Automatic Generation of Artifacts for Two Web Application Testing Models

**Submitted by:** Sarah Vessels

**Advisor:** Dr. Jane Hayes

Department of Computer Science

University of Kentucky

Lexington, KY 40506

Monday, May 9, 2011

**Committee members:**

Dr. Raphael Finkel

Dr. Judy Goldsmith

Dr. Jane Hayes

# Abstract

Web applications are prevalent and it is important that they be of high quality since businesses, schools, and public services rely upon them. Unique testing models designed for web applications can be beneficial since web applications differ greatly from desktop applications. Web applications are accessed via a browser, and a user can manipulate the web application in ways not possible with desktop applications, such as modifying the URI or using the Back button in the browser. It can take a great deal of time to apply a given testing model to a web application, due to the size of the application as well as the many steps in applying a model. This project seeks to decrease the time necessary to apply two particular web application testing models: the Atomic Section Model and the Qian, Miao, Zeng model. Two tools were written using the Ruby programming language, one for each model. The tools take as input the source code of a Ruby on Rails web application, and the URI to a web application written with any framework, respectively. They produce as output test paths that can be traversed manually to ensure good coverage of the web application, and artifacts that can be further manipulated manually to produce test paths, respectively. Through the use of these tools, a web application developer can better see how to test his or her application, and can see all the paths through the application that a user might take.

Contents

[Abstract 2](#_Toc290332933)

[Definitions 4](#_Toc290332934)

[1. Introduction 6](#_Toc290332935)

[2. Requirements for the Qian, Miao, and Zeng Model Tool 7](#_Toc290332936)

[2a. Implementing the QMZ Tool 9](#_Toc290332937)

[2b. Testing the QMZ Tool 13](#_Toc290332938)

[2c. Sample Output from the QMZ Tool 14](#_Toc290332939)

[3. Requirements for the Atomic Section Model Tool 16](#_Toc290332940)

[3a. Implementing the ASM Tool 18](#_Toc290332941)

[3b. Testing the ASM Tool 27](#_Toc290332942)

[3c. Sample Output from the ASM Tool 28](#_Toc290332943)

[Single File Output 28](#_Toc290332944)

[Rails Application Output 30](#_Toc290332945)

[4. Code Statistics 33](#_Toc290332946)

[5. Future Work 34](#_Toc290332947)

[5a. QMZ Tool 34](#_Toc290332948)

[5b. ASM Tool 34](#_Toc290332949)

[6. Conclusions 36](#_Toc290332950)

[7. Bibliography 37](#_Toc290332951)

# Definitions

* **AJAX** - Asynchronous JavaScript and XML; JavaScript code that produces an asynchronous call to the server to retrieve data, without reloading the web page in the browser
* **CSS** - Cascading Style Sheets; a language used to describe the appearance of a web page
* **CSS** **selector** - a pattern matching rule in CSS that is used to determine which styles apply to which elements in an HTML page
* **ERB** - "a templating system that embeds Ruby into a text document... often used to embed Ruby code in an HTML document, similar to ASP, JSP and PHP" [[[1]](#endnote-1)]; sample file extensions of an ERB file include .rhtml and .html.erb
* **ERB HTML view** - a Ruby on Rails view written using embedded Ruby and HTML; see also **Ruby on Rails view**
* **ERB output tag** - found in an ERB file, this tag allows Ruby code to be evaluated and the result to be output when the ERB file is used as a template; the result of the Ruby code within an ERB output tag is displayed as part of the web page when it is included in a Ruby on Rails ERB view; it is started by <%= and ended with %>
* **ERB tag** - found in an ERB file, this tag allows Ruby code to be embedded in an ERB file; the results of the evaluation of the Ruby code will not be included as part of the source of the output file once the ERB is applied as a template; an ERB tag is started by <% and ended with %>
* **form transition** - a form within a web page that has a destination page and a button that, when activated, navigates the browser to the destination page
* **gem** - a Ruby library, often installed through the use of the RubyGems package manager[[2]](#footnote-1)
* **HTML** - HyperText Markup Language; "the predominant markup language for web pages... the basic building-blocks of webpages" [[[3]](#endnote-2)]
* **hyperlink** - a link within a web page that, when followed, will navigate the browser to the target of the hyperlink
* **RHTML** - see ERB
* **Ruby on Rails** - a web application development framework using the Model View Controller paradigm [3]
* **Ruby on Rails partial view** - Ruby on Rails has the concept of partials, or partial views, which are Ruby on Rails views that do not represent an entire web page; a partial might contain only a small portion of a web page, and another view within the Rails application can embed that partial within its own contents (referred to as rendering the partial); see also **Ruby on Rails view**
* **Ruby on Rails view** - can refer to either the source code for a page within a Ruby on Rails application, or the rendering of that page by Rails within a web browser; may contain HTML, Javascript, and CSS; a view can represent an entire page in the web application, or a part of a page (see **Ruby on Rails partial view** ); a view's source code is contained within an ERB file
* **s-expression** - symbol expression, a.k.a. sexp; s-expressions are "list-based data structures that represent semi-structured data" [[[4]](#endnote-4)]; the Ruby gem ruby\_parser class RubyParser returns s-expressions when Ruby code is successfully parsed [[[5]](#endnote-5)]
* **URI** - Uniform Resource Identifier; used to identify the location of a web site on the Internet
* **web scraper** - a tool that saves the source of a web page for later processing
* **YAML** - YAML Ain't Markup Language; a data serialization format akin to XML; YAML support is included in the Ruby 1.8 standard library

# 1. Introduction

Web applications are prevalent and users expect them to perform reliably and quickly. One of the ways that web developers ensure that users are satisfied is by testing their applications. Web application testing differs from desktop application testing because the primary way of using a web application is through another application: a web browser. As a result, web applications are subject to the user’s manipulation in ways that are not common in desktop applications. For example, a user can manually change the URL in the web browser to point to another page in the web application. Users can use their browser’s Back and Forward buttons to change pages. Different plugins, browser settings, and browser extensions alter the user experience. All these factors can provide the web application with unexpected input, or cause it to perform actions that the developer did not anticipate. It is for this reason that different testing models exist solely for testing web applications. This project implements two models for web application testing, discussed below.

The first model is the Atomic Section Model (ASM) [[[6]](#endnote-6)] by Jeff Offutt and Ye Wu. The ASM represents all possible user interface screens of a web application, just as control flow graphs represent all possible sequences of software program execution. ASM can be divided into two levels: the Component Interaction Model (CIM) and the Application Transition Graph (ATG) [6]. CIM is designed for the individual components of the web application while ATG is for the web application as a whole, in which the individual components come together and function as a system. The key idea of the ASM is atomic sections: a section of code that is delivered as a unit to the browser. This code can contain markup such as HTML and CSS, as well as client-side scripting such as JavaScript. An atomic section has the property that if any part of it is sent to the browser, all of it is sent to the browser. By identifying atomic sections in a page, how the atomic sections are combined in a page, and transitions between pages, paths through the web application can be found.

Another web application testing model is the Qian, Miao, and Zeng model (QMZ) [[[7]](#endnote-7)]. Applying the QMZ model involves the creation of a Page Flow Diagram (PFD) [7], which represents “the relationship among pages of a web application” [7]. A PFD has nodes that represent pages in the web application, connected by edges that represent different transitions that the user can take to reach one page from another. A PFD can have cycles when a page links to a different page, which then links back to the first page. As such, it is necessary to manipulate the PFD into an acyclic structure to determine unique paths through the application. A Page Test Tree (PTT) [7], based upon the PFD, is such a structure, and is a simpler representation of the web application than that provided by the PFD.

As part of this project, I have written tools in Ruby to partially automate the process of applying the ASM and the QMZ models to a web application. Ruby was chosen because it is a language with which I am comfortable, and several libraries used in this project already existed for Ruby, such as for parsing HTML and parsing Ruby. Due to the nature of the QMZ model, the underlying language of the web application being tested does not matter, so the QMZ tool is platform-agnostic. The ASM works by considering the code of web pages in the application, and so it was necessary to pick a programming language/framework to target when writing a tool for generating ASM artifacts. I chose to target the Ruby on Rails web application framework because: 1) it is popular,[[8]](#footnote-2) and 2) Ruby applications for testing are widely and freely available online. Though my tools were written and tested in OS X 10.5.8 and Windows 7 using Ruby 1.8.7, they should be executable in any operating system for which there is a Ruby interpreter. The different Ruby gems that my tools use must also be installed; these include Nokogiri, RubyParser, and Treetop. The use of these gems is described below.

Note that my tools do not fully automate web application testing, which must still be performed manually by a human. The QMZ tool generates test paths to follow (i.e., URIs to access in a particular order) to achieve good coverage of the pages and transitions in a web application, where "good coverage" is defined according to the QMZ model. The ASM tool generates CIMs which can be combined into an ATG and then traversed in various ways to achieve coverage of the web application. Test cases would need to be constructed manually for each model to provide input to forms encountered in the web application, and to describe how each page is expected to appear when it is reached after following a certain path of web pages. Manual testing is required to ensure that, upon following these test paths, the web pages display as expected and that no errors occur. My contribution is to remove the repetitive, tedious work of applying these web application testing models, especially for large web applications with many pages.

The report is organized as follows. Section 2 describes the Qian, Miao, and Zeng model, including my tool's requirements, implementation, and testing. Section 3 describes the Atomic Section Model, including my tool's requirements, implementation, and testing. Section 4 provides statistics about my tools' source code. Future work for both tools is detailed in section 5. Section 6 has my conclusions and the bibliography is in section 7.

# 2. Requirements for the Qian, Miao, and Zeng Model Tool

* The QMZ tool shall take as input a URI to a web application and output a list of test paths that traverse the pages of that web application.
* The test paths produced by the QMZ tool shall include all pages in the web application accessible via other pages in the application via hyperlinks and forms.
* A page shall be considered part of the web application if its URI's host name is the same as the host name of the given root URI of the application.
* The produced list of test paths shall be presented in an HTML page where each URI in each test case is a hyperlink.
* The tool shall scrape web pages within the web application and correctly parse their HTML contents in order to find hyperlinks and forms.
* The tool shall follow all hyperlinks and form transitions to unscraped pages in the web application.
* A PFD shall be constructed that represents the web application, according to the QMZ model.
* A PTT shall be constructed using the PFD.
* The tool shall provide a command-line interface.
* All test paths produced by the tool for a given web application shall be unique, by URI and method of transition.
* The tool shall be able to generate test paths for a web application written with any programming language or framework whose pages are in HTML.
* The tool shall not attempt to scrape images, PDF files, ZIP files, or other non-HTML pages within a web application.
* The tool shall keep traversing the web application upon encountering a broken link, 500 error, 403 error, or other error code when scraping a page.
* The tool shall not repeatedly try to scrape a page that previously resulted in an error code or non-HTML page.
* The tool shall not scrape a page at the same URI more than once.
* The tool shall treat instances of the same URI found in different hyperlinks or form transitions within a page as different instances when constructing a test path.

## 2a. Implementing the QMZ Tool

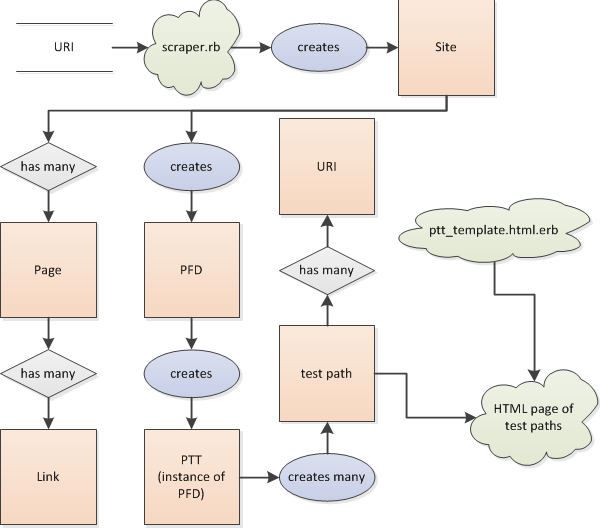


Figure 1 Overview of QMZ tool

To apply the QMZ model to a web application, I had to represent the PFD and PTT for a given site. To construct these, the extent of the web application must be known. This means knowing all the pages within the web application that are accessible via hyperlinks or form transitions on other accessible pages, starting with the root or start page of the application. The root/start page is provided by the user. My QMZ tool first functions as a web scraper to download pages it determines to be part of a given web site, then it extracts the links to other pages within the site from those pages. To execute my QMZ tool, the Ruby script scraper.rb is run on the command line with various parameters. A URI to the web application can be passed with the -u option followed by the URI. For example: scraper.rb -u "http://mywebapp.example.com/". The scraper script constructs a new instance of my Page class, which is then used to construct a new instance of my Site class.

A Page requires a URI either as a Ruby String or an instance of the URI class. Upon creating a new Page, the Page class attempts to download the given URI; if it fails, an ArgumentError is thrown. If the URI is successfully downloaded, the HTML of the page is scraped for links. To the QMZ model, links come in the form of hyperlinks and form transitions. To identify these in a given chunk of HTML code, the Ruby gem Nokogiri[[9]](#footnote-3) was used. Nokogiri is an HTML, XML, SAX, and Reader parser, though only its HTML-parsing features were used in my project. The Nokogiri::HTML module is passed a chunk of HTML that it parses. The result is a Nokogiri::HTML::Document that can be traversed using CSS selectors such as 'a' or 'form.' A list of matching elements from the HTML can be extracted by using the css method on a Nokogiri document. For example, Nokogiri:: HTML(anHTMLString).css('form') would return a list of all the form elements in the given HTML string, case insensitive. The css method can be called on each of these elements in turn to find other elements contained within them, such as buttons in a form. Finally, the values of particular attributes can be extracted from HTML elements through Nokogiri, so the target of a form--its action attribute--can be extracted. A Page finds all forms in the HTML that it scraped from the given URI, then narrows that list down to forms that have a target and have some type of button that the user can activate to submit the form. It further narrows down this list of form elements to those whose targets point to another page in the same site that is being scraped. This is determined by comparing the URI host of the target with the Page URI; relative links (links with a URI that does not specify a host) are always included.

Hyperlinks are similarly extracted from HTML by the Page class using Nokogiri. Hyperlinks in HTML are written using the a tag, so Nokogiri's css method is called with the CSS selector parameter 'a.' The resulting list of hyperlinks is pared down to just those that have a target--the href attribute--that is within the same site that is being scraped.

For both form transitions and hyperlinks, a description of the link is retained for later use. For form transitions, the description is the value of the submit button, or the source of the image if it is an image submit button. For hyperlinks, the description is the link text. These descriptions are retained so that when test paths are presented to the user, the user knows how to access a particular link from a page. This is helpful because there may be multiple ways to access a given page: if a single form had a submit button with the value "Confirm" and another submit button with the value "Reject,” for example. These two buttons would be treated as separate transitions from the source page to the target of the form, thus test paths would be generated to traverse both of them. The server might do different things depending on which button is used, so it is good to test both possibilities.

When a new instance of my Site class is created, the root/home page of the site is passed as a parameter. This is represented by an instance of the Page class, which stores information about the page's URI and the links found within the page's source code. From that home page, the Site class will create other Page instances to represent the other pages in the web application. If two separate pages in a site have a link to the same target page, only one copy of the target page will be represented as a Page. The same is true if two pages have different links that resolve to the same target page, such as when an HTTP redirect is performed by the server.

Two different URIs can point to the same page without a redirect being involved; for example, http://google.com and http://google.com/#comments. The latter URI points to the same page, but to a different marker in that page. To determine if two different URIs point to the same page in order to avoid scraping the same page multiple times, my tool considers the scheme (e.g., http), host (e.g., google.com), and request portion of the URI. The request portion will contain parameters such as ?query=test, which could affect which page is displayed, but not #comments.

If pages link to images, PDF files, or other non-HTML files, these will be skipped. If any link within a page produces a 404 File Not Found or similar error, or is a link to a non-HTML page, that link will be retained in a blacklist of links that should not be traversed. This blacklist helps speed up the process of scraping a web application by not repeatedly attempting to scrape a page that failed previously. The Site.get\_pages method is recursively called to traverse the web application through its form transitions and hyperlinks. The QMZ model knows only about the parts of the web application that are accessible via these links. If a page exists that could be accessed by directly entering a URI in a browser but is not linked to by any other page in the site, that page will be skipped by the QMZ model and my tool.

Once a Site has been constructed, the get\_pfd method is called to construct a PFD that represents the web application. This method returns an instance of my PFD class, which contains pages and links represented by the Page and Link classes, respectively. The QMZ model describes a PFD as containing pages, links, and flows. Flows are modeled in my tool via the Link class, which contains the source URI, the target URI, and the target Page. A list of Link instances is constructed in the get\_pfd method by iterating over each Page in the Site and then iterating in a nested loop over all the link URIs within that Page. A target page is found by extracting the Page with the same URI parts (scheme, host, and request) as the current link URI. A list of Link instances is maintained and used, along with a list of Page instances and the root URI, to create a new PFD instance.

The PFD is immediately converted into a PTT, so as to remove cycles that may exist among pages in the PFD []. This is done via the Site.pfd2ptt class method, which takes in an instance of the PFD class and returns a new instance. A PTT consists of pages and links, like a PFD, and so the same class (PFD) can be used to represent them both. My pfd2ptt method works by making a copy of the provided PFD, such that the original PFD will not be changed, and then applying the PFD2PTT algorithm [] from the QMZ model to that copy. Copies are made of Page instances, as per the algorithm, and an is\_copy flag is set on those pages for later use during traversal. The PFD copy is thus modified to remove cycles and now "is a tree expressing the successive relationship among pages" [].

Once the PTT is acquired, scraper.rb calls the get\_test\_paths method on the PFD instance, which returns an array of test paths. The method works by doing a preorder traversal on the PTT, calling the private PFD method preorder. The preorder method takes in a Page; an integer level; an array of test paths, each of which is itself an array; and a description of the link followed to get to that page. An instance of the LinkText class is used to pair a URI with a description of how the URI was reached; each test path is an array of LinkText instances. The first LinkText in any test path always represents the root page of the site, with the description "Start page." A new LinkText is constructed with its URI set to the given Page's URI and its description set to the given description. The choice of where to put the LinkText is between: 1) the end of the last test path in the array of test paths, or 2) the end of a new test path that begins the same way as the last test path in the array of test paths. If the number of LinkTexts in that last test path is less than or equal to the current level, then the LinkText is appended to the end of that test path. Otherwise, a new test path is created by duplicating a range of the last test path, from the first index up to the levelth index, and then the new LinkText is appended to the end. That new test path is appended to the end of the array of test paths. If the given Page does not have its is\_copy flag set, the page's links are iterated over and preorder is recursively called for each one, with the parameters being the target page of the Link, the current level plus one, the array of test paths as it is now, and the Link description. Because of this recursive call, the array of test paths gets modified. The modified array will either have a new test path appended to the end, or the final test path will get a new LinkText. Finally, the array of test paths is returned from preorder.

The result of the preorder method is an array of arrays of LinkText instances, which is returned by PFD.get\_test\_paths. These represent the test paths a user should take, with each starting at the root page of the site and ending on one of the pages in the site. The test paths differ by both the pages they traverse as well as the transitions used to travel from one page to another. Once scraper.rb gets the test paths, it then creates an HTML page to display them. It styles the page using CSS and stores both the HTML page and the CSS file in a new directory whose name is based off the root page URI of the site. If the root URI of the site were http://google.com/, then the directory containing the test paths would be called http.google.com.

## 2b. Testing the QMZ Tool

To test the QMZ tool, I ran it against known small web applications. The console output was examined followed by the produced HTML file. I went over test paths to ensure that they covered both link and form transitions, and that if two test paths followed the same URIs, it was because they used different transitions. I also sought out different, unknown web sites and ran the QMZ tool against them. This led to the discovery of new defects in the QMZ tool, such as when it tried to follow mailto: links and also when it could not compare an ftp:// link to an http:// link because they were represented by different URI classes: URI::FTP and URI::HTTP, respectively.

The biggest difficulty in testing the QMZ tool was finding web applications small enough that I could manually discover transitions for comparison with the tool's output. It was helpful to run the tool against larger web sites just to test it against new and different HTML, but I did not let the tool run to completion on these because of the sheer number of pages found on a site like Google. Recall that neither the QMZ nor the ASM tool identifies defects in web applications (save for invalid Ruby identified with the ASM tool), rather they each produce artifacts that a human can use to find defects. Thus, testing the QMZ tool means running it against a web application hosted on a server and ensuring that it does not crash when strange code is scraped from a web page or when a page cannot be scraped, and also that the test paths produced fully cover the paths through the web application. The first testing goal was easiest to reach and did not require a site to be known in its entirety: running the QMZ tool against a myriad of sites allowed it to encounter many different HTML coding styles and transition types. This helped identify weak areas in the QMZ tool where it was too restrictive or naive about the input it expected and encountered. The second testing goal of producing paths that fully cover the web site was harder to meet, because it required that I know a web application fully so that I could verify that the paths produced by the tool were all the paths possible. I attempted to partially fulfill this goal by running the QMZ tool against a known web application that I helped write as part of a different project.

My test strategy involved running the QMZ tool against known web applications and manually confirming that the test paths produced for a given site fully traversed the pages in the site. A secondary strategy was to run the QMZ tool against an unknown web site in hopes of encountering new types of HTML, to detect defects in my QMZ tool based on what it could not traverse. This helped to get better branch and statement coverage by causing my QMZ tool to execute code not traversed by the smaller, known applications. However, achieving good statement and branch coverage was not explicitly part of my test strategy.

Web scraping an entire site by its nature is slow, since it makes a request and downloads a full web page, and it does this repeatedly. Because of this, it got very tiresome waiting for the same page to re-scrape after finding and fixing a small bug later in the QMZ tool. To speed up the testing process, I added the ability to serialize the results of the QMZ tool at different points, and save the serializations to files to be read in later. The QMZ tool allows the following to be serialized to YAML: Site instances, PFD instances. Using the YAML::dump and YAML::load methods, Ruby objects can be transformed into their YAML representations with no extra code added to the object's class, then deserialized later into the Ruby object. This allows the QMZ tool to accept input parameters specifying an input file from which either a Site or a PFD is represented in YAML. The user can also specify output files in which to save a Site or a PFD, once they have been constructed, to be loaded later. Then the entire web application need be scraped only once, which greatly speeds up development testing when neither the PFD nor Site class changes.

## 2c. Sample Output from the QMZ Tool

For the web site <http://babyhealthlexington.org/>, the following output was produced:

>ruby scraper.rb -u "http://babyhealthlexington.org/"

Got URI http://babyhealthlexington.org/

Getting pages for site at http://babyhealthlexington.org/...

Got 8 new Pages linked from Page / (13 links)

.Got 2 new Pages linked from Page /wp-admin/ (3 links)

...Got error '405 Method Not Allowed' trying to open URI http://babyhealthlexington.org/wp-comments-post.php, skipping...

Got 3 new Pages linked from Page /2009/01/21/hello-world/ (14 links)

..Got 1 new Page linked from Page /2009/01/21/hello-world/comment-page-1/ (14 links)

.Got 6 new Pages linked from Page /es/2009/01/21/hello-world/comment-page-1/ (14 links)

.............Got 21 pages for site at http://babyhealthlexington.org/

Pages in site rooted at http://babyhealthlexington.org/:

/

/wp-admin/

/2009/01/21/hello-world/

/category/updates/

/es/

/links/

/people/

/mission-statement/

/history/

/wp-login.php

/wp-login.php?action=lostpassword

/2009/01/21/hello-world/trackback/

/2009/01/21/hello-world/comment-page-1/

/es/2009/01/21/hello-world/

/es/2009/01/21/hello-world/comment-page-1/

/es/2009/01/21/hello-world/trackback/

/es/category/updates/

/es/links/

/es/people/

/es/mission-statement/

/es/history/

Getting PFD for site http://babyhealthlexington.org/...

Converting PFD to PTT...

Writing HTML file with test paths to http.babyhealthlexington.org/index.html...

File successfully written

The HTML file of test paths that was created looks like this:

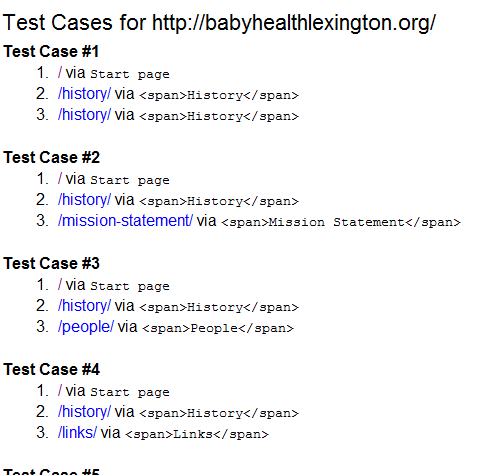


Figure Screenshot of HTML page produced by QMZ tool

In the figure above, test case #1 has the same URI twice, accessed via the same page flow component: /history/ via the <span>History</span> link. This is correct because the History page has the same global navigation as the rest of the web application, and it contains a link to itself. The QMZ model finds this link and includes accessing the History page via a link on the History page as part of a test case. This is part of ensuring test cases exist for each path through the web application.

Expected output of each test case is not determined by the QMZ tool, nor would it be possible for the tool to know what expected output should be. "Expected output" for a web page encompasses many things, including the exact appearance and performance of the web page in different browsers under different conditions. The same page can look very different when viewed in Microsoft Internet Explorer versus Mozilla Firefox, and it can perform differently based on the JavaScript engine in the web browser, plugins the user has enabled, and other settings the user has in the web browser. The human tester is expected to know what content should be displayed on a given web page, and whether the page's appearance and performance in Internet Explorer, Firefox, or another web browser is correct. It is possible the web page will not have the "correct" appearance or functionality in any browser, under any configuration.

In total, 185 test paths were produced for this web site with 21 pages. It is left to the user to determine which test paths to develop into full test cases (deciding input for forms encountered on the test path, which browsers to use, etc.).

# 3. Requirements for the Atomic Section Model Tool

* The ASM tool shall produce as output a Component Interaction Model for each provided Ruby on Rails ERB view file.
* The ASM tool shall accept as input one of the following: the path to a single Ruby on Rails view file, or the path to the root directory of a Ruby on Rails web application. The ASM tool shall also accept as input the URI to where the Ruby on Rails application will be hosted.
* The ASM tool shall provide a command-line interface.
* The ASM tool shall be able to parse an HTML ERB file.
* The ASM tool shall correctly identify atomic sections within an HTML ERB file.
* The ASM tool shall construct a correct component expression for a given HTML ERB file, using its previously determined atomic sections.
* The ASM tool shall correctly identify instances of selection, aggregation, iteration, and sequence [6] with an HTML ERB file.
* The ASM tool shall correctly identify if, elsif, else, unless, case, and when branches in valid Ruby code, including the ERB content contained within those branches.
* The ASM tool shall correctly identify nested branches in valid Ruby code.
* The ASM tool shall be able to parse valid Ruby code.
* The ASM tool shall identify hyperlink and form transitions within HTML when an a tag or form tag is encountered.
* The ASM tool shall identify a form transition in an ERB tag when a call to one of the following Rails methods is encountered: form\_tag, form\_remote\_tag, form\_for, remote\_form\_for, form\_remote\_for.
* The ASM tool shall identify a redirect transition in an ERB tag when a call to one of the following Rails methods is encountered: redirect\_to, redirect\_to\_full\_url.
* The ASM tool shall identify a hyperlink transition in an ERB output tag when a call to one of the following Rails methods is encountered: link\_to, link\_to\_remote, link\_to\_unless\_current, link\_to\_unless, link\_to\_if, link\_to\_function.
* The ASM tool shall identify iteration in an ERB tag when a call to one of the following methods is encountered: each, each\_with\_index, each\_cons, each\_entry, each\_slice, each\_with\_object, upto, downto, times.
* The ASM tool shall identify iteration in an ERB tag when one of the following key words is encountered: while, for, until.
* The ASM tool shall identify aggregation in an ERB tag when a call to one of the following methods is encountered: render.
* The ASM tool shall skip parsing of any file that it cannot read.
* The ASM tool shall be able to traverse the directory structure of a Ruby on Rails application and locate all HTML ERB view files when they are within the app/views directory.
* The ASM tool shall be able to output its parse of the given input file as well as any atomic sections it identified.
* The ASM tool shall only identify transitions to URIs within the web application.
* A URI shall be considered part of the web application if its host name matches the host name of the provided root URI of the Ruby on Rails application.

Some functionality that falls outside the scope of this project is as follows: the ASM tool does not embed the component expression for a Ruby on Rails partial view within the component expressions of other views that render that partial. The tool does not embed the component expression for a Ruby on Rails view within the component expression of the layout used with that view when the view is rendered by Rails in a web browser. The tool considers an ERB output tag as producing a single piece of output, and does not identify selection within that ERB output tag when generating the component expression for the HTML ERB file containing the tag.

## 3a. Implementing the ASM Tool

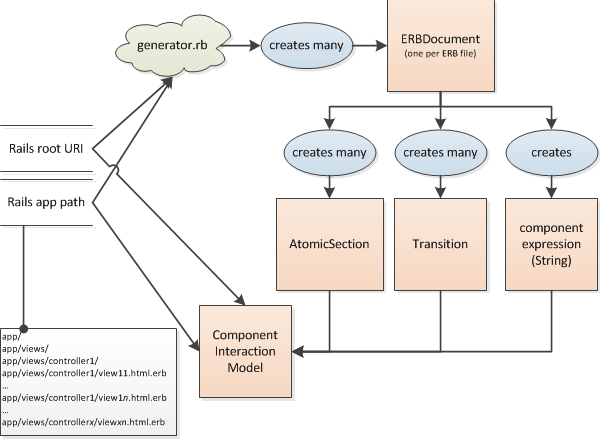


Figure 3 Overview of ASM tool

To implement the Atomic Section Model, I had to make a choice about the target language/framework. Once the decision was made to target Ruby on Rails, it was necessary to identify atomic sections within the ERB files that make up the output of a Rails web application. After atomic sections have been identified, component expressions must be identified. Transitions can also be identified after atomic sections have been found. The combination of atomic sections, transitions, and component expressions forms a Component Interaction Model for the page. A list of CIMs for the pages in the Ruby on Rails web application is the result of my ASM tool.

Component expressions represent the structure of a page in the Atomic Section Model. They list atomic sections in the order in which they appear in the page, and indicate whether the sections are in a sequence, selection, iteration, or aggregation. A sequence of sections p1 and p2 is represented in the output of my tool by p1.p2, and it indicates that atomic section p2 follows p1 in the page. A selection of sections p1 and p2 is represented by p1|p2, and indicates that either p1 or p2 will be displayed when the page is rendered in the browser, but not both. An iteration of section p1 is represented by p1\*, and indicates that section p1 will be displayed consecutively an arbitrary number of times. Finally, an aggregation of sections p1 and p2 is represented by p1{p2}, and indicates that p2 is included as part of p1.

The most difficult hurdle to overcome in the design of the ASM tool was how to generate a component expression for a given page. The tool can handle a combination of HTML, Ruby, and JavaScript in an ERB file, but it requires the Ruby to be valid. If a page in the web application being tested has invalid Ruby, the web application will generate an error when that invalid Ruby code is reached, so my tool is still fulfilling its purpose of helping identify errors in the application. Thus, it seems a reasonable requirement for the provided Ruby code to be valid.

The tool must be able to identify where an ERB tag or an ERB output tag occurs in the file, and treat those differently than HTML or JavaScript. To know where these ERB tags occur, a custom grammar was written for ERB files using the Treetop gem[[10]](#footnote-4). Treetop is described as "a language for describing languages" on its web site; it allows grammars to be written using Ruby code. My ERB grammar was built using parts of the RHTML grammar found in the Erector framework[[11]](#footnote-5) as well as parts of the complex\_html grammar[[12]](#footnote-6) provided as a sample for Treetop use.

Based on the code within an ERB tag, the component expression for the page may change. Consider the following ERB:

Some text

<% if condition %>

<p>Some optional text</p>

<% end %>

<strong>Final text</strong>

In this example, the atomic sections would be as follows:

|  |  |
| --- | --- |
| **Atomic Section** | **Contents** |
| **p1** | Some text |
| **p2** | <p>Some optional text</p> |
| **p3** | <strong>Final text</strong> |

The two ERB tags do not belong to any atomic section because they will not be displayed when this Rails page is rendered in the browser. Section p2 will only show up some of the time, since it is within the body of a Ruby if expression. We must therefore treat p2 differently than if it were within a Ruby while loop or within a Rails form\_tag block. The component expression for the above ERB would be p1.(p2|NULL).p3, meaning section p1 will appear in the browser, followed by either p2 or nothing at all, followed by p3.

So to determine component expressions, Ruby code must be parsed in addition to the HTML, XML, JavaScript, etc. that can appear in an ERB file. In order to parse Ruby code, the ruby\_parser gem[[13]](#footnote-7) was used. This gem allows strings of Ruby code to be parsed, and returns an s-expression when the provided Ruby code is valid. The s-expression is then analyzed to determine if the ERB indicates that contained atomic sections will be selected, iterated, or aggregated. Knowledge of both Ruby and the Ruby on Rails framework was necessary to identify when iteration would occur, for example. Built-in Ruby loop constructs such as for and while had to be identified, as well as methods such as each, each\_with\_index, and downto. One shortcoming is that my tool cannot detect when iteration will occur within a user-defined method. Consider the following Ruby code:

<%

def looper

1.upto(3) { |i| yield i }

end

looper do |j| %>

This is iteration <%= j %>

<% end %>

Here, the s-expression for the looper do |j| block does not indicate that the call to the looper method iterates. The component expression for this ERB comes back as p1, and not p1\*. It is not sufficient to simply treat all blocks as iterations since, for example, the Rails method form\_tag[[14]](#footnote-8) takes a block but does not iterate.

The two main scripts for the ASM tool are generator.rb and single\_file\_generator.rb. The former expects as input the path to the root directory of a Ruby on Rails application, while the latter expects the path to a single ERB file. They also each take a URI to the root page of the web application if the application is hosted online. This is so that a guess can be made as to the URI to the page, based on known Rails URI patterns.

The single\_file\_generator.rb script first reads the file at the given path and passes it to the Parser class. This class uses the Treetop ERB grammar to attempt to parse the given ERB code. It also performs much of the cleanup necessary to construct a CIM for the file, such as identifying atomic sections and nesting code. Parser constructs a class ERBGrammarParser by making a call to Treetop.load and providing the Treetop ERB grammar file. This ERBGrammarParser class is used to parse the provided ERB code, and it returns the root node of the results of its parse: an instance of my ERBDocument class. ERBDocument inherits from Treetop's Treetop::Runtime::SyntaxNode class, as do the other node types represented in my grammar. Most of the work of my ASM tool is done within these classes. Because Ruby has open classes [[[15]](#endnote-8)], I was also able to add extra methods and properties to the Treetop::Runtime::SyntaxNode for use in inherited classes.

It is necessary to identify blocks of code so as to know when an atomic section is within, for example, an if block. Lines of ERB code must be parsed to see if they form valid Ruby code when concatenated. The RubyParser class will raise an exception when invalid code is passed, so it is possible to successively parse lines of Ruby code found in ERB tags to determine when a full block has been found. Then, any code found between the first line of the block and the last line of the block are considered the children of that block. Consider the following:

<% 1 + 1 %>

<p>Text</p>

<% while condition %>

<%= getValue() %>

<strong>Sample</strong>

<% end %>

In this example, the first ERB tag containing Ruby code 1 + 1 would be valid code by itself. It is a single line so it has no children. In this case, my tool would move on to the next ERB tag, while condition. When this code is parsed, RubyParser would raise an exception because this by itself is not valid Ruby. When the next ERB tag's content is appended, however, the while condition with end are valid Ruby. Then my tool knows that the ERB output tag getValue() and the HTML <strong>Sample</strong> are nested within that while loop.

Some code blocks contain others, such as in the following example:

|  |  |
| --- | --- |
| 1 | <% if outer\_condition %> |
| 2 | <% for i in 1..5 do %> |
| 3 | Test |
| 4 | <% end %> |
| 5 | <% else %> |
| 6 | Value |
| 7 | <% end %> |

Here my tool would have to parse all ERB tags concatenated together before RubyParser would accept the input as valid code. After that, the contents of the code block would be lines 2 through 6, inclusive. That is, the RubyParser does not identify lines 2-4, inclusive as being in the true condition of an if statement, nor does it identify line 6 as being in the false condition of an if statement. This must be handled later by my tool via the split\_branches method.

After code blocks have been found, atomic sections can be identified. This is done by iterating over all the nodes in the ERBDocument and appending content to an instance of the AtomicSection class. When an ERB tag is encountered, or when a call to the Rails method render is encountered within an ERB output tag, a new atomic section is started. A new section must be started for a render because that is an instance of aggregation, where the contents returned by render will be embedded in the page. Atomic sections are retained within a list in ERB nodes such as the ERBDocument and ERBTag classes.

After atomic sections are identified, they are nested within code blocks. An index is used to indicate where in the original ERB document a node was found, and these indices are used to determine the parent of an atomic section. An atomic section's range is the range of indices of its contents. Atomic sections are identified at the top-most level, the ERBDocument, and then moved to be contained in the separate atomic section lists that ERBTag nodes maintain. This is useful later when determining component expressions, to see if an atomic section is within, for example, an ERB tag with a loop. If an atomic section's range is indices 3-5, for example, and an ERBTag has range 2-6, the atomic section will be nested within that ERBTag.

Once atomic sections have been nested, branches can be split in ERBTag nodes. This is necessary for identifying selections in component expressions. For each node that looks like a selection (for example, its s-expression has an if or a case statement), the split\_branch method is called on it. This method determines which of the node's contents belongs in which branch by analyzing s-expressions.

The method first checks that the node on which it is operating looks like a selection by calling the selection? method. By convention, the names of Ruby predicates (methods that return a Boolean value) end in a question mark. The selection? method checks the node's s-expression for :if, :case, or :when--Ruby symbols which indicate a branch. Ruby's unless keyword is represented by :if in an s-expression. If the node is not a branch, the split\_branch method returns without doing anything.

The direct ERBTag children of the node are considered, and split\_branch is called recursively on each of these first. This is to split nested branches, such as an if statement within an if statement, first. The node's list of branch content, maintained as a field within the class, is initialized to an empty array. This array will eventually be an array of arrays, where each sub-array represents a branch in the node. For example, if the node's Ruby source were if condition1 then blah; elsif condition2 then bleh; else bloh; end, then its branches would be [[blah], [bleh], [bloh]].

If the node has no ERBTag children, the atomic sections in the node are the node's only branch, so they are appended to the list of branches and the method returns. If there are children, the method next identifies pivot points in the node. Consider the following Ruby:

|  |  |
| --- | --- |
| 1 | case some\_variable |
| 2 | when 5 |
| 3 | value1 |
| 4 | when 4 |
| 5 | value2 |
| 6 | else |
| 7 | value3 |
| 8 | end |

Here, the case some\_variable would be the node being considered, and all other lines would be content within that node, by virtue of previous code block identification. The pivots in this code would be lines 2, 4, and 6; these are identified by selecting all the ERBTag children of the node whose s-expression is :invalid\_ruby,meaning that the child's Ruby source could not be parsed by itself.

The content of the first branch consists of all atomic sections and child ERBTag instances within the node whose index is between the node's index and the first pivot's index. In the Ruby code above, there would be no such first branch because there are no children between lines 1 and 2. However, with an if statement, the first branch would consist of children between the if condition and the first elsif or else. This first branch content is appended to the node's list of branch content.

Next, a prev\_pivot variable is initialized depending on the first pivot child: if the first pivot is invalid Ruby, such as line 2 above, then prev\_pivot is initialized to that first pivot. Otherwise, prev\_pivot is initialized to the current node itself. The list of pivots is then iterated over, and branch content is determined for the different branches. As before, a branch's ERBTag and atomic section content is selected by comparing the index against the prev\_pivot's index and the index of the current pivot in the iteration. If prev\_pivot is not the current node, then the branch content is moved such that its parent is no longer the current node, but rather prev\_pivot is the new parent node. Then an array containing only the prev\_pivot node is appended to the current node's list of branch content. The prev\_pivot variable is set to be the current pivot, and the next iteration of the loop begins.

This pivot loop over the Ruby code above would proceed as follows:

prev\_pivot = when 5

// Loop iteration 1:

cur\_branch\_content = [value1]

move value1 to be a child of when 5 instead of case some\_variable

append [prev\_pivot] to branch\_content for case some\_variable

prev\_pivot = when 4

// Loop iteration 2:

cur\_branch\_content = [value2]

move value2 to be a child of when 4 instead of case some\_variable

append [prev\_pivot] to branch\_content for case some\_variable

prev\_pivot = else

After this loop completes, there may still be a remaining branch between the last pivot and the close of the current node. In the sample Ruby above, this last branch consists of value3 between the else and the end. If the current node has a close, content from the current node is selected where the content's index is between the last pivot's index and the closing node's index. This content is moved from being contained in the current node to being contained in the last pivot node. Then, the last branch's content is appended to the list of branch content for the current node.

This technique for branch splitting correctly identifies nested branches. Consider the following ERB:

<% case var %>

<% when 5 %>

<% case var2 %>

<% when 6 %>

It's six!

<% when 6.5 %>

It's six and a half!

<% else %>

It's 6.75!

<% end %>

<% when 7 %>

<% if condition1 %>

<html>

<% elsif condition2 %>

<%= some\_var %>

<%= method\_call() %>

<B>A bold statement!</B>

<% else %>

<p>Paragraph</p>

<% end %>

<% else %>

It's neither 5 nor 7!

<% end %>

The following output from the single\_file\_generator.rb script shows how the ASM tool understands the above ERB code:

>ruby single\_file\_generator.rb tests\fixtures\nested\_case\_when.html.erb "http://example.com"

Parsing ERB file tests\fixtures\nested\_case\_when.html.erb...

Could not interpret path tests\fixtures\nested\_case\_when.html.erb as a Rails URL, skipping transition identification

Source file: tests\fixtures\nested\_case\_when.html.erb

0-26: case var

Content and sections:

1: when 5

Content and sections:

2-9: case var2

Content and sections:

3: when 6

Content and sections:

4: Atomic Section #1 (indices 4..4):

4: It\'s six!

5: when 6.5

Content and sections:

6: Atomic Section #2 (indices 6..6):

6: It\'s six and a half!

7: else

Content and sections:

8: Atomic Section #3 (indices 8..8):

8: It\'s 6.75!

Close:

9: end

10: when 7

Content and sections:

11-23: if condition1

Content and sections:

12: Atomic Section #4 (indices 12..12):

12: html

13: elsif condition2

Content and sections:

14: Atomic Section #5 (indices 14..18):

14: <%= some\_var

15: <%= method\_call()

16: b

17: A bold statement!

18: /b

19: else

Content and sections:

20: Atomic Section #6 (indices 20..22):

20: p

21: Paragraph

22: /p

Close:

23: end

24: else

Content and sections:

25: Atomic Section #7 (indices 25..25):

25: It\'s neither 5 nor 7!

Close:

26: end

Component Interaction Model

Start page: tests\fixtures\nested\_case\_when.html.erb

Start URL: http://example.com/UNKNOWN/nested\_case\_when

Component expression: ((p1|p2|p3)|(p4|p5|p6)|p7)

Transitions:

In the output above, indentation is very important and indicates the tool's understanding of code nesting. Three outer branches are correctly identified, and the nested inner branches are also parsed correctly.

After branches have been split, transitions can be identified. This is done using the same code that the QMZ model uses, i.e., with Nokogiri. There are different node classes in the ASM model that can have transitions, such as an ERBTag with a form\_tag call or an HTMLOpenTag with a hyperlink. There are several modules in my ASM tool that contain code to be shared between classes, including instance methods and class methods. This uses the mixin feature[[16]](#footnote-9) of Ruby to share members without multiple inheritance (Ruby uses single inheritance). Since multiple classes need access to the HTML parsing methods that use Nokogiri, these are placed in a SharedHTMLParsing module. They could not be placed in a base class because the node classes in the ASM tool must inherit from Treetop::Runtime::SyntaxNode.

Node classes that can possibly have transitions have a get\_local\_transitions method that differs by class. Each implementation takes a source parameter, though, intended to contain the URI of the page where the hyperlink or form transition is located. It is the job of the get\_local\_transitions method to identify any transitions within its own code, which the Treetop::Runtime::SyntaxNode retains, and return Transition instances to represent these. The Transition class has three children: FormTransition, RedirectTransition, and LinkTransition. These three classes correspond to the form link transition, redirect transition, and simple link transition in the Atomic Section Model. A RedirectTransition is created when, for example, an ERBTag has a call to the Rails method redirect\_to, which would cause the browser to redirect to another page.

The get\_local\_transitions method defined in ERBTag looks for form and redirect transitions by looking at the s-expression of the code of that ERBTag. It does not need to look for link transitions because even if a Rails method such as link\_to were used in an ERB tag to produce a hyperlink, the result of that code would not be included in the final web page because it is in an ERB tag and not an ERB output tag.

To find local form transitions in an ERBTag, the names of known <form>-producing methods are iterated over, and the s-expression is checked for each to see if a call is made to one of them. One improvement that could be made in future work is to move this list of methods to an external file that could be read at runtime. This would allow greater flexibility in the event that Rails changes or adds methods that affect transition generation. If a call is detected within the s-expression, the arguments to the call are extracted from the s-expression and used to determine the destination of the transition. For a form transition, the destination is the contents of the action attribute in HTML. The s-expression is checked for calls to the Rails method url\_for, as well as the assignment of values to keys controller and action in a hash. Once the destination of the form transition has been determined, a new instance of the FormTransition class can be constructed.

Other transitions are identified differently based on the node class in which they are identified. For example, the HTMLOpenTag class uses Nokogiri to parse its HTML source to find link and form transitions. The ERBOutputTag class checks the s-expression for its Ruby source code to find calls to Rails methods such as link\_to and link\_to\_unless\_current.

Once transitions have been identified, a new instance of the ComponentInteractionModel class can be constructed with the atomic sections, component expression, and list of transitions for that ERB file. The final output of my ASM tool is a single CIM or a list of CIMs, depending on whether the single\_file\_generator.rb or the generator.rb script was used.

## 3b. Testing the ASM Tool

An automated test suite was developed using Ruby's Test::Unit[[17]](#footnote-10). Most unit tests are designed to check that the component expression produced for a given ERB input file matches the expected, manually determined component expression. Since generating a component expression is one of the last steps of the ASM tool and it is a single string, it is a good baseline to ensure the tool is functioning correctly. The component expression summarizes how the ERB parser and the tool's use of RubyParser interpreted the given input, so if the component expression is wrong, it can indicate previous steps, such as atomic section identification or code nesting, went awry.

All tests are runnable at once via executing the tests/runner.rb script. There are different test classes that contain unit tests for the different node classes, such as ERBDocument and ERBTag. Particular ERB files are stored as test fixtures in the project, and used by the different tests. These ERB files represent code that previously caused errors in the ASM tool, or that did not have their component expressions correctly generated. A sample unit test follows:

def test\_multiple\_erb\_lines\_unequal\_ifs\_component\_expression

assert\_component\_expression(fixture('\_in\_progress.html'),

'\_in\_progress.html.erb',

'((((p1|NULL).p2)|p3)|NULL).p4.p5.p6\*.p7.p8\*.p9

.p10.p11.p12.(p13|p14)')

end

This reads the contents of the fixture with file name prefix \_in\_progress.html and passes it to the assert\_component\_expression helper method:

def assert\_component\_expression(erb, file\_name, expected)

doc = Parser.new.parse(erb, file\_name, URI.parse('/'))

assert\_not\_nil doc

actual = doc.component\_expression()

assert\_equal expected, actual,

"Wrong component expression for " + file\_name

end

## 3c. Sample Output from the ASM Tool

### Single File Output

Consider the following ERB:

<div id="game\_area">

<% #Write out a hidden AJAX (remote) form to submit cards.

form\_remote\_tag(:url => url\_for(:action => "handle\_card\_submit"), :html => {:id => 'cardform'}) do %>

<%= hidden\_field\_tag 'form', 'card\_submission' %>

<% #Write out the cards for the user to see, along with hidden checkboxes.

@cards.each\_with\_index do |card, count| %>

<%= display\_card(card, session[:user][:color\_blind]) %>

<%= check\_box\_tag('cards[]', card.id, false, :id => "card#{card.id}", :class => 'display\_none') %>

<%

end

end %>

</div><!--game\_area-->

This ERB will produce an HTML form tag that has an AJAX call to the handle\_card\_submit action (meaning the handle\_card\_submit() method in the current Rails controller class, which equates to the handle\_card\_submit/ page). Within that form, there is a hidden form tag as well as a loop over the @cards array. Within the loop, there are two ERB output tags. Wrapped around the entire form is an HTML div tag. Using the single\_file\_generator.rb script, here is the output:

>ruby single\_file\_generator.rb app\vie

ws\test\cards.html.erb "http://example.com"

Parsing ERB file app\views\test\cards.html.erb...

Source file: app\views\test\cards.html.erb

0: Atomic Section #1 (indices 0..0):

0: div id => "game\_area"

1: #Write out a hidden AJAX (remote) form to submit cards.

2-9: form\_remote\_tag(:url => url\_for(:action => "handle\_card\_submit"), :html

=> {:id => 'cardform'}) do

Content and sections:

3: Atomic Section #2 (indices 3..3):

3: <%= hidden\_field\_tag 'form', 'card\_submission'

4: #Write out the cards for the user to see, along with hidden checkboxes.

5-8: @cards.each\_with\_index do |card, count|

Content and sections:

6: Atomic Section #3 (indices 6..7):

6: <%= display\_card(card, session[:user][:color\_blind])

7: <%= check\_box\_tag('cards[]', card.id, false, :id => "card#{card.id}

", :class => 'display\_none')

Close:

8: end

Close:

9: end

10: Atomic Section #4 (indices 10..11):

10: /div

11: <!--game\_area-->

Component Interaction Model

Start page: app\views\test\cards.html.erb

Start URL: http://example.com/test/cards

Component expression: p1.p2.p3\*.p4

Transitions:

Form Transition

</test/cards> --> </test/handle\_card\_submit>

Underlying code: form\_remote\_tag(:url => url\_for(:ac

tion => "handle\_card\_submit"), :html => {:id => 'cardform'}) do

The single\_file\_generator.rb script shows its parse of the input ERB, which is useful for debugging. Even though the opening div, the results of the form\_remote\_tag call, and the hidden\_field\_tag call will all be displayed together in the resulting HTML when this page is viewed in the browser, the tool does not know that. So the tool treats the form\_remote\_tag ERB tag as a separator between two atomic sections. It would be possible to maintain a list of Rails methods that do not result in conditional or iterative output, such as form\_remote\_tag.Then the tool could be modified to recognize that a form\_remote\_tag ERB tag does not immediately signify the start of a new atomic section;however, that was outside the scope of this project.

The ERB code above also shows another intricacy of the Ruby on Rails framework: some ERB tags, when interpreted by Rails, can produce output. Namely, the form\_remote\_tag method produces an HTML form tag with different attributes depending on the parameters[[18]](#footnote-11). My tool treats all ERB tags as non-output producing, since generally this is the case. The tool does, however, scan ERB tags for form transitions, as evidenced in the output above that shows that a single form transition was found in the ERB.

### Rails Application Output

The following is the result produced by running the generator.rb script against the root directory of a Ruby on Rails application:

>ruby generator.rb ..\set\_game "http://cs499-game.3till7.net"

Looking in directory ../set\_game/app/views

Looking in directory ../set\_game/app/views/login

Parsing ERB file ../set\_game/app/views/login/register.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/login/register.html.erb

Start URL: http://cs499-game.3till7.net/login/register

Component expression: p1

Transitions:

Form Transition

<http://cs499-game.3till7.net/login/register> --> </login/try\_register>

Underlying code:

form\_tag :action => 'try\_register' do

Parsing ERB file ../set\_game/app/views/login/index.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/login/index.html.erb

Start URL: http://cs499-game.3till7.net/login/index

Component expression: p1

Transitions:

Form Transition

<http://cs499-game.3till7.net/login/index> --> </login/try\_login>

Underlying code:

form\_tag :action => "try\_login" do

Link Transition

<http://cs499-game.3till7.net/login/index> --> </login/register>

Underlying code:

link\_to( image\_tag("transparent.png", {:alt => 'Register'}),

{:controller => 'login', :action => 'register'}, :id => 'register\_button')

Looking in directory ../set\_game/app/views/layouts

Parsing ERB file ../set\_game/app/views/layouts/application.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/layouts/application.html.erb

Start URL: http://cs499-game.3till7.net/layouts/application

Component expression: p1.((p2.(p3|p4))|p5).p6.(p7|NULL).(p8|NULL).(p9|NULL).p10

Transitions:

Link Transition

<http://cs499-game.3till7.net/layouts/application> --> </game/top\_records>

Underlying code:

link\_to '<span>Top 20</span>', :controller => 'game', :action => 'top\_records'

Link Transition

<http://cs499-game.3till7.net/layouts/application> --> </login/logout>

Underlying code:

link\_to '<span>Logout</span>', :controller => 'login', :action => 'logout'

Link Transition

<http://cs499-game.3till7.net/layouts/application> --> </game/rules>

Underlying code:

link\_to '<span>Rules</span>', :controller => 'game', :action => 'rules'

Link Transition

<http://cs499-game.3till7.net/layouts/application> --> </login/admin>

Underlying code:

link\_to '<span>Admin</span>', :controller => 'login', :action => 'admin'

Looking in directory ../set\_game/app/views/game

Parsing ERB file ../set\_game/app/views/game/\_waiting.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/\_waiting.html.erb

Start URL: http://cs499-game.3till7.net/game/\_waiting

Component expression: p1.p2

Transitions:

Parsing ERB file ../set\_game/app/views/game/\_top\_list.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/\_top\_list.html.erb

Start URL: http://cs499-game.3till7.net/game/\_top\_list

Component expression: p1

Transitions:

Parsing ERB file ../set\_game/app/views/game/\_poll\_for\_updates.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/\_poll\_for\_updates.html.erb

Start URL: http://cs499-game.3till7.net/game/\_poll\_for\_updates

Component expression: {p1}

Transitions:

Parsing ERB file ../set\_game/app/views/game/\_in\_progress.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/\_in\_progress.html.erb

Start URL: http://cs499-game.3till7.net/game/\_in\_progress

Component expression: ((((p1|NULL).p2)|p3)|NULL).p4.p5.p6\*.p7.p8\*.p9.p10.p11.p12.(p13|p14)

Transitions:

Form Transition

<http://cs499-game.3till7.net/game/\_in\_progress> --> </game/handle\_card\_submit>

Underlying code:

form\_remote\_tag(:url => url\_for(:action => "handle\_card\_submit"),

:html => {:id => 'cardform'}) do

Form Transition

<http://cs499-game.3till7.net/game/\_in\_progress> --> </game/handle\_call\_set>

Underlying code:

form\_remote\_tag(:url => url\_for(:action => "handle\_call\_set"),

:loading => "$('systemWorking').show();") do

Form Transition

<http://cs499-game.3till7.net/game/\_in\_progress> --> </game/handle\_no\_sets

>

Underlying code:

form\_remote\_tag(:url => url\_for(:action => "handle\_no\_sets"),

:loading => "$('systemWorking').show();") do

Parsing ERB file ../set\_game/app/views/game/\_finished.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/\_finished.html.erb

Start URL: http://cs499-game.3till7.net/game/\_finished

Component expression: ((p1|p2)|NULL).(p3|NULL).(p4|NULL).p5.p6\*.p7

Transitions:

Parsing ERB file ../set\_game/app/views/game/top\_records.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/top\_records.html.erb

Start URL: http://cs499-game.3till7.net/game/top\_records

Component expression: p1.(p2|(p3.{p4}.p5)).p6.(p7|(p8.{p9}.p10)).p11

Transitions:

Parsing ERB file ../set\_game/app/views/game/rules.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/rules.html.erb

Start URL: http://cs499-game.3till7.net/game/rules

Component expression: p1

Transitions:

Parsing ERB file ../set\_game/app/views/game/join.html.erb...

Component Interaction Model

Start page: ..\set\_game/app/views/game/join.html.erb

Start URL: http://cs499-game.3till7.net/game/join

Component expression: p1.{p2}.p3

Transitions:

Parsing ERB file ../set\_game/app/views/game/index.html.erb...

Component Interaction Model

Start page: ../set\_game/app/views/game/index.html.erb

Start URL: http://cs499-game.3till7.net/game/index

Component expression: p1.p2.p3.(p4|NULL).p5.((p6.(p7.(p8|(p9.p10\*.p11)).

p12)\*.p13)|NULL)

Transitions:

Form Transition

<http://cs499-game.3till7.net/game/index> --> </game/new>

Underlying code:

form\_tag :action => "new" do

Link Transition

<http://cs499-game.3till7.net/game/index> --> <//join>

Underlying code:

link\_to game.name, :action => 'join', :game\_id => game.id

This shows a variety of page structures using form and link transitions, and component expressions with selection, aggregation, and iteration. Some pages were very simple, with only a single atomic section and no transitions. Other pages had much more complicated component expressions and several transitions.

# 4. Code Statistics

The ASM tool consists of 2,699 lines of code and 241 lines of comments across 47 files. The QMZ tool consists of 590 lines of code and 35 lines of comments across 8 files. See Appendix A for the source code of my tools.

# 5. Future Work

## 5a. QMZ Tool

The QMZ tool can be expanded to take credentials and other parameters for forms, so that when it encounters a form in a page it has scraped, it can populate the form and submit it. This would allow the tool to traverse the web site as much as possible when unauthenticated, for example, and then traverse the web site again after it has submitted a login form and authenticated. The PTT could be modified such that a node indicates whether any form on the page was submitted first, so that one branch of the PTT might be the unauthenticated traversal of the web site, and another branch might be the authenticated traversal. This would greatly increase the usefulness of the QMZ tool since so many web applications require some form of authentication to access the full contents. If the particular forms and their desired input were known in advance, these could be provided to the QMZ tool initially, and it could submit the appropriate parameters to the specified forms. It might also be possible to make the QMZ tool interactive such that if it found a form within a scraped page, it would prompt the user to get parameters. Form submission could be handled by the Ruby gem Mechanize[[19]](#footnote-12) or something similar. The Mechanize gem can also maintain state via cookies, which many web sites use to differentiate an authenticated user from an unauthenticated user.

The QMZ might also be modifiable to automate part of the testing process. Using a Ruby gem like Selenium[[20]](#footnote-13), it would be possible to take screenshots of web pages. Once the PTT is constructed, test paths could be followed and a screenshot could be taken of each successive page in the path. These screenshots could then be presented to the user in the HTML page of test paths, alongside the associated URI in the test path. It would be necessary to take a different screenshot for each URI even if the same URI appeared in several test paths, since the page might change depending on the referring page, or on the transition used to reach the page. While having a screenshot alone might not be sufficient for all test cases, it would be helpful when the test purpose is to ensure correct appearance and content.

The combination of screenshot and form support in the QMZ tool could be very helpful in terms of automating test execution. A user would still be necessary to manually examine each screenshot, and to provide form parameters. The tool would go through the steps of entering parameters into forms, traveling to each page in the test path, and taking screenshots along the way of the results of following each transition.

## 5b. ASM Tool

A few changes could be made to how component expressions are generated, based on how Ruby on Rails applications generally use views. If a view called application.html.erb exists in the app/views/layouts/ directory, it is used as the main layout for the entire web application. This can be overridden in individual controller classes, and the layout file does not necessarily have the .html.erb extension. It would be helpful to be able to specify that a particular file is used as an application- or controller-wide layout, and then combine its component expression with the component expressions of other pages that use that layout. Consider a page with component expression (p1|p2). Consider a layout used on this page with component expression p1.p2.p3, and a call to yield in an ERB output tag between sections p2 and p3. When Rails shows the page, its contents will be shown where the yield statement is within the layout. The ASM tool could be modified to take this into account, such that the component expression for the page would be p1.p2.(p3|p4).p5, where p1 and p2 are the same sections from the layout, p3 and p4 are the sections from the page, and p5 is the final section from the layout.

Partials are another concept in Ruby on Rails with the idea that a small chunk of a page can be included within another page. For example, a single line of code displaying a user name could be its own partial, and a page might include this partial within a loop. Then when the page is rendered, a list of user names appears because the partial is used many times as the page loop iterates. Partials appear in Rails code within ERB output tags, via calls to the render method. The ASM tool currently notices such calls and treats them as aggregation, in the terminology of the Atomic Section Model, but it does not follow up the partial and try to embed the partial's component expression within the parent page's component expression. Likewise, the ASM tool currently does not try to embed a page's component expression within the component expression of the page's layout.

Currently, the ASM tool considers the output of an ERB output tag to be a single atomic section, or part of another atomic section if it follows HTML content or another ERB output tag. A finer-grained approach could be used to check for selection within an ERB output tag and possibly identify more than one atomic section as the result of the ERB output tag. Consider the following ERB output tag: <%= 'some content' if condition %>. The current ASM tool would produce the component expression p1 for this code. A more accurate component expression would be (p1|NULL), since the atomic section is really the string 'some content' and not the entire ERB output tag, and this string can either appear or not.

The ASM tool can be modified to combine a collection of CIMs into an Application Transition Graph for the entire web application. Once the ATG is constructed, different coverage methods could be applied to generate test paths that traverse the entire web application, such as prime path coverage [] or invalid path coverage []. These test paths could be presented in the same way as the QMZ model's test paths, and could even be merged and displayed in a single page.

Further work could be done to the ASM tool to make it more flexible to changes in the Ruby on Rails framework, such as pulling lists of key methods out of node classes and into external files. A YAML file would be easily parsable in Ruby and could contain method names know for generating link, form, and redirect transitions. The different ERB node classes could refer to this YAML file when identifying transitions.

A combination of the QMZ and ASM tools would be helpful. Once an ATG is constructed in the ASM tool and test paths are found through some form of path coverage, a separate tool could be used to combine test paths from the QMZ and ASM tools. This new tool could actually follow the test paths, using gems like the aforementioned Selenium and Mechanize to take screenshots and submit forms. An HTML page of results could be presented to the user, showing the test paths that were followed and screenshots of each page in each test path. The parameters submitted to each form could also be displayed. It might be useful to allow multiple sets of parameters to be used for a single form, in order to simulate different test cases.

# 6. Conclusions

Through the QMZ tool, it is possible to find paths a user might take through your application via the transitions you have provided. Through the ASM tool, it is possible to generate artifacts which can be used to find paths that show engage every possible chunk of code that could be output to the user in their browser. The combination of these two tools can provide a developer with a good overview of what to test. Previously, applying these models involved much manual generation of artifacts. For the manual application of the QMZ model, links and forms needed to be manually found in each web page within the web application. Then, a tree had to be constructed of these transition elements, and then traced to find possible test paths. To manually generate CIMs for the Atomic Section Model for a Ruby on Rails application, each view file in the source code had to be analyzed manually to identify atomic sections. Once atomic sections were labeled, one would have to manually determine the component expression of that page, as well as a list of transitions found within the page.

My contribution is these tools that remove much of the manual work from finding good test paths for a web application. The tools will be released under an open source license to the public, so others may use them freely to help test their code. It is desirable to have correctly functioning web applications since accomplishing even daily tasks depends more and more on the Internet. With the help of my tools, testing web applications will be faster and simpler. This convenience will hopefully lead to well-tested and better performing web applications.

# 7. Bibliography

1. Wikipedia contributors, "ERuby," *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=ERuby&oldid=417473164 (accessed March 17, 2011). [↑](#endnote-ref-1)
2. <http://rubygems.org/> [↑](#footnote-ref-1)
3. Wikipedia contributors, "HTML," *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=HTML&oldid=419043476 (accessed March 17, 2011). [↑](#endnote-ref-2)
4. Wikipedia contributors, "S-expression," *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=S-expression&oldid=418259285 (accessed March 17, 2011). [↑](#endnote-ref-4)
5. Ryan Davis, "parsetree's ruby\_parser-2.0.6 Documentation," *parse\_tree & ruby\_parser*, http://parsetree.rubyforge.org/ruby\_parser/ (accessed March 17, 2011). [↑](#endnote-ref-5)
6. Offutt, Jeff and Wu, Ye, “Modeling Presentation Layers of Web Applications for Testing”, Springer’s Software and Modeling, DOI: 10.1007/s10270-009-0125-4. [↑](#endnote-ref-6)
7. Qian, Miao, and Zeng, “A Practical Web Testing Model for Web Application Testing”, Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, DOI 10.1109/SITIS.2007.16. [↑](#endnote-ref-7)
8. <http://rubyonrails.org/applications> [↑](#footnote-ref-2)
9. <http://nokogiri.org/> [↑](#footnote-ref-3)
10. <http://treetop.rubyforge.org/> [↑](#footnote-ref-4)
11. <http://erector.rubyforge.org/> [↑](#footnote-ref-5)
12. <http://github.com/threedaymonk/treetop-example/blob/master/complex_html.treetop> [↑](#footnote-ref-6)
13. <http://rubyforge.org/projects/parsetree/> [↑](#footnote-ref-7)
14. <http://api.rubyonrails.org/classes/ActionView/Helpers/FormTagHelper.html#method-i-form_tag> [↑](#footnote-ref-8)
15. Ruby contributors, "Ruby from Other Languages", *Ruby A Programmer's Best Friend,* http://www.ruby-lang.org/en/documentation/ruby-from-other-languages/ (accessed March 17, 2011). [↑](#endnote-ref-8)
16. <http://www.ruby-doc.org/docs/ProgrammingRuby/html/tut_modules.html> [↑](#footnote-ref-9)
17. <http://ruby-doc.org/stdlib/libdoc/test/unit/rdoc/classes/Test/Unit.html> [↑](#footnote-ref-10)
18. <http://apidock.com/rails/ActionView/Helpers/PrototypeHelper/form_remote_tag> [↑](#footnote-ref-11)
19. <http://mechanize.rubyforge.org/mechanize/> [↑](#footnote-ref-12)
20. <http://selenium.rubyforge.org/> [↑](#footnote-ref-13)