Unravel

Rapid Web Application Reverse Engineering via Interaction Recording, Source Tracing, and Library Detection

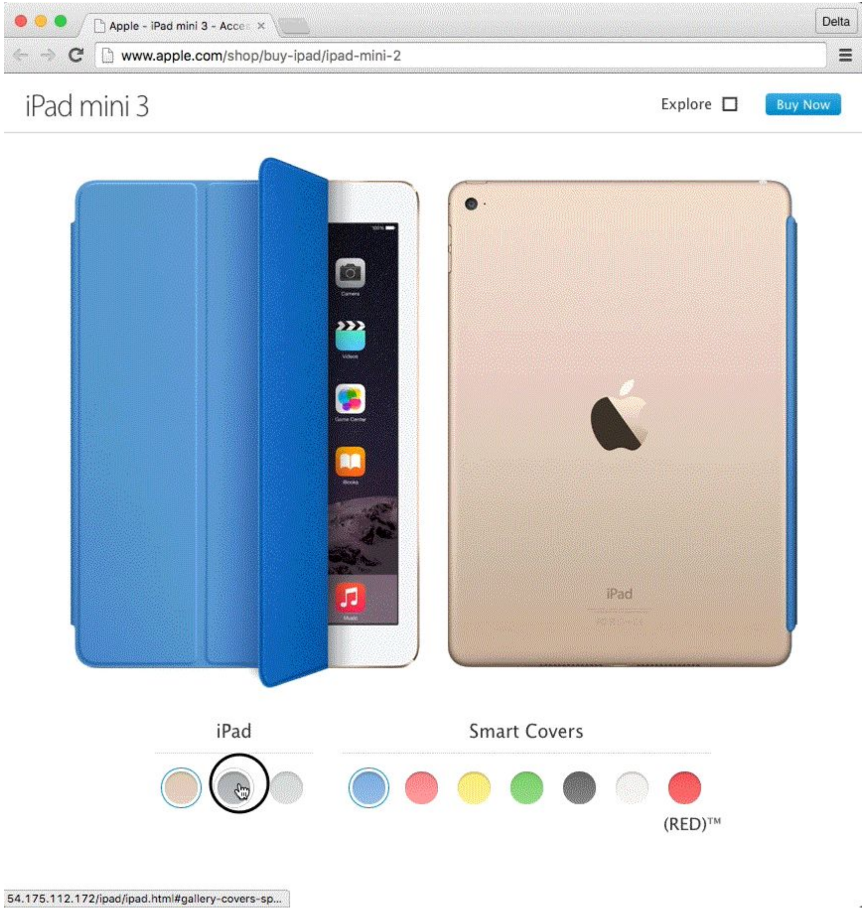
Joshua Hibschman

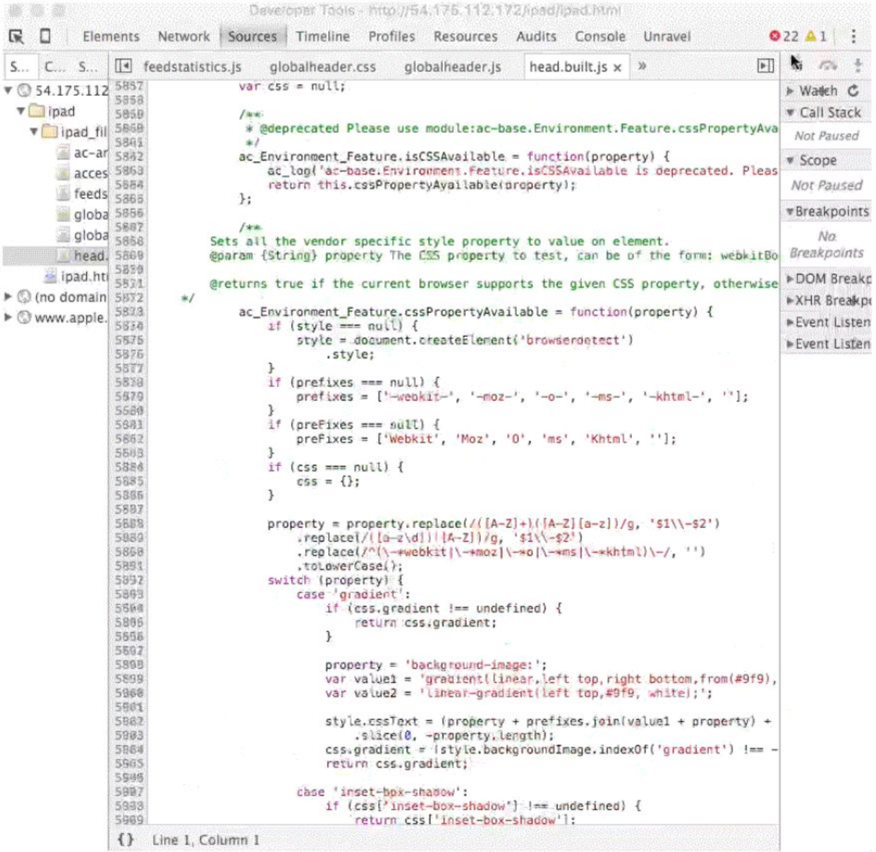
Haoqi Zhang

NORTHWESTERN UNIVERSITY

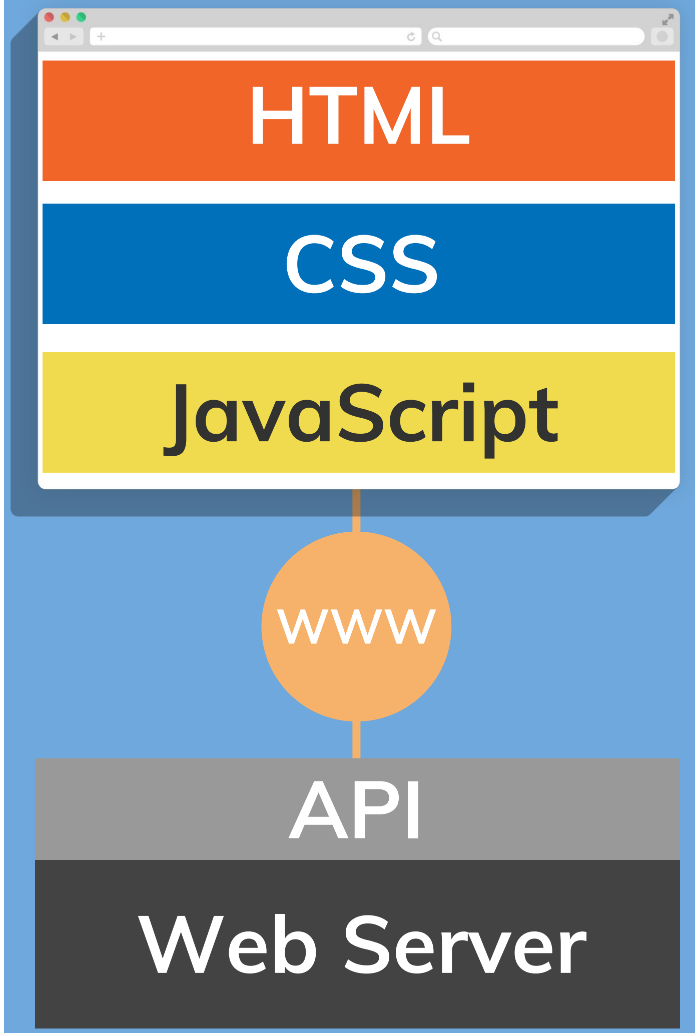
DELTA LAB

Inspection is Overwhelming


Apple - iPad mini 3 - Acce
Delta
www.apple.com/shop/buy-ipad/ipad-mini-2
iPad mini 3
Explore
Buy Now
iPad
iPad
Smart Covers
(RED)™
54.175.112.172/ipad/ipad.html#gallery-covers-sp...



```
        var css = null;

        /**
         * @deprecated Please use module:ac-base.Environment.Feature.cssPropertyAva
         */
        ac_Environment_Feature.isCSSAvailable = function(property) {
            ac_log('ac-base.Environment.Feature.isCSSAvailable is deprecated. Pleas
            return this.cssPropertyAvailable(property);
        };

        /**
    Sets all the vendor specific style property to value on element.
    @param {String} property The CSS property to test, can be of the form: webkitBo

    @returns true if the current browser supports the given CSS property, otherwise
*/
        ac_Environment_Feature.cssPropertyAvailable = function(property) {
            if (style === null) {
                style = document.createElement('browserdetect')
                    .style;
            }
            if (prefixes === null) {
                prefixes = ['-webkit-', '-moz-', '-o-', '-ms-', '-khtml-', ''];
            }
            if (preFixes === null) {
                preFixes = ['Webkit', 'Moz', 'O', 'ms', 'Khtml', ''];
            }
            if (css === null) {
                css = {};
            }

            property = property.replace(/([A-Z]+)([A-Z][a-z])/g, '$1\\-$2')
                .replace(/([a-z\d])([A-Z])/g, '$1\\-$2')
                .replace(/^(\-*webkit|\-*moz|\-*o|\-*ms|\-*khtml)\-/, '')
                .toLowerCase();
            switch (property) {
                case 'gradient':
                    if (css.gradient !== undefined) {
                        return css.gradient;
                    }

                    property = 'background-image:';
                    var value1 = 'gradient(linear,left top,right bottom,from(#9f9),
                    var value2 = 'linear-gradient(left top,#9f9, white);';

                    style.cssText = (property + prefixes.join(value1 + property) +
                        .slice(0, -property.length);
                    css.gradient = (style.backgroundImage.indexOf('gradient') !== -
                    return css.gradient;

                case 'inset-box-shadow':
                    if (css['inset-box-shadow'] !== undefined) {
                        return css['inset-box-shadow'];
```



HTML

CSS

JavaScript

www

API

Web Server

Pilot Study

Setup

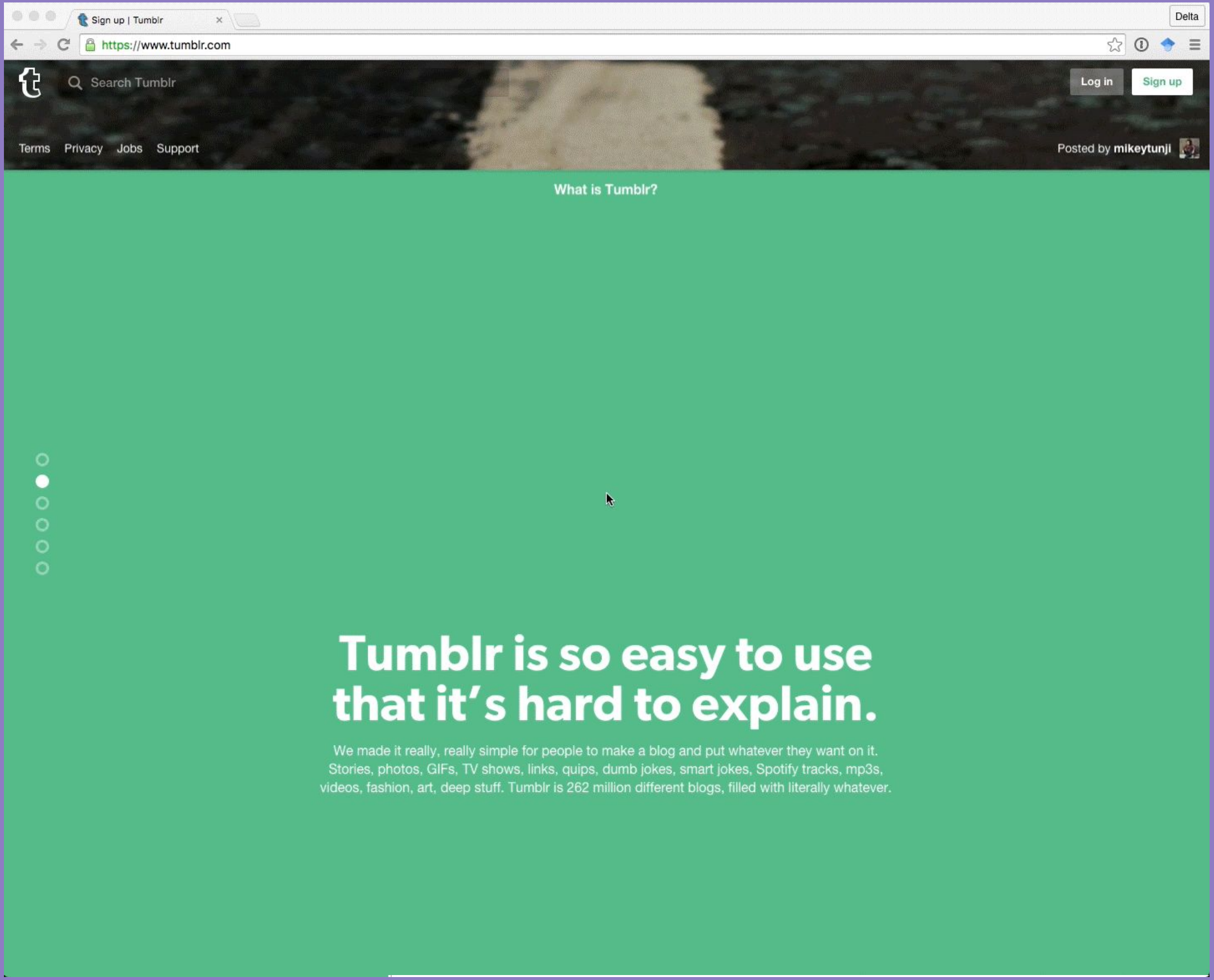
- 4 developers
- 20 mins
- Observe Strategy

Findings

- Find-all 60K LOC JS
- Copy-paste
- Poor tool utilization

Discussion

- Unravel the HTML, JS
- Bubble relevancy
- Hand-off inspection
- Detect libraries



Tumblr is so easy to use that it's hard to explain.

We made it really, really simple for people to make a blog and put whatever they want on it. Stories, photos, GIFs, TV shows, links, quips, dumb jokes, smart jokes, Spotify tracks, mp3s, videos, fashion, art, deep stuff. Tumblr is 262 million different blogs, filled with literally whatever.

Foundational: Inspection, Recording, Replay

FireCrystal Oney, Myers 2009	Web UI Replay with Causal JS
Dynamic Web Breakpoints Barton, Odvarko 2010	DOM Changes set JS Breakpoints
Theseus Lieber, Brandt, Miller 2014	Real-time JS Traces from Browser to Editor
Scry Burg, Ko, Ernst <10m ago	Web UI Replay with Causal JS/HTML/CSS and diffs

Contributions of Unravel in Related Work

Unravel Hibschman, Zhang (Right Now!)	JS HTML traces & libs while recording, then aggregate

Show me what happened first, inspect later

Be portable

Be extendable

Study how it’s used

Unravel [System]

HTML Observations

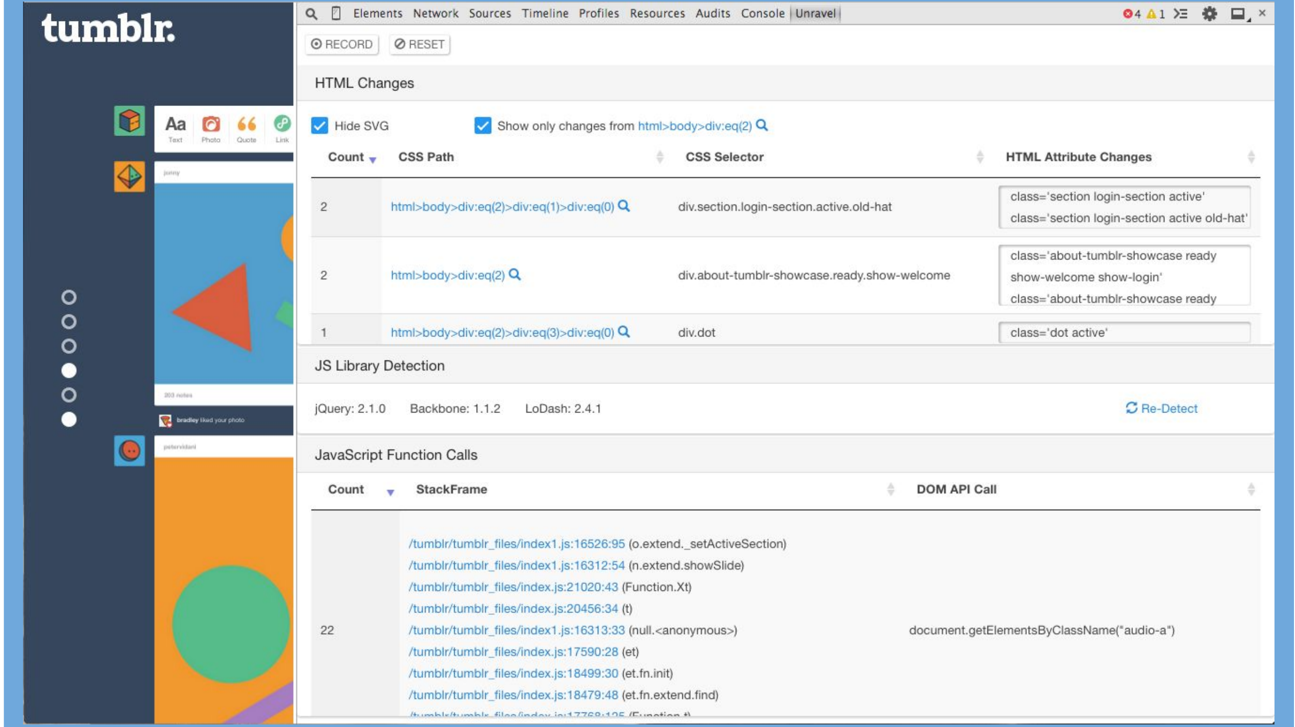
- Observe DOM
- Gather, Reduce, Filter
- Handoff Inspection

JavaScript Traces

- Trace Document API
- Gather, Reduce, Filter
- Handoff Inspection

Library Detection

- APIs
- Syntactic Sugar
- Shims, Polyfills



HTML Changes

☑ Hide SVG ☑ Show only changes from html>body>div:eq(2)

Count	CSS Path	CSS Selector	HTML Attribute Changes

2	html>body>div:eq(2)>div:eq(1)>div:eq(0)	div.section.login-section.active.old-hat	class='section login-section active'
			class='section login-section active old-hat'

2	html>body>div:eq(2)	div.about-tumblr-showcase.ready.show-welcome	class='about-tumblr-showcase ready show-welcome show-login'
			class='about-tumblr-showcase ready

1	html>body>div:eq(2)>div:eq(3)>div:eq(0)	div.dot	class='dot active'

JS Library Detection

jQuery: 2.1.0 Backbone: 1.1.2 LoDash: 2.4.1 Re-Detect

JavaScript Function Calls

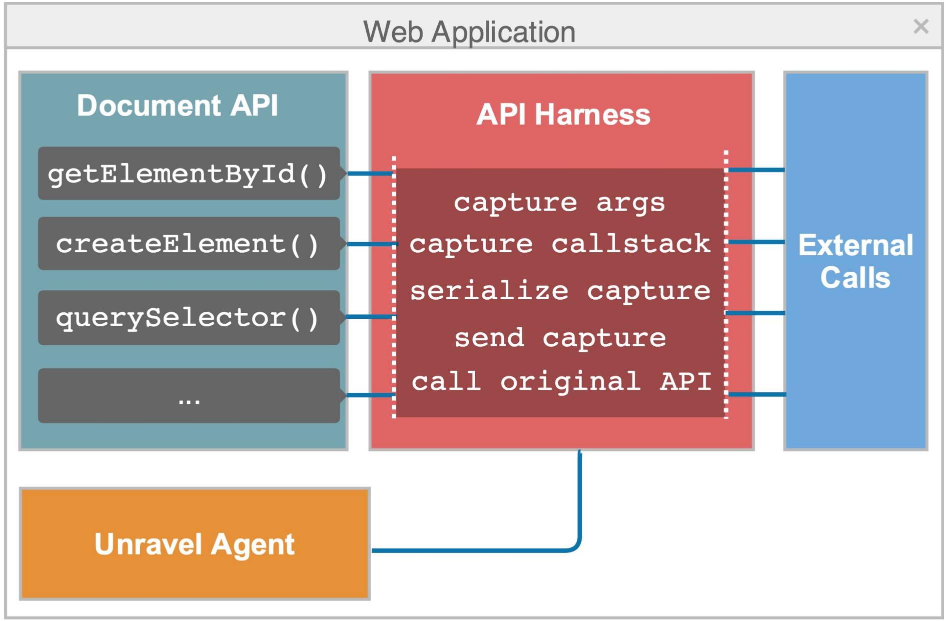
Count	StackFrame	DOM API Call
22	/tumblr/tumblr_files/index1.js:16526:95 (o.extend._setActiveSection) /tumblr/tumblr_files/index1.js:16312:54 (n.extend.showSlide) /tumblr/tumblr_files/index.js:21020:43 (Function.Xt) /tumblr/tumblr_files/index.js:20456:34 (t) /tumblr/tumblr_files/index1.js:16313:33 (null.<anonymous>) /tumblr/tumblr_files/index.js:17590:28 (et) /tumblr/tumblr_files/index.js:18499:30 (et.fn.init) /tumblr/tumblr_files/index.js:18479:48 (et.fn.extend.find) /tumblr/tumblr_files/index.js:17768:125 (Function.t)	document.getElementsByClassName("audio-a")

Live Demo!

Capturing Traces: API Harness

Monitor all invocations of an API

```
for func in window.document
   baseFunc = closure func
   func = function(args) {
     capture args
     throw error
     catch error, get stack
     publish stack and args
     return baseFun(args..context)
   }
```



Web Application

Document API

getElementById()

createElement()

querySelector()

...

API Harness

capture args

capture callstack

serialize capture

send capture

call original API

External Calls

Unravel Agent

User Study

Users

- 13 Junior & Senior developers
- Familiar with Chrome DevTools

Research Questions

- How is strategy altered?
- What is the feature utilization?
- Which barriers were overcome?

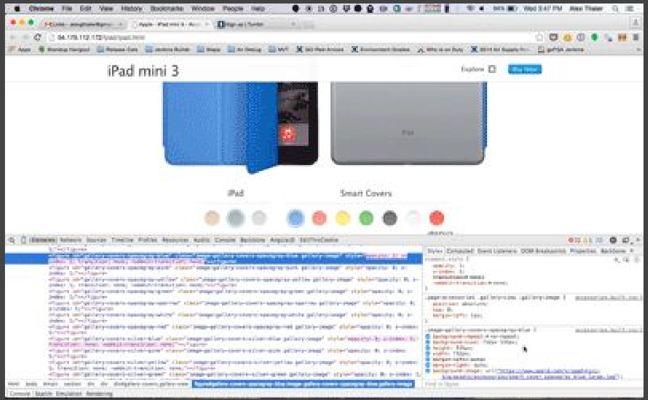
Task: How does a feature work (5)?

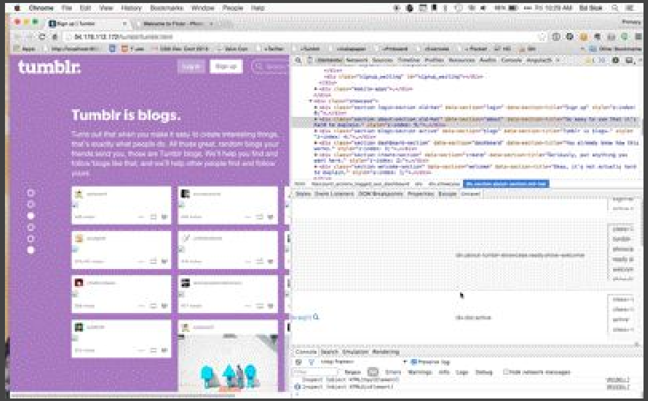
- 15 minutes site A (Unravel)
- 15 minutes site B (Control)
- 15 minutes followup

Milestones

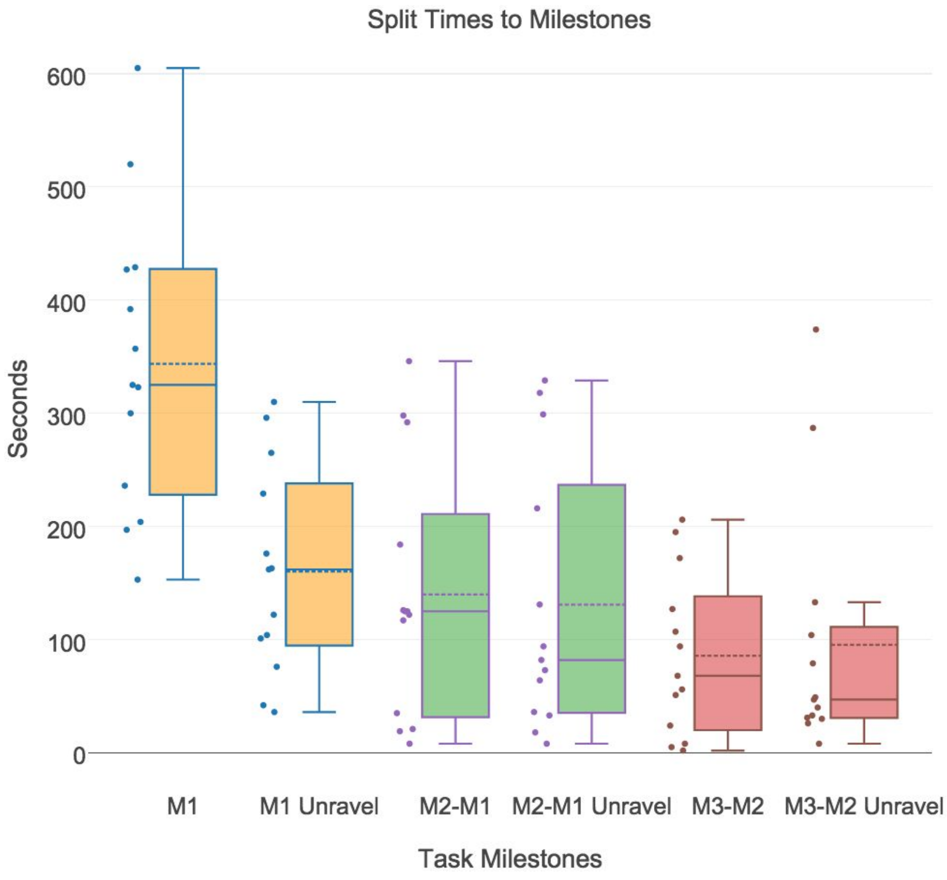
- M1: Key Source Code
- M2: Second Key Source
- M3: “Ah-ha” moment

Stunning photos made easy.
The New York Times
Made with KICKSTARTER
iPad
Smart Covers
Seriously, put anything you want here.
kindle paperwhite $119 $99





Study Findings



53.4% Decrease in time to M1, not M2 or M3

Users looked through 80% less JavaScript

JavaScript Traces were the most helpful feature in user discussions

7 out of 13 users learned a new strategy during the exercise

Discussion

Unravel helped users find a starting point of understanding the code

Overcome Design, Information, Understanding Barriers

Unravel drastically decreased the dependence on code search and inspection

Served both intermediate and expert developers equally well

Limitations

Only the client-side of the story

Doesn’t deal with SVG animation well

Won’t capture WebGL transformations

Closured Document API references hide from Unravel

Minification leaves some variables a mystery while giving names to others

Only tested on a small number of users

Video of the Future!

Unravel

Rapid Web Application Reverse Engineering via Interaction Recording, Source Tracing, and Library Detection

Joshua Hibschman

Haoqi Zhang

NORTHWESTERN UNIVERSITY

DELTA LAB

Spare Slides >>

Implementation [light dependencies]

Inject monitor agent before other scripts load

Listen for events

Aggregate reduce and filter on-the-fly

Deep dive into the murky waters of script loading

By Jake Archibald

Published: June 5th, 2013

Comments: 57

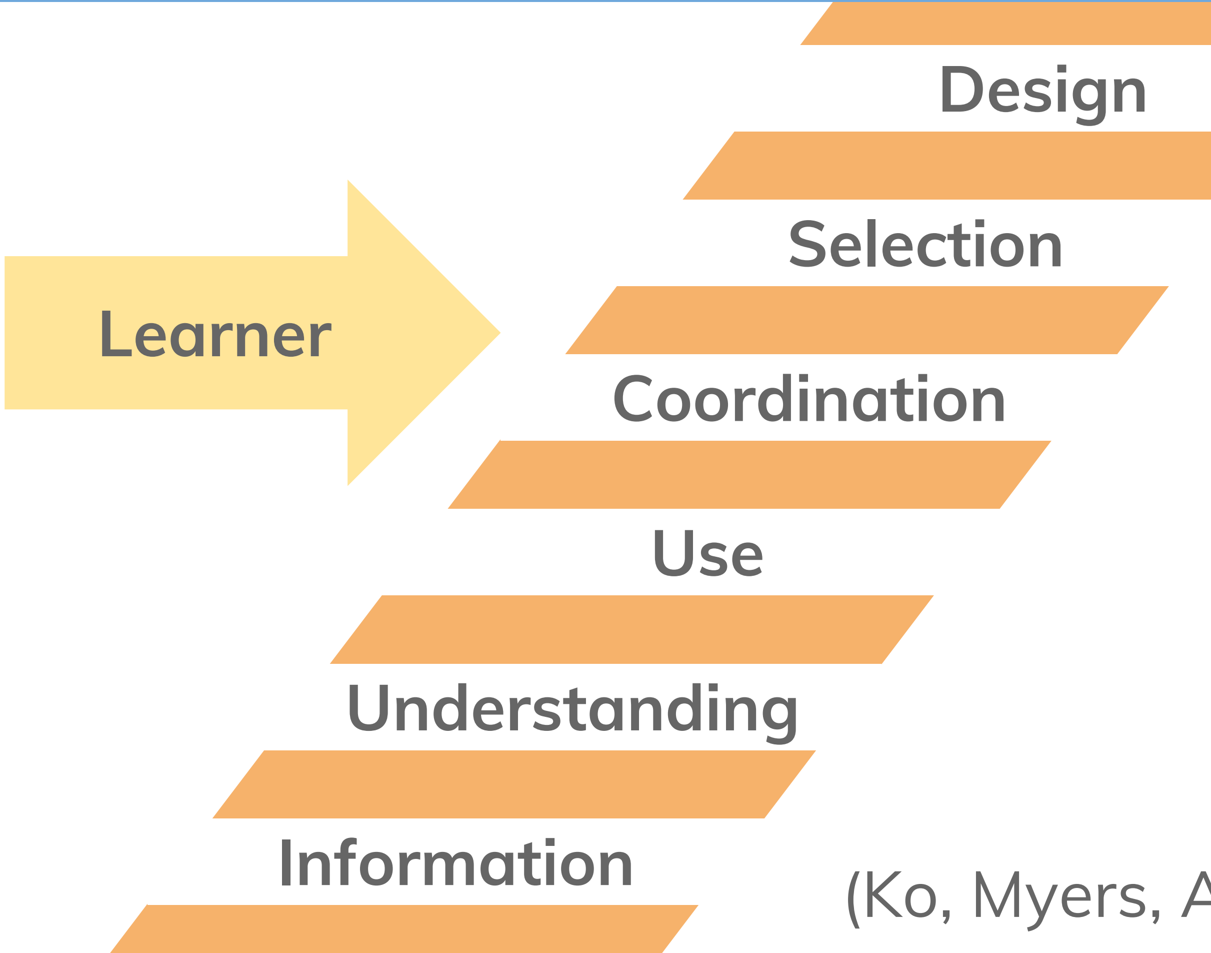
Design Goal: Promote thick authentic learning

THICK

Personally meaningful	How did they do that?
Relate to the real-world	Professional websites
Think in modes of a discipline	A programmer, user, tester, reverse engineer
Assessment reflects the learning process	Integrate and reuse professional web techniques

(Shaffer and Resnick, 1999)

Design Goal: Overcome Learning Barriers



Design

Selection

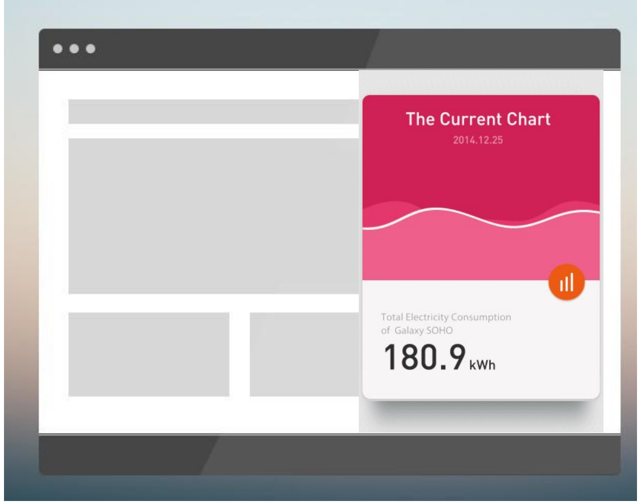
Learner

Coordination

Use

Understanding

Information



(Ko, Myers, Aung, 2004)

Understanding Changes in Dynamic UI

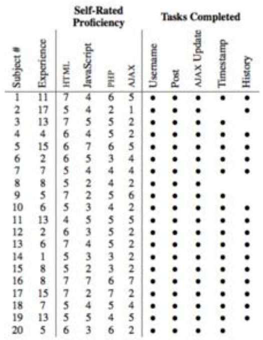
Web Foraging
(Brandt et al 2009)

Timelapse
(Burg et al 2013)

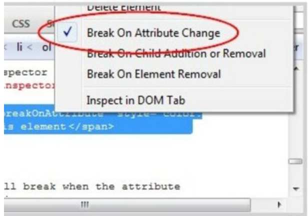
Gliimpse
(Dragicevic et al 2011)

Scotty
(Eagan et al 2011)

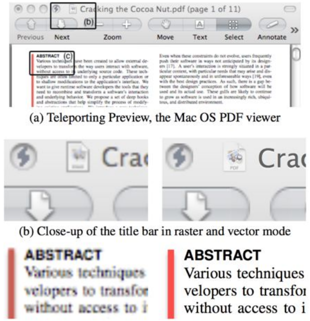
Mimic
(Breslav et al 2014)



		Self-Rated Proficiency				Tasks Completed				
Subject #	Experience	HTML	JavaScript	PHP	AJAX	Username	Post	AJAX Update	Timestamp	History
1	11	7	4	6	5	•	•	•	•	•
2	17	5	4	2	1	•	•	•		•
3	13	7	5	5	2	•	•	•	•	
4	4	6	4	5	2	•	•	•	•	•
5	15	6	7	6	5	•	•	•	•	•
6	2	6	5	3	4	•	•	•	•	•
7	7	5	4	4	4	•	•	•	•	•
8	8	5	2	4	2	•	•	•		
9	5	7	2	5	6	•	•	•	•	
10	6	5	3	4	2	•	•	•	•	•
11	13	4	5	5	5	•	•	•	•	•
12	2	6	3	5	2	•	•	•	•	•
13	6	7	4	5	2	•	•	•	•	•
14	1	5	3	3	2	•	•	•	•	•
15	8	5	2	3	2	•	•	•	•	•
16	8	7	7	6	7	•	•	•	•	•
17	15	7	2	7	2	•	•	•	•	•
18	7	5	4	5	4	•	•	•	•	•
19	13	5	5	4	5	•	•	•	•	•
20	5	6	3	6	2	•	•	•	•	



```
spector
nspector
reakOnAttribute style="color:
s element</span>

ll break when the attribute
```

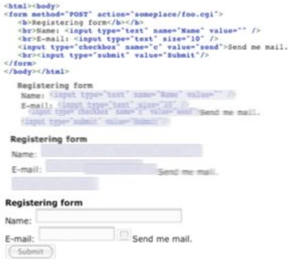


(a) Teleporting Preview, the Mac OS PDF viewer

(b) Close-up of the title bar in raster and vector mode

ABSTRACT
Various techniques
velopers to transfor
without access to i

ABSTRACT
Various techniques
velopers to transfor
without access to i

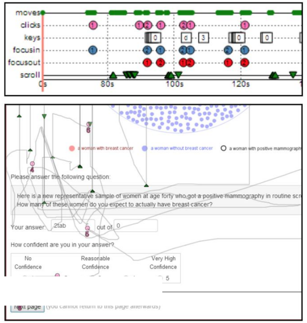


```
<html><body>
<form method="POST" action="someplace/foo.cgi">
   <b>Registering form</b></b>
   <br>Name: <input type="text" name="Name" value="" />
   <br>E-mail: <input type="text" size="10" />
   <input type="checkbox" name="c" value="send">Send me mail.
   <br><input type="submit" value="Submit"/>
</form>
</body></html>
```

Registering form
Name: <input type="text" name="Name" value="" />
E-mail: <input type="text" size="10" />
<input type="checkbox" name="c" value="send">Send me mail.
<input type="submit" value="Submit"/>

Registering form
Name:
E-mail: Send me mail.

Registering form
Name:
E-mail: □ Send me mail.
Submit



Please answer the following question:

Here is a new representative sample of women at age forty who got a positive mammography in routine scr
How many of these women do you expect to actually have breast cancer?

Your answer: 2tab out of 0

How confident are you in your answer?

(you cannot return to this page afterwards)