit into any of the above categories.

3.5.1 Sauce Labs

Sauce Labs [13] is an online service that gives you access to a variety of browsers and browser options. They outline the options available to the developer as follows: Selenium, JavaScript, Mobile, Manual. Sauce Labs will start up a Selenium server and run the developer’s Selenium tests, they can run Unit Tests using their JavaScript service, they allow access to a variety of mobile device browsers and resolution, and they offer the ability to manually look at a web application in different browser configurations.

Every CI framework we looked at had built-in configurations for Sauce Labs so that one of the CI steps could be running your tests on Sauce Labs.

Sauce Labs is an helpful tool for making sure a web application behaves as expected on a wide spectrum of devices and configurations and saves the developer the hassle of acquiring multiple devices, operating systems, and browsers.

3.5.2 Istanbul

As already mentioned briefly, Istanbul [22] is a tool used for code coverage metrics. It was built into Intern.io but getting it running with other frameworks is easy as well.

Getting Istanbul to output code coverage is as easy as installing it and then calling the `istanbul` function from the commandline with your test executable. For example: `istanbul cover mocha -- . -R spec` will run the mocha tests in the current directory and output coverage statistics. To break that down a little further:

`istanbul cover` tells istanbul to run coverage on the tests.

`mocha --` specifies that the executable to be run is mocha and that the developer is done giving commandline arguments to istanbul, that's the `--`

Finally, `. -R spec` is just the flags for the mocha command to run the appropriate test suites.

Istanbul offers an easy way to get code-coverage on your tests and is painless to get working right away.

3.5.3 Grunt.js

Grunt.js [3] is a task-runner for JavaScript. What Grunt does is automate routine tasks like linting, testing, copying files, and updating modules. The developer writes Grunt scripts that outline the necessary tasks and then can run them on the commandline with the `grunt` command.

Grunt is widely supported and has all kinds of built-in or installable plugins that make writing common tasks nearly trivial. As an example, a few built-in Grunt tasks that we use include: `jshint` for linting, `uglify` for code minimization, and `watch`

for automatically re-running tests when certain files change. Each of those plugins requires very little configuration before they are ready to be used.

Of particular interest is that last feature, `watch`. We use Grunt's `watch` capability to run tests constantly. This feature gives immediate feedback to the developer when they change existing code or write new code after having written a test for it. As long as all the tests are passing, the developer has a good idea that the changes made are regression-tested as well as freshly tested and are ready for publishing or integration.

3.5.4 Node Inspector

Node Inspector [27] is a GitHub project for debugging Node.js server-side code. The inspector runs in the browser and is started by running `node-debug` instead of `node` when starting the application. Simple as that. Then you can set breakpoints using the normal browser developer tools.

Currently Node Inspector only runs on Chrome and Opera but since those browsers are available on all OSes, it shouldn't be a problem for developers to use.

3.6 Framework Decision Matrices

In an effort to further expedite the selection of tools, the following decision matrices can be quickly referenced by new developers to decide on which tools are right for them.

| Framework | Assertion Styles | Test Interfaces | Reporters | AMD | Active Development* |
|-----------|------------------|-----------------|-----------|-----|---------------------|
| Jasmine | | | | | X |
| Mocha | X** | X | X | | X |
| Qunit | | | | | X |
| Intern | X | X | X | X | X |

Table 3.1: Decision Matrix for Unit Testing Frameworks

* Active Development is defined as any commit within the last month.

** Mocha requires a separate assertion, which leaves the options for assertion styles entirely up to the developer.

| Framework | GitHub | Other Repos | Multi-Language | Active Development** | GUI Configuration |
|-----------|--------|-------------|----------------|----------------------|-------------------|
| Jenkins | X* | X* | X* | X | X |
| Travis | X | | X | | |
| Semaphore | X | | X | X | X |
| MagnumCI | X | X | X | X | X |
| Drone.io | X | X | X | X | X |
| Manual | | X | X | X | |

Table 3.2: Decision Matrix for Continuous Integration Frameworks

* Jenkins has the ability to do all of the things listed with enough time devoted to setup.

** Active Development is defined as any commit within the last month.

CHAPTER 4

Framework Selection

In this chapter, we will outline our final choices for frameworks and why we chose them.

4.1 Chosen Tools

After evaluating a number of different combinations and tools, we came to the following setup:

- Unit Testing: Mocha with Chai.js
- Continuous Integration: MagnumCI
- Browser Testing: Nightwatch
- Misc:
 - Coverage: Istanbul
 - Task Runner: Grunt
 - Package Manager: NPM (sort of a given when working with Node.js)

4.1.1 Mocha with Chai.js

Our choice of using Mocha with Chai.js was made because Mocha offers the flexibility of choosing one's own assertion library and testing interface while still having a default structure that all developers can read and code.

Mocha was chosen over Jasmine because of Mocha's additional flexibility and the way that Mocha handles asynchronous tests.

Mocha was chosen over QUnit because of the ability to structurally portray test suites. Although QUnit produced a more thorough test summary, the ability to scan code more efficiently was of greater importance to us.

The final decision came down to Mocha versus Intern. Mocha was eventually chosen over Intern because of the ability to choose an assertion library and because of the issues we had with Intern. Both frameworks offer impressive flexibility regarding the way in which tests can be structured and written. Mocha ended up allowing for a bit more flexibility and caused no problems for us. Perhaps with more time and patience we would have chosen Intern, but Mocha served our needs well.

4.1.2 MagnumCI

MagnumCI has all of the features that a developer might need for free. We are still confident that the features needed by an open-source developer will continue to be free after MagnumCI ends its open-beta period.

Given enough resources and time, Jenkins will offer more support and functionality than the rest of the CI frameworks. However, there's something to be said for ease of setup and total time to first use of a tool. MagnumCI offered quick initialization through web pages and automatic repository connections that allow users to quickly configure and run the CI framework.

MagnumCI was chosen over Travis because our evaluation application was not on GitHub. Travis is still an incredible CI framework that has some of the best documentation among the CI frameworks. If a developer has their project on GitHub, we would recommend Travis over MagnumCI.

Semaphore’s lack of support for JavaScript and Node.js projects as well as the GitHub requirement kept us from choosing it.

Drone.io, while offering a number of repository location options, did not offer the ability for a repository to be hosted at a custom location.

MagnumCI was chosen over manual CI for much the same reasons as it was chosen over Jenkins. Manual CI requires an extraordinary effort on the part of the developer if they wish to have all of the features and history of the other CI frameworks.

Overall, the CI frameworks were all impressive and usable. The main driving factor behind our decision was that our application used a custom repository and that at the time of this writing our application is closed-source. We have every intention of opening it to the public as soon as we get a version that is more polished.

4.1.3 Nightwatch

The choice for Nightwatch was an easy one. First and foremost there were not many options and the options we did have sorted themselves out.

Nightwatch, being a wrapper around Selenium already, was chosen over Selenium because of its concise way to express tests. We have yet to come across any situation that is not covered with Nightwatch as the developer of Nightwatch has seemed to implement nearly all WebDriver calls, even if some are undocumented.

Nightwatch was chosen over Intern’s UI testing tool because of the issue with Intern running its own server. If Intern did not impose its own server for what it calls “functional” testing, we would probably have gone with Intern for both Unit and UI testing. Intern is a strong choice for an all-in-one testing tool.

4.1.4 Miscellaneous Tools

Generally the miscellaneous tools were chosen because they were already integrated into our project or were the obvious choices for other reasons.

Istanbul was chosen because of the ease of set up in getting it running using our existing toolset and project. There are other code coverage tools but Istanbul required so little setup that we chose to not look further.

Grunt was already a part of the evaluation application prior to the testing effort we set out on. With a bit of looking into Grunt, it was an obvious choice to keep it as our new functionality was easily integrated into the existing Grunt scripts.

NPM is the go-to choice for acquiring modules for Node.js applications.

CHAPTER 5

Testing and Development Workflow

In this section we will mention our development setup, discuss how to set up the chosen tools, and outline our recommended workflow using the tools. It is our hope that after finishing this section, the reader will be able to install and run the chosen tools and have a good grasp of what the next steps are for their particular project.

5.1 Implementation Configuration

The following are the specs of the laptop that we used for the entirety of this research and implementation:

- Model: 15-inch Macbook Pro, Late 2008
- OS: OS X 10.9.2, Mavericks

The point of listing the laptop specs is to show that our research and workflow was completed on an average, everyday laptop and does not require any kind of special hardware or software configurations. In addition, our setup instructions will be OS X based, but nearly all commands will work the same on another UNIX-based system.

To carry out the development we used a combination of the OS X terminal, Sublime Text, and Netbeans. The terminal was used mainly for starting Node services, testing tools, Grunt scripts, and for updating modules. Sublime Text was used to write the Unit Tests and the UI tests. Netbeans was used to quickly navigate through the code as it had already been set up as a series of Netbeans projects.

5.2 Tool Setup

Getting the tools up and running is a fairly simple task. It involves installing modules with NPM, writing or copying a few configuration scripts, and then using the terminal commands for running the tests.

We decided to install the node modules globally, but it's not a requirement. Tests can be executed using locally installed node modules. Instead of using the command by itself, the developer would just need to supply the full path to the executable, which is generally `./node_modules/MODULE_NAME/bin/MODULE_NAME` instead of `MODULE_NAME`. For example, a globally installed Mocha executable can be run with `mocha ...` while a locally installed Mocha module is run with `./node_modules/mocha/bin/mocha ...`.

Following the setup instructions will be a more thorough gruntfile that can serve as a template and starting point for developers. The gruntfile will have tasks that will run the Unit Testing and the UI Testing.

5.2.1 Setting up Mocha

To install Mocha, type into the terminal: `NPM install -g mocha`. The `-g` means to install the module globally. The developer need simply skip the `-g` in all of the following install commands if they want the modules installed locally.

Next, to install Chai, type into the terminal: `NPM install -g chai`.

To get Chai running with Mocha, just `require(...)` the appropriate chai assertion method in a test file and start writing. For example, to get the Assert library:

```
var assert = require('chai').assert;
```

The tests can be put in any folder, but something to keep in mind is that the tests

will reference the source code. The further away from the source code the tests are, the longer the paths to the source code in the test files.

For reference here's what a simple test file might look like:

```
// repo/src/tests/myTestFile.js

var assert = require('chai').assert;

describe("My Test Suite", function () {
  var myModule = require('../my_module.js');

  it("should run my function", function () {
    var result = myModule.myFunction();
    var expected_result = 10;

    assert.equal(result, expected_result);
  });
});
```

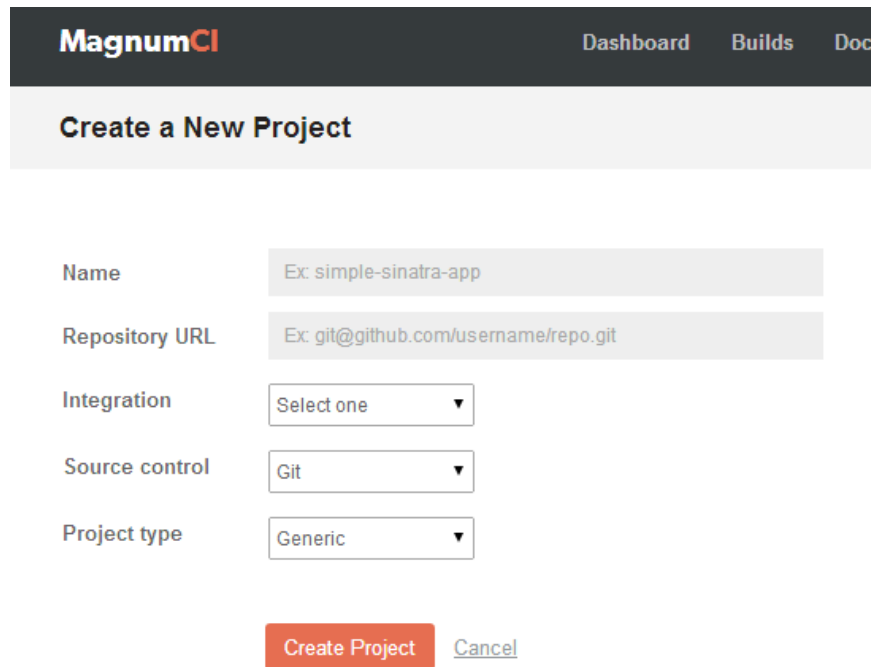
After the tests have been written, the developer can run them using the `mocha` command. There are a number of command-line arguments available to the `mocha` command which are well documented and which we won't talk about here. We will talk about the one that we used consistently, that being `-R`. The `-R` argument allows the developer to specify how the test results will be printed and formatted. Mocha offers seventeen different reporters with more reporters available online and the ability to create custom reporters. We chose the `spec` reporter because it outlined each individual suite and test case. Other reporters generally provide less information such as the `progress` reporter which just shows a progress bar. All together, our mocha test command was `mocha -R spec .`, the final `.` being the directory where the

tests are located.

5.2.2 Setting up MagnumCI

Setting up MagnumCI is all done online. Visit “magnum-ci.com”, create an account, sign in, click on “Add a New Project”.

Here is the new project page:



The screenshot shows the 'Create a New Project' page on the MagnumCI website. At the top, there is a dark navigation bar with the 'MagnumCI' logo on the left and links for 'Dashboard', 'Builds', and 'Doc' on the right. Below the navigation bar is a light gray header section with the title 'Create a New Project'. The main form area contains several input fields: 'Name' with a placeholder 'Ex: simple-sinatra-app', 'Repository URL' with a placeholder 'Ex: git@github.com/username/repo.git', 'Integration' with a dropdown menu showing 'Select one', 'Source control' with a dropdown menu showing 'Git', and 'Project type' with a dropdown menu showing 'Generic'. At the bottom of the form, there are two buttons: a red 'Create Project' button and a blue 'Cancel' button.

Fill out the fields with the appropriate values for a given project and click “Create Project”. Depending on what the setup is for your project, you’ll see slightly varying web pages after you create your project. One thing in common, though, is the button at the bottom of the steps list that says “Customize Build”. After following the instructions, click “Customize Build”:

There are a number of configuration options and panes that can be accessed on the left side. Our focus was making sure the “Build Configuration” tab was properly set up. This involved selecting the correct branch, the correct Node.js runtime version, the correct database from the “Resources” area, and adding the appropriate commands in the build steps. In our case we wanted “mongoDB” and “Xvfb” for UI testing, and then `NPM install -g` in the “Dependencies Installation” to make sure our modules are available when we need them. Finally, we run Nightwatch using `nightwatch -c nightwatch.json` (the nightwatch flags will be outlined in the next section) in the “Test suite commands” box. All of that makes our build configuration page look like this:

Build Configuration










Build on branches

List branches

master

Runtime Version

Resources

Environment variables

One per line. Example: FOO=BAR

Build Steps

Before installation:

Dependencies installation:

npm install -g

Before test execution:

Test suite commands:

nightwatch -c nightwatch.json

After test suite execution:

[Save Changes](#) [Back to Project](#)

After that setup, the project is ready to go. Anytime a commit is pushed from then on, MagnumCI will run and notify the developer of successful and unsuccessful builds.

5.2.3 Setting up Nightwatch

Installing Nightwatch consists of just installing another NPM module. NPM `install -g nightwatch`. In addition to Nightwatch, the developer needs Selenium. The easiest way is to just download the “selenium-server-standalone” JAR straight from the Google repository here:

<http://selenium-release.storage.googleapis.com/index.html>. Our tests ran using “selenium-server-standalone-2.40.0” but any version 2.40+ should work. Once the developer has the JAR and places it in a location they will remember, generally along with other

similar executables in a bin or library directory, Nightwatch is ready to be configured and run. Nightwatch configuration happens in a JSON file that is pointed to when the developer executes Nightwatch.

The configuration file has plenty of options to get Nightwatch to do exactly what the developer wants. Some of the more important ones include the `selenium` object that holds information about where the executable is located and what ports to run on, the `src_folders` attribute to specify a source directory, and the `test_settings` object that contains configurations for default web pages, differing environments, and some desired capabilities.

To run Nightwatch, navigate to the project folder and run `nightwatch -c path_to_nightwatch_c`. This will use the `src_folders` attribute to determine where to look for the tests, start the Selenium server, and run the tests. When debugging certain functions of even test cases, tests can be isolated with the `-t` flag: `nightwatch -c nightwatch.json -t folder1/test1.js`. Nightwatch offers a number of other command-line arguments that are discussed in their developer guide [49].

5.2.4 Setting up Istanbul

As mentioned in the Istanbul section, getting it running is simple. Install Istanbul with NPM: `NPM install -g istanbul` and then run the `istanbul` command along with any others needed. In our case, we ran `istanbul cover mocha -- . -R spec` to run Istanbul on our Mocha Unit Tests. For a more detailed view of possible commands and arguments, the developer can check out the Istanbul page [22].

5.2.5 Setting up Grunt

Getting Grunt set up just the way the developer wants it is the most involved, but there are a myriad of examples and sample scripts to help them through. Installation

is the same as we've seen before, `NPM install -g grunt-cli`. After installation the developer will need to create a "gruntfile" which is just a JavaScript file named "gruntfile.js". Inside the gruntfile are all of the command names and procedures that the developer wants. As the documentation outlines: "A gruntfile is comprised of the following parts: 1) The wrapperfunction 2) Project and task configuration 3) Loading Grunt plugins and tasks 4) Custom tasks"[3].

Here is an example gruntfile:

```
// 1) Wrapper Function
module.exports = function(grunt) {

    // 2) Project configuration.
    grunt.initConfig({
        pkg: grunt.file.readJSON('package.json'),
        uglify: {
            options: {
                banner: '/*! <%= pkg.name %> <%= grunt.template.today("yyyy-mm-dd")
                    %> */\n'
            },
            build: {
                src: 'src/<%= pkg.name %>.js',
                dest: 'build/<%= pkg.name %>.min.js'
            }
        }
    });

    // 3) Load the plugin that provides the "uglify" task.
    grunt.loadNpmTasks('grunt-contrib-uglify');
```

```
// 4) Default task(s).  
grunt.registerTask('default', ['uglify']);  
};
```

In this simple gruntfile, all that is happening is some configuration and the creation of an “uglify” task that will obscure the code when run.

A more complicated gruntfile simply consists of more tasks. Taken one at a time, it’s easy to pick apart a gruntfile. Task names can refer to one or more commands, so the command `grunt testMyStuff` may run one command, or may run five commands. The task names are the ones that are used as the first parameter of the `registerTask` call.

5.3 Suggested Workflow

In addition to just a list of tools and installation instructions, we’d like to propose a way to use those tools that is efficient and nearly autonomous. Our goal is to bring testing tools into use with as little resistance as possible. With a bit of effort at the outset, testing can become a crucial and enjoyable part of development.

What we recommend is a Unit Test suite that runs on save. While this may seem excessive, it provides invaluable, instant feedback for regression assurance and new functionality. The way to set this up with the above tools is have a Grunt task that watches for changes in certain files and runs a script when the files change. Here’s what our Grunt task looks like:

```
module.exports = function(grunt)  
{  
  grunt.initconfig({  
    watch: {
```

```

    test: {
      options: {
        spawn: false
      },
      files: ['**/*.js'],
      tasks: ['mochaTest']
    }
  },
  mochaTest: {
    test: {
      options: {
        reporter: 'spec',
        clearRequireCache: true
      },
      src: ['PolyXpressAuthor/test/unit/functional/*.js']
    }
  }
});

// Specify the test directory for Watch to keep track of
var defaultTestSrc = grunt.config('mochaTest.test.src');
grunt.event.on('watch', function(action, filepath) {
  grunt.config('mochaTest.test.src', defaultTestSrc);
  if (filepath.match('test/functional')) {
    grunt.config('mochaTest.test.src', filepath);
  }
});

```

```
grunt.loadNpmTasks('grunt-contrib-watch');
grunt.loadNpmTasks('grunt-mocha-test');

// Register the mocha test task
grunt.registerTask('test', ['mochaTest', 'watch:test']);
}
```

Our gruntfile is more complex than just that, but those are the parts that make the test-on-save functionality work. As we said earlier, Grunt has a lot of plugins that are available. We use one here to make our lives easier, `grunt-mocha-test`. It is installed, once again, through NPM and gives us access to the now-built-in task `mochaTest`. The run-on-save functionality comes from the `watch` task which has `mochaTest` in its `tasks` array.

Once that setup is done, the developer just runs `grunt --force test` once which will execute the tests and then watch for changes and continue testing on save. The `--force` flag is used so that the Grunt task continues after “failure” which would occur if any tests did not pass.

For running the browser tests, we recommend using the CI tool. Most CI setups have Xvfb (a windowless X11 process) and some even have Firefox pre-installed. Nightwatch requires both to be installed. If Firefox needs to be installed, the developer can use the CI framework’s package manager to install it. For example, if the package manager is `apt-get`: “`apt-get install firefox`”. In the CI setup, after installing all the prerequisites, the developer should start their application. Next, export the `DISPLAY` environment variable, export `DISPLAY=:99` (the numbers need to match between the xvfb call and the export). Then, run `Xvfb :99 -ac -screen 0 1280x1024x24 &` to start the “window” that will be used. Finally, run Nightwatch, `nightwatch -c path_to_nightwatch_config` and enjoy UI testing that doesn’t take

over the workstation.

5.3.1 Template Files

In order to get developers up and running as quickly as possible, we've included a couple template files below.

The first file is the `package.json` file that includes the dependencies for the Unit Testing, UI testing, and grunt portions of the workflow:

```
{  
  "name": "MODULE_NAME", // Edit Me  
  "version": "SOME_NUMBER", // Edit Me  
  "dependencies": {  
    "grunt": "latest",  
    "grunt-contrib-watch": "latest",  
    "grunt-shell": "latest",  
    "grunt-mocha-test": "^0.10.2"  
    "mocha": "^1.18.2",  
    "istanbul": "^0.2.7",  
    "nightwatch": "^0.4.9",  
    "chai": "^1.9.1"  
  }  
}
```

The developer should put this file in the main folder of their project, the folder where they will run commands from. It is not necessary for the folder to be the same folder that the tests are in, just a central location where the developer expects the most work to be done. This file can be copied into other directories if the developer expects to run the testing and Grunt commands from multiple directories.

Next is an example `gruntfile.js` that will get Grunt up and running quickly:

```
module.exports = function(grunt)
{
    // Configure Grunt
    grunt.initConfig({
        pkg: grunt.file.readJSON('package.json'),
        shell: {
            NPM: {
                command: 'NPM update',
                options: {
                    stdout: true
                }
            },
            istanbul: {
                command: 'istanbul cover _mocha -- PATH_TO_TESTS -R spec',
                // Edit Me
                options: {
                    stdout: true
                }
            }
        },
        watch: {
            test: {
                options: {
                    spawn: false // So watch actually works!
                },
                files: ['**/*.js'],
                tasks: ['mochaTest']
            }
        }
    });
}
```

```

    },
    coverage: {
      options: {
        spawn: false // So watch actually works!
      },
      files: ['**/*.js'],
      tasks: ['shell:istanbul']
    }
  },
  mochaTest: {
    test: {
      options: {
        reporter: 'spec',
        clearRequireCache: true
      },
      src: ['ABSOLUTE_PATH_TO_TESTS'] // Edit Me
    }
  }
});

var defaultTestSrc = grunt.config('mochaTest.test.src');
grunt.event.on('watch', function(action, filepath) {
  grunt.config('mochaTest.test.src', defaultTestSrc);
  if (filepath.match('RELATIVE_PATH_TO_TESTS')) { // Edit Me
    grunt.config('mochaTest.test.src', filepath);
  }
});

```

```

// Load libs
grunt.loadNpmTasks('grunt-contrib-watch');
grunt.loadNpmTasks('grunt-mocha-test');
grunt.loadNpmTasks('grunt-shell');

// Register the default tasks
grunt.registerTask('default', ['mochaTest']);

// Register update task
grunt.registerTask('update', ['shell:NPM']);

// Register the mocha test task
grunt.registerTask('test', ['mochaTest', 'watch:test']);

// Register the code coverage task
grunt.registerTask('test_with_coverage', ['shell:istanbul',
    'watch:coverage']);
};

```

The developer will want to put the gruntfile in the same folder as the `package.json` file. This template file comes with a `default` task, an `update` task, a `test` task, and a `test_with_coverage`. The `default` will run the Mocha tests one time, with no additional watch functionality. The `update` task that will update the dependencies. We recommend that the developer run the `update` task at least once a day to keep dependencies and executables up to date. The `test` task will run the Mocha tests with watch functionality so that the tests will run on every save of the source files. The `test_with_coverage` task will run the Mocha tests with *coverage* in addition to the watch functionality.

With these two files, the developer will have a testing workflow set up with minimal effort.

CHAPTER 6

Evaluation

After gathering the tools into a loose toolkit, we wanted to use that toolkit on an existing project of respectable size. We chose an application developed at Cal Poly called PolyXpress.

6.1 PolyXpress (PX)

Our major evaluation application was an in-house application called PolyXpress (PX)[10]. As the About page says, “PolyXpress allows you to create location-based stories, build eTours, or create restaurant guides. It is the tool that will bring people to locations in order to entertain, educate, or provide amazing deals.”[10] Users can play or create “Stories”, each of which consists of “Chapters”. Each chapter consists of various “Events”. Each “Event” generally gives the user some sort of information, in the form of text, video, audio, or a picture. After the player has chosen a story, they must physically reach certain checkpoints to unlock chapters and events. Since the application requires physical movement, it works solely on devices with GPS, generally phones and tablets. Rather than put a series of pictures showing PolyXpress here, we will put them in the Appendices at the end of the paper. To get a better sense of what PolyXpress does and looks like, we recommend looking at the Appendices now.

PolyXpress consists of three smaller application: 1) The Player. 2) The Author. 3) The PointFinder.

The Player is what the user open to actively complete stories. Using the Player, the user travels around and get notification when things are within range and can explore the story further. When the user opens the Player they can choose which story they want to play, including any they've made. The Player also gives them access to the PX Marketplace where all other public stories are available for download.

The Author is a tool to create one's own stories. Our tests revolved around the Author, so we'll be discussing it in depth in the following section.

The PointFinder is a tool for users to select GPS points based on their current location. Setting points when authoring stories can be done by clicking on a map or using points saved in the PointFinder.

6.1.1 PolyXpress Author

Our tests covered the Authoring portion of PolyXpress. PolyXpress is a completely open application where users can publish and share their stories with the world.

PolyXpress is meant for the general public and so there's no convoluted programming involved to create a story, it's all done through the application's Authoring tool. Screen captures of the authoring tool are in the appendix once again. The flow for creating a story is to create the Events first, the Chapters second, and finally tie it all together and create a Story. The flow is this way because the tool offers lists of existing Events in the Chapter creation page and lists of existing Chapters in the Story creation page. Doing the flow out of order or backwards would require a lot of editing of elements already created.

An Event consists of a title, a list of authors, a list of keywords, a version number, whether or not it's a recurring event, a geographical location (either entered manually as latitude and longitude, or using a map and clicking), a range at which this event can be triggered, and a number of assets to show the user when this event is triggered.

A Chapter consists of almost all the same things as an Event with the addition of an overview, an image, and a list of Events associated with that Chapter.

A Story consists of most of the same things as a Chapter except that the list of Events is a list of Chapters and the user chooses whether or not they want to publish this story. If the user chooses to share the story with the world, it is placed in the marketplace, otherwise it is stored within their own list of stories and is inaccessible to anyone else.

In regards to the underlying code, the Author tool is made up of three parts, the Controller, the Model, and the Server. The Model is more-or-less just a list that stores events, chapters, and stories and has retrieval and insertion methods. The Controller orchestrates the interactions between the Server, Model and the web page. The Server stores and retrieves information from the backing database.

6.1.2 Writing Unit Tests for Web Applications

Writing Unit Tests for normal applications can be quite challenging. When there is the added overhead of network communication, it becomes exceptionally difficult to test units in isolation, i.e. without having to worry about bugs and time related to network code. There is also the issue of asynchronous calls that we talked about earlier. Although the testing frameworks offer solutions, the developer still must recognize and properly write tests for asynchronous calls.

The goal of a Unit Test is to test the function, module, or “unit” in isolation. While this sounds simple in theory, the situation is often more complicated when it comes to actual code. The method we used to alleviate this issue was to create mock functions and objects. Mocks are stand-in pieces of code that get called instead of the real implementation. Mocks are highly controlled and often just simply return the same value regardless of parameters or prior calls. Mock functions are just as

their name implies, a function with a known return value regardless of parameters that is named the same things as what the unit is calling. Mock objects are objects with attributes and functions that the unit interacts with instead of the real object it will interact with in production.

Most of our mocks were used in the Controller Unit Tests because the controller continually reaches into the Model and the Server. In order to test the Controller in isolation, we had to mock almost an entire new Model and new Server. This mock kept track of previous data, parameters, and calls in order to test that certain calls were made and made in the right order within the Controller.

All of the Unit Testing code for the PolyXpress Author is available in the appendix. The mocks are at the top of each test file, though the Model does not have any mocks.

6.2 Test Metrics

Below are several tables of metrics we collected after completing our tests. The tables are: Lines of Code, Code Coverage, Separate Tests, Asserts, and Mock Objects and Functions.

The Lines of Code table measures how many lines of test code were written per line of source code. The Code Coverage table shows the coverage statistics for our tests including statement, branch, function, and line coverage. The Separate Tests table shows how many tests were written for each part of the Authoring tool we tested. The Asserts table breaks down the tests even further into how many asserts were written for each part of the Authoring tool. Finally, the Mock Objects and Functions table shows how many mock objects and stubbed functions we had to create to ensure proper Unit Testing of the Controller and Server modules. The Model was self-contained enough that we did not need to create mock objects or functions for it.

| | Source LOC | Test LOC | Test LOC per Source LOC |
|------------|------------|----------|-------------------------|
| Unit Tests | | | |
| Model | 93 | 104 | 1.11 |
| Controller | 862 | 438 | 0.51 |
| Server | 321 | 206 | 0.64 |
| UI Tests | | | |
| Main | N/A* | 16 | N/A* |
| Event | N/A* | 140 | N/A* |
| Chapter | N/A* | 133 | N/A* |
| Story | N/A* | 102 | N/A* |
| Cleanup** | N/A* | 46 | N/A* |

Table 6.1: Lines of Code

* The source for the UI is dynamic based and so no measurement is possible.

** Though not direct testing, the Cleanup code was important to bring the application back to its starting state.

| | Coverage | Elements Covered / Total Elements |
|-----------|----------|-----------------------------------|
| Statement | 98.78% | 565 / 572 |
| Branch | 50%* | 4 / 8 |
| Function | 96% | 120 / 125 |
| Line | 98.77% | 564 / 571 |

Table 6.2: Code Coverage

* Branches not covered were diagnostic in nature.

| | | |
|------------|------------|-------|
| | | Tests |
| Unit Tests | | 50 |
| | Model | 15 |
| | Controller | 23 |
| | Server | 12 |
| UI Tests | | 29 |
| | Main | 1 |
| | Event | 9 |
| | Chapter | 9 |
| | Story | 7 |
| | Cleanup | 3 |

Table 6.3: Separate Tests

| | | |
|------------|------------|---------|
| | | Asserts |
| Unit Tests | | 237 |
| | Model | 33 |
| | Controller | 128 |
| | Server | 76 |
| UI Tests | | 43 |
| | Main | 5 |
| | Event | 13 |
| | Chapter | 13 |
| | Story | 12 |
| | Cleanup | 0 |

Table 6.4: Asserts

| | | |
|----------------------|--------------|----------------|
| | Mock Objects | Mock Functions |
| Controller Unit Test | 13 | 32 |
| Server Unit Test | 4 | 11 |

Table 6.5: Mock Objects and Functions

CHAPTER 7

Related Work

We were unable to find any works that attempted to bring existing technologies together to create a web testing framework. Instead, we came across a number of proprietary tools and research. Compared to these proprietary tools, the framework we already laid out is free and each part is supported by a community. There is no need to contact any paper authors to get software or support. We will, however, still outline them as they provided insight into what the problems in existing frameworks were, gaps in coverage, what was important in a framework, and inspiration and acknowledgment that this is a real problem that many people are trying to solve.

We also found research into common bugs, general testing techniques, and finding bugs that greatly improved our test cases and understanding.

7.1 Testing Frameworks

The following are papers that were written about creating proprietary testing frameworks.

7.1.1 “A Framework for Automated Testing of JavaScript Web Applications”

A collection of IBM Researchers and two university students came up with a framework called “Artemis” for testing web applications [52]. What is novel about their tool is that it generates test cases automatically based on execution patterns and

feedback. They have attempted to solve the problem that writing test cases by hand is both time-consuming and often difficult. Their automatic test generation lead to an average of 69% test coverage with enough tests generated.

Though it was an interesting and somewhat effective method, 69% coverage is not great, we would have preferred around 80%, and this particular framework only covers the client-side, so it was incomplete for our purposes.

7.1.2 “A Multi-Agent Software Environment for Testing Web-based Applications”

Two students and a person from Lanware Limited brought AI into the web testing world with an agent-based environment for testing web applications [46]. In order to split testing into manageable tasks for agents to carry out, they created an ontology for web testing using XML. Each agent is set up to handle one particular kind of testing with certain test data, i.e., a Unit Tester with data for one or two functions. This agent may then communicate with another agent who needs the results from that test, such as a test coverage agent or agent that is keeping track of test success and failure. This communication is done through a message-passing intermediary layer and a set of brokers who shuffle messages between agents.

This system is exciting and intricate but far too complex for a general web testing toolkit that just about any developer could pick up. It was important to weed techniques like this out as we wanted a fairly easy to use and understand toolkit.

7.1.3 “A 2-Layer Model for the White-Box Testing of Web Applications”

Two researchers at the Center for Scientific Research and Technology in Povo, Italy came up with a model for white-box testing of web applications [54]. (White box testing is when the developer is given access to the underlying code.) This paper breaks up the testing process into two abstraction levels, the navigation model, the

way in which a webpage goes from page to page, and the control flow model, the way in which information is passed and stored on a given page or between pages. These constitute the 2-layers in the title. The navigation model represents high-level test cases, like asserting that a given link redirects to the correct location, while control flow represents low-level test cases, like making sure data is persisted between pages.

This 2-layer approach is interesting and provided us with some insight, but the system was made for PHP applications and we weren't looking to implement a brand-new system and test PolyXpress with it all in one year.

7.1.4 “Invariant-Based Automatic Testing of AJAX User Interfaces”

Two individuals from the Software Engineering Research Group at Delft University in the Netherlands came up with a way to automatically test AJAX UI [41]. The core of this idea is using a crawler to infer a flow graph. This paper outlines an plugin for an automated tool, ATUSA, for creating state validators and test suites for covering paths discovered during crawling (with a separate tool, CRAWLJAX). Their tests show that with minimal manual effort, the use of ATUSA can lead to high code coverage and error discovery.

This tool showed promise as an AJAX testing tool, but we wanted something more generic. Their crawler, CRAWLJAX, however, was very interesting and while we didn't use it, the techniques it used are used in other tools we have.

7.1.5 “JSART: JavaScript Assertion-based Regression Testing”

Two students from the University of British Columbia created a tool called JSART that uses run-time code instrumentation and analysis to infer invariants that can be used for regression testing [42]. When the code runs, JSART creates a number of test cases that will be run the next time the code executes. The point being to catch

any discrepancies between old code and new code. They ran their tool on nine web systems with fairly strong results. The tool was not perfect and occasionally created inappropriate tests, but generally the tests were appropriate and useful.

This is a useful concept that could be integrated into the toolkit at a later time.

7.1.6 “DOM Transactions for Testing JavaScript”

Three people from Albert Ludwigs University in Germany came up with a way to deal with test fixtures that are built and torn down frequently [45]. When testing JavaScript, many of one’s tests will change the underlying web page in some way. This is often different from testing traditional software where components are separated. In order to test JavaScript, test fixtures for setting up the page and tearing down the page are exceptionally important. Because these fixtures are being made and torn down frequently, this paper discusses a way to utilize transactional memory to quickly get the web page into the right state without constant setup and tear down.

This could be integrated with one of the existing tools we will use, but our goal was to collect existing tools that did not need alteration.

7.1.7 “Testing Web Applications in Practice”

Four students from the University of Seville, Spain published a paper with an overview of web testing [35]. This paper was more of an introduction to testing, outlining the concepts of Unit Testing, Integration Testing, Regression Testing, testing the client versus testing the server, and a few others. It also describes how to test an actual web application using PHPUnit, a library similar to JUnit, for those developers familiar with Java. When writing their tests, they suggest taking more abstract actions and breaking them up into their functional parts and testing each of those. As an example, inserting a customer into the database (albeit not that abstract) is broken into tests

for the function that inserts the customer and tests for making sure the customer makes it into the database.

The paper was good refresher on testing concepts and a hands-on application of those concepts. There wasn't any new information or tools to put toward our toolkit, but it was a good reminder and outlined some testing procedures.

7.1.8 “Contract-Driven Testing of JavaScript Code”

Two individuals from the Albert Ludwigs University in Germany created a tool, JSConTest, that provides a framework for making contracts for a JavaScript program, allowing for guided random testing on inputs and outputs [30]. Contracts in JSConTest are comments above each function that are of the form “type -> type”. More complicated forms can introduce parenthesis for function types. In general, JSConTest is useful for detecting type errors on input or output based on the operations performed within a function on the input and the result being output.

Although JSConTest could be useful in a toolkit, its source is nowhere to be found and there were other, more robust, tools out there for our purposes.

7.1.9 “JAWS: A JavaScript API for the Efficient Testing and Integration of Semantic Web Services”

A researcher from Ford Motor Company published a paper regarding an API, called JAWS (JavaScript, AJAX, Web Service), to facilitate testing and integration of Semantic Web Services [44]. Semantic Web Services (SWS) are the server side of machine-to-machine interaction via the internet. This is opposed to machine-to-human interaction such as web pages that one visits personally. SWS use markup that is detailed and sophisticated which conforms to certain standards but is not human-readable.

Another interesting API and project, but we are not focusing on SWS and so including a tool or API like JAWS would be overkill and beyond trying to keep the toolkit as simple and robust as possible.

7.1.10 “WebMate: A Tool for Testing Web 2.0 Applications”

Four people from Saarland University in Germany created a tool, WebMate, for automatically navigating web-pages with dynamic content and data [55]. WebMate’s goal is to output a usage model for testing purposes. This usage model allows tests to be focused on navigation as well as functionality. This dual usage leads to the ability to test the experience as well as the functions. In addition, WebMate can be used for cross-browser testing by running it on multiple browsers and comparing the results. In essence, WebMate is a web application crawler seeking to automatically suss out the layout and interaction between web pages to provide a model to the tester, be they a developer or the automatic testing capabilities of WebMate.

WebMate was another interesting project that could have seen use in our toolkit if crawlers weren’t already available in other, more mainstream, tools.

7.1.11 “Continuous Testing with Ruby, Rails, and JavaScript”

This well known book by Rady and Coffin [47] talks about Continuous Testing and has lots of examples. The last two chapters of the book, six and seven, talk about setting up a Continuous Testing environment for JavaScript using Node.js and a number of other tools. It fit almost perfectly with our project and was used as the basis of our testing setup. They use JSLint for linting, Jasmine for test cases, and Watchr for running the tests continuously. Lastly it discussed writing effective tests, which is always important.

Not all of those tools made it into the eventual final toolkit, but they were all

important to look at in the context of making the overarching toolkit.

7.2 Finding Bugs

One of the most important aspects of testing is actually finding the bugs in the program. A number of sources provided guidelines and suggestions for finding bugs in web applications.

7.2.1 “Finding Bugs In Dynamic Web Applications”

Seven researchers from the MIT Computer Science and Artificial Intelligence Lab wrote a paper addressing testing in dynamic web applications [51]. Common testing techniques were not made to work with dynamic content that can be generated at the drop of a hat or click of a button. Therefore, common testing techniques are inadequate for testing dynamic web applications. This paper discusses ways to find bugs in these new-fangled dynamic web applications. They use a new technique and extended a tool called Apollo. This technique generates dynamic test cases based on concrete and symbolic execution. “The basic idea is to execute an application on an initial input (e.g., an unconstrained or randomly chosen input), and then on additional inputs obtained by solving constraints derived from exercised control flow paths.”

Although they test PHP applications, the extension to JavaScript is fairly straightforward. We didn’t use Apollo, but the techniques described within allowed for more focused and productive testing.

7.2.2 “Testing Web Services: A Survey”

Three researchers at the Centre for Research on Evolution, Search & Testing at King’s College in London wrote a survey of web frameworks for testing web services [43]. Their focus is on Service-Oriented Computing, a paradigm in which web sites or applications provide small pieces of functionality usable by “anyone”. These services are then combined to create more fully featured applications. What they note in particular is that when using web services, the developer rarely has access to the source code and must therefore black-box test them. Corollary to that is that the developer has to implicitly trust whoever made the service. They go through and discuss a number of frameworks and come to the conclusion that it’s most certainly not a solved problem for the following reasons: The frequency of testing required, testing without disrupting the operation of the service, determining when testing is required and which operations need to be tested. Regarding the frequency of testing, web services often change in effort to make improvements, but those improvements can change the services and break tests. Testing without disrupting service is difficult because these services often limit accesses or simultaneous access of both a test and a running web application. Finally, determining when and what to test comes back to the trust issue. Some services provide their own testing metrics that include code coverage, number of tests, etc, but the developer must trust that the service is thoroughly tested, not just tested enough to provide interesting metrics.

7.2.3 “A Framework for Testing RESTful Web Services”

Two people from the University of Dakota School of Aerospace Studies outline a framework for testing RESTful web services [48]. RESTful web services, short for Representation State Transfer, are, in the context of web applications, a name for web APIs that conform to a set of constraints. In particular they must offer four request

methods: GET, POST, PUT, and DELETE. These methods generally connect to a database or other storage system in order to store, retrieve, or delete entries therein.

7.3 Writing Effective Test Cases

Although not too distinct from *finding bugs*, writing effective test cases is its own art form. The important of writing concise, correct, productive test cases cannot be overstated. The difference between writing one test case that effectively tests multiple things and writing a single test case for every little detail is the difference between developers wanting to write tests at all and skipping it due to time constraints or frustration.

7.3.1 “Going Faster: Testing The Web Application”

Two employees of Evant Software wrote an article for IEEE Software regarding the testing of web applications [33]. Mostly this paper was an overview and implementation of Test Driven Development (TDD) and eXtreme Programming, but in the middle of the article was a good section about the difficulties of testing the web and how to deal them. A few of the remarks regarding the difficulty of web testing is that often the client-side and server-side portions of web applications are written in different languages. This wasn’t an issue for our application, but is still an issue for many applications. Some of the most useful advice was: “First, test those parts of the server-side code that are not directly concerned with being part of the Web application, without involving Web peculiarities. ... Second, test those parts of the client-side code that have no server interaction. This is typically code that contains little or no business logic. ... Third, write functional tests of a low grain in the server-side language (for example, Java or C++) that simulate the request/response Web environment. ... Using this framework you can write walk-through tests that

simulate the user moving through the UI”.

This step-by-step walkthrough of how to test was an excellent starting point to look at different tools and how they handled this issue.

CHAPTER 8

Conclusion

It is our hope that after reading through the tools and workflow described here, any developer would be able to hone in on the tools they want without having to scour the web for the best tool for their job. We feel that we've covered all of the latest and greatest tools that offered support and ease of use to new developers.

CHAPTER 9

Future Work

Given unlimited time and resources, we would have implemented a number of other things.

Full PX coverage

It's no surprise that given enough time we'd like to extend this to all of PolyXpress so that when new code is written, it doesn't matter which part of the codebase the developer is in, they'll have tests for all of it.

“Real-world” Evaluation

Although PolyXpress is a complex application that has been used by people outside of Cal Poly, it would have been nice to try out our framework on an existing open-source project and on a brand new project to see how the workflow would go on an established project or from the ground up.

Headless Browsers

The notion of headless browsers is extremely interesting and worthy of exploration. PhantomJS appeared to be quite promising for a full DOM implementation or ZombieJS for more simple DOM emulation. A side-by-side comparison of tests running through a full-fledged browser and tests running through a headless browser would

be interesting and enlightening as to whether or not headless browsers could have a real place in a testing framework.

Sauce Labs

Sauce Labs, the online web testing site, offered a phenomenal set of features that we did not have time to take advantage of. The ability to test web applications across multiple browsers and browser configurations without having to set all of them up is enticing.

BIBLIOGRAPHY

- [1] Browser Usage Statistics. <http://gs.statcounter.com/#desktop-browser-ww-monthly-201402-201402-bar>.
- [2] ChaiJS. <http://chaijs.com/>.
- [3] Grunt: The Javascript Task Runner. <http://gruntjs.com/>.
- [4] intern.
- [5] Jasmine. <http://jasmine.github.io/>.
- [6] Jenkins Organization. <https://github.com/jenkinsci>.
- [7] jQuery. <http://jquery.com/>.
- [8] MagnumCI. <https://magnum-ci.com>.
- [9] Mocha. <http://visionmedia.github.io/mocha/>.
- [10] PolyXpress. <http://pxdev.cfapps.io/>.
- [11] QUnit. <http://qunitjs.com/>.
- [12] Samsung RF4289HARS. <http://www.samsung.com/us/appliances/refrigerators/RF4289HARS/XAA>.
- [13] Sauce Labs. <https://saucelabs.com/>.
- [14] Selenium. <http://docs.seleniumhq.org/>.
- [15] Semaphore. <https://semaphoreapp.com/>.

- [16] Semaphore Blog. <https://semaphoreapp.com/blog/>.
- [17] Usage of JavaScript for websites. <http://w3techs.com/technologies/details/cp-javascript/all/all>.
- [18] Why AMD? <http://requirejs.org/docs/whyamd.html>.
- [19] Ieee standard for software unit testing. *ANSI/IEEE Std 1008-1987*, 1986.
- [20] Systems and software engineering – vocabulary. *ISO/IEC/IEEE 24765:2010(E)*, pages 1–418, Dec 2010.
- [21] Software and systems engineering software testing part 1:concepts and definitions. *ISO/IEC/IEEE 29119-1:2013(E)*, pages 1–64, Sept 2013.
- [22] Krishnan Anantheswaran. Istanbul. <https://github.com/gotwarlost/istanbul>.
- [23] Assaf Arkin. Zombiejs. <http://zombie.labnotes.org/>.
- [24] Scott Bellware. Behavior-Driven Development, June 2008.
- [25] Travis CI. Travis. <https://travis-ci.org/>.
- [26] Travis CI. Travis Plans. <https://travis-ci.com/plans>.
- [27] Danny Coates. Node Inspector. <https://github.com/node-inspector/node-inspector>.
- [28] World Wide Web Consortium. Plan 2014. <http://dev.w3.org/html5/decision-policy/html5-2014-plan.html>, 2012.
- [29] Karthik Pattabiraman Frolin S. Ocasriza Jr. and Benjamin Zorn. JavaScript Errors in the Wild: An Empirical Study. http://ece.ubc.ca/~frolino/projects/jser/tech_report.pdf, 2011.

- [30] Phillip Heidegger and Peter Thiemann. Contract-Driven Testing of JavaScript Code. <https://proglang.informatik.uni-freiburg.de/jscontest/jscontest/>, 2010.
- [31] Miko Hevery. jasmine-node. <https://github.com/mhevery/jasmine-node>.
- [32] Ariya Hidayat. Phantomjs. <http://phantomjs.org/>.
- [33] Edward Hieatt and Robert Mee. Going Faster: Testing The Web Application. *IEEE Software*, pages 60–65, 2002.
- [34] Hyperakt and Vizzuality. The Evolution of the Web. <http://www.evolutionoftheweb.com/>.
- [35] Manuel Mejas Jess Torres Javier Jess Gutierrez, Maria Jos Escalona. Testing web application in practice. In *First International Workshopl, WWV*, 2005.
- [36] Inc. Joyent. Node.JS. <http://nodejs.org/>.
- [37] Kohsuke Kawaguchi. Jenkins. <http://jenkins-ci.org/>.
- [38] Vidar Kongsli. Security Testing with Selenium. <http://www.kongsli.net/oopsla/OOPSLA'07%20Security%20Testing%20with%20Selenium%20v1.0.pdf>, 2007.
- [39] Azat Mardanov. Todo App with Express.js/Node.js and MongoDB. <http://webapplog.com/todo-app-with-express-jsnode-js-and-mongodb/>.
- [40] Steve McConnell. Code Complete, 2nd Edition, 2004.
- [41] Ali Mesbah and Arie van Deursen. Invariant-Based Automatic Testing of AJAX User Interfaces. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.152.9899&rep=rep1&type=pdf>, 2009.

- [42] Shabnam Mirshokraie and Ali Mesbah. JSART: JavaScript Assertion-based Regression Testing. <http://www.ece.ubc.ca/~amesbah/docs/icwe12.pdf>, 2012.
- [43] Mark Harman Mustafa Bozkurt and Youssef Hassoun. Testing Web Services: A Survey. Technical report, King’s College London, 2010.
- [44] David A. Ostrowski. JAWS: A Javascript API for the Efficient Testing and Integration of Semantic Web Services. <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-248/paper9.pdf>, 2007.
- [45] Annette Bieniusa Phillip Heidegger and Peter Thiemann. DOM Transactions for Testing Javascript. <https://proglang.informatik.uni-freiburg.de/jscontest/trans.pdf>, 2010.
- [46] Hong Zhu Qingning Huo and Sue Greenwood. A Multi-Agent Software Environment for Testing Web-based Applications. In *International Computer Software and Applications Conference*, 2003.
- [47] Ben Rady and Rod Coffin. *Continuous Testing with Ruby, Rails, and JavaScript*, chapter Creating a JavaScript CT Environment. The Pragmatic Bookshelf, 2011.
- [48] Hassan Reza and David Van Gilst. A Framework for Testing RESTful Web Services. In *Seventh International Conference on Information Technology*, 2010.
- [49] Andrei Rusu. Nightwatchjs. <http://nightwatchjs.org/>.
- [50] Brad Rydzewski and Thomas Burke. Drone.io. <https://drone.io/>.
- [51] Julian Dolby Frank Tip Danny Dig Amit Paradkar Shay Artzi, Adam Kiezun and Michael D. Ernst. Finding Bugs In Dynamic Web Ap-

- plications. <http://dspace.mit.edu/bitstream/handle/1721.1/40249/MIT-CSAIL-TR-2008-006.pdf>, 2008.
- [52] Simon Holm Jensen Anders Moller Shay Artzi, Julian Dolby and Frank Tip. A Framework for Automated Testing of JavaScript Web Applications. <https://cs.uwaterloo.ca/~ftip/pubs/icse2011artemis.pdf>, 2011.
- [53] Travis Tidwell. Zombie Phantom. <https://github.com/travist/zombie-phantom>.
- [54] P. Tonella and F. Ricca. A 2-layer model for the white-box testing of web applications. In *Telecommunications Energy Conference, 2004. INTELEC 2004. 26th Annual International*, 2004.
- [55] Tobian Orth Valentin Dallmeier, Martin Burger and Andreas Zeller. WebMate: A Tool for Testing Web 2.0 Applications. <http://www.testfabrik.com/assets/pdf/webmate-jstools12.pdf>, 2012.

CHAPTER 10

Appendices

The following appendices include the test code for the ToDo application, screenshots of the various PolyXpress modules, and the PolyXpress test code.

10.1 PolyXpress Tutorial

We also created a thorough tutorial for future PolyXpress developers that is available at <https://github.com/tgashby/PolyXpress/wiki/Tutorial>

The tutorial makes no assumptions and goes from getting the code from the repository all the way through running the tests that are already there. In addition, there are detailed sections on writing Unit Tests using Mocha with Chai.js and writing UI tests with Nightwatch.

10.2 ToDo App Test Code

The ToDo application and related tests are available at <https://github.com/tgashby/ThesisTestApp>

The Unit Testing code is located in the `test` folder and consequent sub-folders.

Unfortunately most of the CI frameworks used GUI configurations that are not included in the repository.

10.3 PolyXpress Player Screenshots

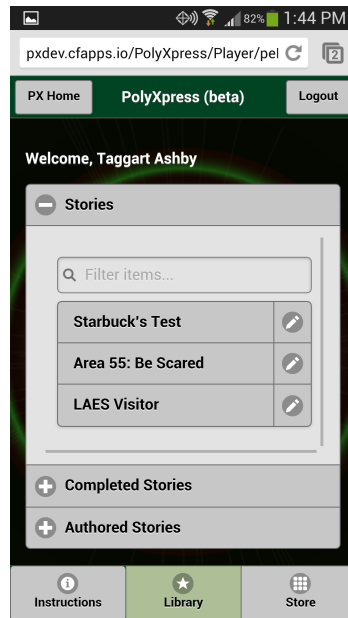


Figure 10.1: Story selection

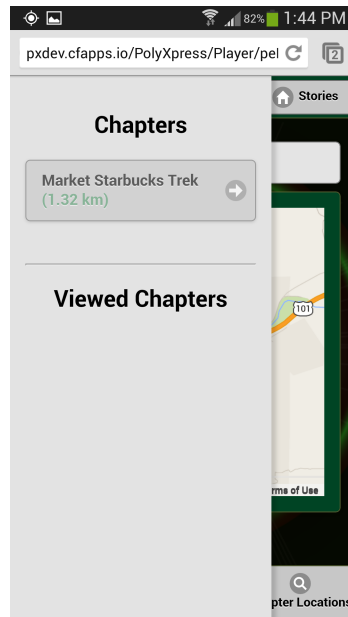


Figure 10.2: Chapter pane with approximate distances to each chapter

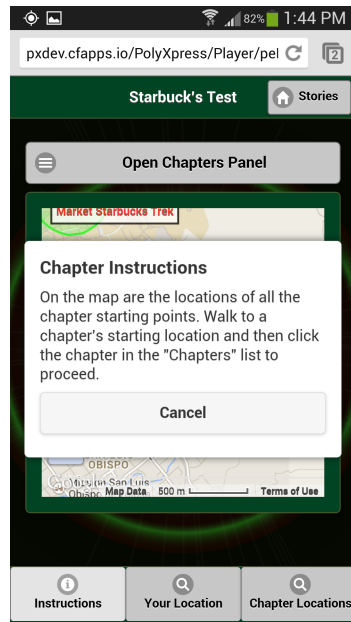


Figure 10.3: Each page and element of PolyXpress has instructions, here they are for Chapters

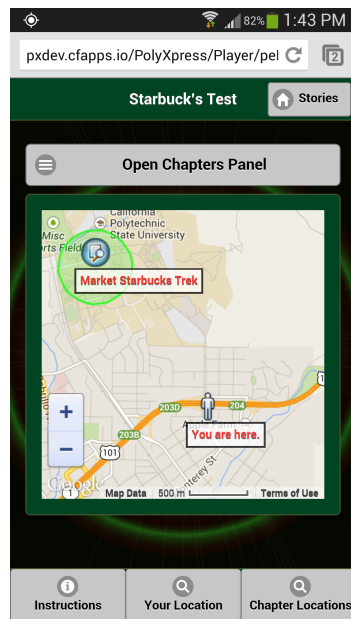


Figure 10.4: A map of the chapter location in relation to the player

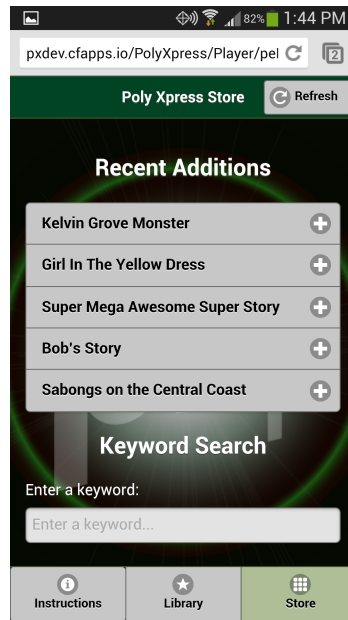


Figure 10.5: PolyXpress marketplace, to find other published stories

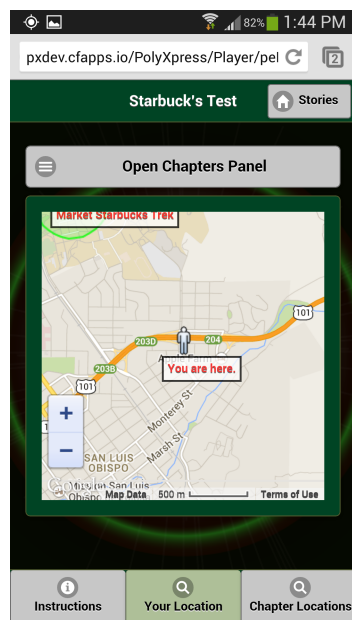


Figure 10.6: A map centered around the player

10.4 PolyXpress Author Screenshots

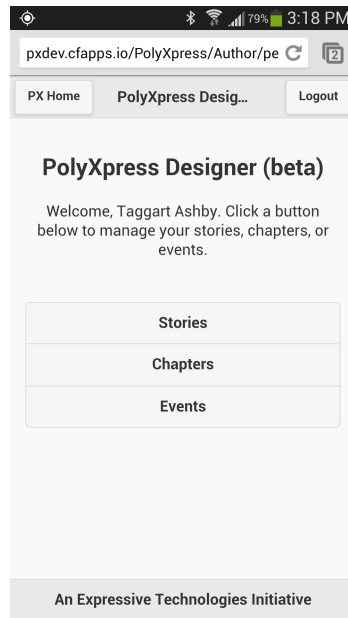


Figure 10.7: PolyXpress Authoring Tool

A screenshot of the 'Create Event' form within the PolyXpress Designer mobile app. The status bar at the top shows a battery level of 78% and the time 3:18 PM. The address bar displays 'pxdev.cfapps.io/PolyXpress/Author/pe'. The navigation bar shows 'PolyXpress Desig...'. The main heading is 'Create Event'. Below this is a section titled 'Administration' containing an 'Author Management' button. Underneath are input fields for 'Authors:' (containing the value '5283b1180a6c701f00000003'), 'Version:', and 'Keywords:'. There is also a 'Repeats?' section with a checkbox and the text 'Click if this a reoccurring event.' At the bottom are 'Save' and 'Cancel' buttons.

Figure 10.8: Form to create an Event

pxdev.cfapps.io/PolyXpress/Author/pe

Repeats?

☐ Click if this a reoccurring event.

Context

Title:

Geographic Location Information

Open Map View

(35.29066743) Latitude:

(-120.65294036) Longitude:

Save Cancel

Figure 10.9: Bottom of form to create an Event

pxdev.cfapps.io/PolyXpress/Author/pe

PolyXpress Desig...

Create Chapter

Administration

Author Management

Authors:

5283b1180a6c701f00000003

Version:

Keywords:

Context

Save Cancel

Figure 10.10: Form for creating Chapters

The screenshot shows the PolyXpress Design app interface. At the top, the status bar displays the time as 3:19 PM and battery level at 78%. The app title 'PolyXpress Design...' is visible. The main section is titled 'Geographic Location Information' and contains an 'Open Map View' button. Below this, there are input fields for Latitude (pre-filled with '(35.29060952)') and Longitude (pre-filled with '(-120.65290728)'), and a 'Range(km):' field. The 'Event Management' section follows, with an 'Events:' input field and an 'Existing Events:' section that is currently empty. At the bottom, there are 'Save' and 'Cancel' buttons.

Figure 10.11: Events are selected manually for inclusion in Chapters

The screenshot shows the PolyXpress Design app interface for creating a story. The title 'Create Story' is at the top. Below it is the 'Administration' section, which includes an 'Author Management' button. The 'Authors:' field contains the text '5283b1180a6c701f00000003'. The 'Version:' field is empty. The 'Publish?' section has two radio buttons: 'Click if the story is ready to share with the world.' and 'Click if you would like to test the story in PolyXpress Player.' At the bottom, there are 'Save' and 'Cancel' buttons.

Figure 10.12: Story creation asks if the user wants their story published