

AUTOMATED TESTING FOR SEO with Mike King & Hamlet Batista



 \checkmark

Your Speakers



Mike King Managing Director at iPullRank Hamlet Batista CEO at RankSense \checkmark

Has this ever happened to you?





♡ 1

ılı

Follow

 \sim





Replying to @iPullRank

I got threatened with being sued, a week after a client signed a contact, because their Google traffic went to zero. I hadn't started as they had asked me to wait a week as their site redesign was finished. That redesign included adding Disallow: / to robots.txt

Follow









Alter California

and the same same and

A REAL PROPERTY AND

And in case of the local division of the loc

Sec. Sec.

No. of Concession, Name

5/4

street, some different

TO THE ADDRESS TATION.

- ----

and the state of the

and a start of the start of the

all and I found part and

 \checkmark

What is Automated Testing?



So first, let's talk just a little bit about modern web development and how it works.

Why Does This Matter?



A lot of modern software frameworks follow this design pattern that allows for a separation of concerns between data, logic and frontend.



In this example the PHP framework Laravel manages the Model and the Controller and the JavaScript framework React acts as the View.

URLs point to controllers which then return the right views featuring data from the right models.

This is why segmented crawls are important to developers being able to implement recommendations.



Render HTML in the client's browser

Waterfall Rarely Makes Sense Anymore

It's unlikely that the websites you work on will follow a waterfall methodology. Everything is agile now.



(

Most Dev Teams Follow Agile Methodology

Small pieces of functionality are tackled in short time frames or "sprints."



()

Continuous Integration



Often engineering teams are following a Continuous Integration model where teams and individuals are autonomously pushing their code as they complete it.



Deployment Pipelines feature Automated Tests

Continuous Deployment



Deployment pipelines allow for safeguards to be put in place to ensure that everyone is following the same process to get things deployed.

Add Failure Condition Fail build on specific text in build log ∇ Fail build if its build log:⁽²⁾ contains exact text ٧ • error The exact text to look for. Note: The time and block name prefixes preceding each message in the build log are ignored. Failure message: The message to display in the UI and the build log. Test on finished build Save Cancel

Certain tests are deemed as critical to the infrastructure of the website and will cause the deployment process to fail and the developer has to go back and fix it before they can continue with their work.



Mashable





ILLUSTRATED BY SEGUE TECHNOLOGIES

There are three flavors of testing that are most relevant to us as SEOs: Unit Tests, Integration Tests and GUI Tests.

25



Unit tests can be used to determine the presence of tags and attributes in structure of code to be served to the client, but not the rendered code.

Unit Testing

With Unit tests you are testing an individual function or procedure to ensure that that it works.

These are generally small and fast. Codebases can have many of these.

With integration testing you're testing how your logic comes together to build the bigger system. It can involve external data such as APIs and databases or how the various pieces of the systems work together.

This is significantly slower than unit testing, but depends on how much of the system needs to be rendered in order to complete the test.



Integration tests can be used to determine tags, features and content of completed HTML source being served to the client. However, there can be overlap with what can be viewed in an UI test.

ر*ه



UI tests can be used to determine tags, features and content of the rendered DOM being served to the client.

Integration Testing

UI Testing is the testing of the frontend interface of your website. This is largely used for ensuring front end components work as intended.

This is the slowest version of testing because it requires the page to be fully rendered.

This is What Headless Browsers are Actually For







Certain tests are deemed as critical to the infrastructure of the website and will cause the deployment process to fail and the developer has to go back and fix it before they can continue with their work.

Read Why the Chrome DevTools Team Made Puppeteer

FAQ

Q: Who maintains Puppeteer?

The Chrome DevTools team maintains the library, but we'd love your help and expertise on the project! See Contributing.

Q: What are Puppeteer's goals and principles?

The goals of the project are:

- Provide a slim, canonical library that highlights the capabilities of the DevTools Protocol.
- Provide a reference implementation for similar testing libraries. Eventually, these other frameworks could adopt Puppeteer
 as their foundational layer.
- · Grow the adoption of headless/automated browser testing.
- Help dogfood new DevTools Protocol features...and catch bugs!
- Learn more about the pain points of automated browser testing and help fill those gaps.

Test Driven Development



Some modern developers follow this methodology wherein they write their tests first and then write their code until the tests pass. Unit tests are primarily used for this, but any type of test could be used. Note: UI Tests will dramatically slow down deployment.

Ψ

How Does This Apply to SEO?

Last Year I Talked about Using Task Runners

Search Engine Land SEO SEM MOBILE LOCAL RETAIL GOOGLE BING SOCIAL MORE EVENTS

How to get developers to implement SEO recommendations

The best SEO strategy is meaningless if there's no one to implement it! Columnist Michael King shares his recommendations for how to best work with your developers to get things done.

Michael King on August 16, 2017 at 12:34 pm



The hardest problem in doing SEO isn't the algorithm updates. It isn't having access to the enterprise tools. It's not even whether or not you have the experience to determine where to focus your efforts.

No, the hardest problem in SEO is getting developers to actually execute recommendations.

https://searchengineland.com/get-developers-implement-seo-recommendations-280318





This is a scale of which the developers you interface with are somewhere between an underachieving known-it-all and overachieving do-it-all.

@BritneyMuller brings up a great point

My goal with that post and talk is to introduce you to another way to integrate developers and ensure we're all moving towards working better together.









Replying to @iPullRank

I'd love to flip this convo:

There's A TON we can learn from developers!

Lack of communication & ownership is often to blame for dev mishaps.

Hope to build content about cultivating healthier dev/SEO relationships very soon. 8:39 PM - 27 Sep 2018

🐔 🚇 🕼 🌰 🧝 6 Likes QZ 17 M

ر*ه



Let's go back through those **horror stories**.




Replying to @iPullRank

Can't imagine much worse than a noindex tag being added to the live site...

6:11 PM - 27 Sep 2018 2 Likes
↓
2 Like

This is a critical issue so this test should fail the build.

A Test Can Catch This

Unit Test Failure Condition – Presence of meta noindex tag on functions that impact the <head>.

Integration Test Failure Condition – Presence of meta noindex on routes that should not have them.

Not Sure I'd Test for this One

Integration Test Failure Condition –

H1 tag is the same in the HTML on every accessed page in our test dataset.

UI Test Failure Condition -

H1 tag is the same in the rendered DOM on every accessed page in our test dataset.



Replying to @iPullRank

They used "[name of company] Blog" in an H1 on every blog page as part of the template. The post title was H2.



I would not recommend having this test fail the build though.



Replying to @iPullRank

I got threatened with being sued, a week after a client signed a contact, because their Google traffic went to zero. I hadn't started as they had asked me to wait a week as their site redesign was finished. That redesign included adding Disallow: / to robots.txt

6:15 PM - 27 Sep 2018

Robots.txt Test

Unit Test Failure Condition –

Presence of "User-agent: * Disallow: /" on prod in a function for the robots.txt route.

Integration Test Failure Condition –

Presence of "User-agent: * Disallow: /" on prod in robots.txt URL.

SSR Test

Integrated Test/UI Test Failure Condition -

Key body content is not available in the HTML version.

This one requires the rendered DOM as a comparison so it technically could be a UI and/or Integrated Test.



Replying to @iPullRank

Multi-million page site having massive deindexing (millions of pages) due to bad JS on relaunch. Revenue model collapsed.

6:27 PM - 27 Sep 2018 from Kailua, HI



This should definitely fail the build.





Replying to @iPullRank

Reverting canonical tags to old versions of URLs. No one noticed until the old URLs. started showing up in GA.

6:49 PM - 27 Sep 2018

0 M O 11 \bigtriangledown

This is a critical issue so this test should fail the build.

Rel-canonical

V

V

UI Test Failure Condition –

URLs that should have a self-referencing canonical do not in the DOM.

Integration Test Failure Condition –

URLs that should have a self-referencing canonical do not in the HTML.

Unit Test Failure Condition -

Route does not set canonical tag on the server side.

UI Test Failure Condition –

Rendered DOM features tags in the <head> sub-tree that are invalid.

Integration Test Failure Condition -

HTML features tags in the <head> that are invalid or HTML does not pass W3C validation.

Unit Test Failure Condition -

Function that returns the <head> features tags that don't fit a list of valid <head> tags.



Rogue tag in <head> breaking all hreflang. Fun.

9:39 PM - 27 Sep 2018

9 tl 🛛 🗹 🗹

This is a critical issue so this test should fail the build.

ر*ه



Replying to @iPullRank

I also once had a dev let a site update go live without consulting the SEO team, so all site arch., schema, metas, and the whole nine yards were wiped out. That was a NIGHTMARE

6:14 PM - 27 Sep 2018

1 Like

But Testing is Not a Magic Bullet

Sure, we can develop tests for to look for the presence of structured markup and to enforce metadata specifications, but when things go this wrong, you have to put your team member, project management, consultant, management hat on and get to the bottom of what went wrong. (*)

Who Usually Does this Work?



Developer

QA Engineer





Doing an SEO Audit is basically doing similar work to that of a QA Engineer.

Crawling HTML is Somewhat like Integration Testing

You're looking at the results of what the software spits out when all the systems come together.



(



Crawling JavaScript is more like UI Testing

You're looking at the results when the user interface is rendered.

...And you should be comparing the differences

<view rendered source>

URL: https://www.coindesk.com/

Tip: Source code is beautified so formatting may not match real source

Show/hide: @ Raw @ Rendered @ Difference Comparison

Raw (1	🛄 As mobile	Сору	Rendered	1	Copy	Difference Comparison (1) (0 of 42) Copy					
<pre>ci+-[if it IE 7]> (ntml class="no-js ie ie5 oldie" lang="en"> ci[en</pre>	dif]> if IE 7? 0]> <html class="no
" ie"="" lang="en" no-js=""></html>	-js <1	<pre><l<l[endif]><head></head></l<l[endif]></pre>	.com/sax2/epsteg.js">		<pre>it[if it IE 7]> chtml class="mo-js ie ie6 oldie" lange"en"> <?endif?><?if IE 7?> chtml class="mo-js ie ie7 oldie" lange"en"> <!--(endif)--><?if IE 8?> chtml class="mo-js ie le8 oldie" lange"en"> <!--(endif)--><?if IE 9?> chtml class="mo-js iendif]><?if (gt IE 9)]!(IE)?><!-----><!--(if (gt IE 9)]</pre--></pre>					
<title>CoinDesk - Leader in blockchain news.</title>			apikey+coindesk.com8ampjuuid+c36e8c4f-006d-4bf2-99f8- 8180e2f56b188ampjuu1=https://www.coindesk.com/&jcallback=10uery112408402335583664964_15380			<pre>ctct[endif]>cnead></pre>					
k rel="dns-prefetch" href="//media.coindesk.com">			936793066amp;_+1536093679307">			<pre><script aax2="" apstag.js"="" async*"="" c.amazon-adsystem.com="" src*"=""></script></pre>					
<pre>clink rel="alternate" type="application/rss=xml" title="Coi bref="https://feeds.feedburner.com/CoinDesk"></pre>	nDesk RSS Feed"		<pre></pre>	front.net/code/ptrack-v1.1.1-engaged	dtime-	<pre><script)<="" async="" href="1" icon"="" pre="" shortcut="" src="https://api.persely.com/v2/profile) apixey=coindesk.com&uuid=c36e6c4f-do6d-4bf2-99f8-</pre></td></tr><tr><td colspan=3><pre>(link rel="></td><td><pre><script async="" type="text/javascript" s</pre></td><td>rc="//a.pub.network/core/prebid-ana</td><td colspan=4><pre>8180e2*560186amp;ur1*https://www.coindesx.com/∓callback*jQuery11240840235583664964_15380 936793066amp;_*1538093679307*></script></pre>					
<pre></pre>	1711.css" rel="styles"	eet"	1.19.0.js"> (script async="" type="text/javascript"	Service and an and a service of the		<pre><script src="https://dlzjf7jlzjs58.cloudfront.net/code/ptrack-v1.1.1-engagedtime-
slots.js"></script></pre>					
(meta charset="utf-8")			<pre>src="//www.googletagservices.com/tag/js/got.js"></pre>	/script>		<pre>cscript async="" type="text/javascript" src="//a.pub.network/core/prebid-analytics-</pre>					
<pre>cmeta name="p:domain_verify" content="266ef71c542343fed103b</pre>	1f5f9ebba6e">		<pre><script async*"="" src="https://s.skimresources.com/is/100652K157442</pre></td><td>S.seimlines.is" type="text/javascript">c/script></td><td></td><td>1.19.0.js"></script></pre>								
<pre>(meta property="twitter:account_id" content="1410565129"></pre>			<pre>cscript async+"" type+"text/javascript"</pre>	anterest difference who defer the date		<pre>(script async="" type="text/javascript" src="//www.googletagservices.com/tag/js/gpt.js"></pre>					
<pre>cmeta name="alexaVerifyID" content="C6q9LThNt82vzHpLosc35-a</pre>	rucht>		SPC- Mttps://acexiendinice.cloud#ront.net/js/colm	Deskcon/verkeywee.win.js->s/script>		<pre><script <="" async="" pre="" type="text/javascript"></td></tr><tr><td><pre><meta name="msapolication-config" content="none"></pre></td><td></td><td></td><td><pre>snc="https://sb.scorecardresearch.com/beacon.js"></pre></td><td></script></pre>		<pre>src+"https://s.skimresources.com/js/100652X1574425.skimlinks.js"></pre>			
<pre>cmeta name="viewport" content="width-device-width.initial-s</pre>	Sevice-width, initial-scale=1">		<pre><script src="https://srv-2018-09-28-00.config.parsely.com/config/coindesk.com"></pre></td><td colspan=4><pre>cscript async="" type="text/javascript" src="https://dcBxl@ndzn2cb.cloudFront.net/is/coindeskcom/v@/keywee.min.is">c/script> </pre></td></tr><tr><td><pre>(meta name="format-detection" content="telephone=no"></pre></td><td></td><td></td><td colspan=2></script></pre>			condet approximate property (approximate)					
<pre>cmeta mame+"microtip" content+"161pCNLKZCZde4dgznZDE5wiikQeyOSeuh" data-</pre>			<title>CoinDesk - Leader in blockchain news.</title>			<pre>src+"https://sb.scorecardresearch.com/beacon.js"></pre>					
currency= ptc deterreciptent= coincesc >			k rel="dns-prefetch" href="//hedia.coindesk.com">			<pre><script src="https://srv-2018-09-28-00.config.parsely.com/config/coindesk.com"></script></pre>					

John Hogg from iProspect shared his diff comparison Chrome extension that computes the DIFF between HTML and rendered DOM. When it's released I encourage you to check it out.

Your Role is To Help Plan the Tests



You'll want to work with a QA Engineer, a Product Manager or a Developer to develop test cases that can then be built into.

Here are Few Examples

Test Case ID	Test Scenario	Test Type	Test Steps	Test Data	Expected Results	Actual Results	Pass/Fail
SEO01	Presence of Meta Descriptions	Unit Test	Check for presence of meta description tag in HTML	Page template code	All URLs should have meta description	Product Detail Page is missing meta description	PASS
SEO02	Viable Internal Links	Functional Test	1.Render pages2.Open allinternal links3.Reviewresponse codes	Crawled URL data	All links return 200 response code	Many links to redirects and 404s	FAIL
SEO03	Average Page Speed Less than 2 Seconds	Functional/Integ ration Test	1.Render pages2.Capture pagespeed3.Determineaverage pagespeed per pagetype	Render all page types from URL list	All page types should return an average of 2 seconds load time	Homepage takes 5 seconds to load	FAIL

These examples of tests should be enough to get you started on defining your own potential tests for SEO.

 \heartsuit

Tools

Some tools you can use to get your automated testing for SEO off the ground

If your developers are doing Continuous Integration or Continuous Delivery, it's likely that they are using Jenkins. https://jenkins.io/



Jenkins

ر*ه

Get Failure Alerts in Slack

Jenkins plugin for posting r	notifications to a Slack channel	UII magnis	
@ 296 commits	¥ 2 branches	🛇 10 releases	11 51 contributors
Icanch: master • New pull requ	vest	Create new file Uploa	d files Find file Clone or download *
This branch is 265 commits alree	ad, 2 commits behind tinyspeckmaster.] #400 💮 Compan
Sign bekanda and kmadel Add mul	tiple channel comment to slackSend docs (#399)		Latest commit 1875926 6 days ago
Ba src	Add multiple channel comment to slackSen	d docs (#399)	6 days ago
D .gitignore	add .iml files to gitignore (for IDEA)		3 years ago
CHANGELOG.md	Update CHANGELOG for #294		2 years ago
README.md	Fixed getSlackService test		a year ago
RELEASE.md	add Jenkins Update Center to RELEASE che	cklist	3 years ago
E pomami	(maven-release-plugin) prepare for next de	velopment iteration	a year ago
E README.md			



Provides Jenkins notification integration with Slack or Slack compatible applications like RocketChat and Mattermost.

Install Instructions for Slack

Get a Slack account: https://slack.com/
 Configure the Jenkins integration: https://my.slack.com/services/new/jenkins-ci



Jenkins APP 9:51 PM

Christopher Rung

Test Project, build #2

Failed

Christopher Rung

Branch christopher.rung_JIRA-123 Test Results Passed: 9, Failed: 1, Skipped: 0

JIRA-123: This is a test commit message

Failed Tests

Last Commit

testInvalidReturnsNull: com.rung.testProject.ClassTest > testInvalidReturnsNull FAILED java.lang.AssertionError: expected:null but was:<1> at org.junit.Assert.fail(Assert.java:88) at org.junit.Assert.assertEquals(Assert.java:631)

Show more...

Or in Jenkins' logs



You could also just regularly the logs that Jenkins provides.

I

Jenkins logging directly to MySQL database

tags: groovy jenkins mysql

Written on January 27, 2017

There are cases if you like to send the log output from Jenkins directly to a database for better querying or just as a step for further processing them.

As Jenkins is not really admin-friendly with redirecting logs, we have to use the init.d feature of Jenkins (https://wiki.jenkins-ci.org/display/JENKINS/Configuring+Jenkins+upon+start+up). Thus you can simply put an init.groovy file into the jenkins home directory. In this init file you can do a lot of stuff like setting some variables, disabling the sending of usage statistics and so on. And you can hook into the LogManager of Jenkins: https://wiki.jenkins-ci.org/display/JENKINS/Logging

In this example we are using Jenkins 1.651.3

Jenkins is using the java.util.logging package, so first of all we need a handler for sending LogRecords to a MySQL database:

@Grapes([

@Grab(group='mysql', module='mysql-connector-java', version='5.1.40'), @GrabConfig(systemClassLoader = true)

1)

```
import java.util.logging.*;
import java.sql.*;
```

```
class MySqlHandler extends Handler {
```

def driverString = "com.mysql.jdbc.Driver"; def insertSQL = "insert into log (level,logger,message,sequence,sourceClass,sourceHethod,threadID,t def connectionString; def connection; def prepClear; def prepInsert; public MySqlHandler(String connectionString) { try {

this.connectionString = connectionString; Class.forName(driverString);

Or Pipe it into MySQL

Sending the Jenkins logs into MySQL gives you the opportunity to monitor and visualize failures and successes of builds with respect to SEO in a variety of ways such as through Google Data Studio.

You'll Also Need a Crawler

https://github.com/yujiosaka/headlesschrome-crawler

Headless Chrome Crawler	paramy Conscilutoper modeled
PT Examples Tips Code of Conduct Contributing Changelog	
Distributed crawler powered by Headless Chrome	
Features	
Traviers based on simple requests to HTML files are generally fast. However, it ometimes ends up capturing empty bodies, especially when the websites are suit on such modern frontend frameworks as Angulari5, React and Yue js.	200
owered by Headless Chrome, the crawler provides simple APIs to crawl these	
tynamic websites with the tollowing features:	Headlesss Chromo Crawler
Distributed crawling	Chrome Crawler
 Configure concurrency, delay and retry 	
 Support both depth-first search and breadth-first search algorithm 	
 Pluggable cache storages such as Redis 	
 Support CSV and JSON Lines for exporting results 	
 Pause at the max request and resume at any time 	
 Insert jQuery automatically for scraping 	
 Save screenshots for the crawling evidence 	
 Emulate devices and user agents 	
 Priority queue for crawling efficiency 	
Obey robots.txt	
 Follow sitemapumil 	
[Promise] support	
Setting Started	

https://github.com/antivanov/js-crawler

-	EADML md
	js-crawler
	Web crawler for Node.JS, both HTTP and HTTPS are supported.
	Installation
	ope install js-crasler
	Usage
	The crawler provides intuitive interface to crawl links on web sites. Example:
	ver Grader = require("fs-crader").default;
	<pre>new Grawler().configure([depth: 3]) .crawl("http://waw.pogle.com", function onSuccess(page) (.coms0.confere.rul);</pre>
	Di
	The call to configure is optional, if it is omitted the default option values will be used.
	orduccess callback will be called for each page that the crawler has crawled, page value passed to the callback will contain the following fields:
	• se1 - URL of the page
	centent - body of the page (usually HTML) status - the HTTP status code
	Extra information can be retrieved from the rest of the page fields: error, response, body which are identical to the ones passed to the callback of request, invocation of the Request module, referenc field will reference the url of the page that le the crawler to the current page.
	Options-based API

Alternative APIs for passing calibacks to the cract function.

You'll want a headless crawler and a text-based crawler to spin up pages and run tests on and a fixed list of pages that represent all of the public-facing routes.



You'll Also Need a Crawler



Building your crawler with AWS Lambda, Azure Functions or Google Clouds Functions may be a great way to enjoy scale without having to manage additional servers.



Shoutout to **@vvgomes**



Vini's Blog Post on Serverless Architectures

Event Driven Architectures & Serverless

MONDAY, JULY 23, 2018 - 8 MINS

EVENT-SOURCING SERVERLESS

After a 2.5 years journey building an event-driven services platform for a large financial organization, I've recently joined a new team in the middle of their on-going migration to the public cloud. One of the things we've been exploring is the Serverless model in order to release software with close to none operations effort.

Turns out there are significant commonalities between EDAs (Event-Driven Architectures) and Serverless which can lead to interesting advantages from the architectural point of view when applied to the appropriate business context. That was the topic of my recent presentation at the <u>North America XConf</u>.



http://vvgomes.com/edas-and-serverless/



Tips On Testing

Make only mission critical SEO tests potentially fail the build. Most SEO issues will be spotted in your integration tests. Include your SEO tests in every build. Keep your testing suites separate. Do all crawling in parallel. **、*** 力

Things to Play With

Because I never like to leave you without having something you can do

Check out these Video Tutorials on Jenkins

Get started with Jenkins PLAY ALL

In these video tutorials I'll show you how to get started with Jenkins, the open source continuous integration & delivery application. I'll cover topics like installing Jenkins, importing git source code,



Play with Jenkins: https://www.youtube.com/playlist?list=PLzvRQMJ9HDiSaisKr7OnM4FI7JXCDDcmt

Kantu Browser Automation Tool



See how Browser Automation tests work with a chrome extension: https://chrome.google.com/webstore/detail/kantu-browser-automation/gcbalfbdmfieckj Inblleoemohcganoc?hl=en

Codeception



If you dabble in PHP or have a WP site you want to play around with, check out Codeception. https://codeception.com

Puppeteer

Puppe	eer	
build failing	build persong CD persong open v1.10.0	
API FAQ Cont	ributing Troubleshooting	-
Puppeteer DevTools P headless) C	s a Node library which provides a high-level API to control Chrome or Chromium over the rotocol. Puppeteer runs headless by default, but can be configured to run full (non- hrome or Chromium.	~
What can I do?	Pup	be
Most things the examples to g	at you can do manually in the browser can be done using Puppeteer! Here are a few et you started:	22
• Generate	screenshots and PDFs of pages.	
Crawl a S	A (Single-Page Application) and generate pre-rendered content (i.e. "SSR" (Server-Side Rendering)).	
Automate	form submission, UI testing, keyboard input, etc.	
 Create an latest Jav 	up-to-date, automated testing environment. Run your tests directly in the latest version of Chrome us sScript and browser features.	ing
Capture a	timeline trace of your site to help diagnose performance issues.	
Test Chro	me Extensions.	
Give it a spin:	https://try-puppeteer.appspot.com/	
Getting S	tarted	
Installation		
To use Pupper	eer in your project, run:	
nos 1 euros	teer	

The Headless Chrome library has tons of scraping use cases. https://github.com/GoogleChrome/puppeteer

PuPHPeteer

EADME.md			
PuPHPetee			1
php >=7.1 Composer vi	6.0 Node >=8.0.0 NPM v1.4.0 Build Status pess	na	
A Puppeteer bridge for	PHP, supporting the entire API. Based on Rial	to, a package to manage	7-
Node resources from P	IP.		119
Here are some example	borrowed from Puppeteer's documentation	and adapted to PHP's	
syntax:			VTI
Example - navigating t	https://example.com and saving a screensho	t as example.png.	
use Nesk\Puphpeteer	uppeteer;		
\$puppeteer = new Pu	eteer;		
\$browser = \$puppeter	->launch();		
<pre>\$page = \$browser->m</pre>	Page();		
<pre>\$page->goto('https:</pre>	example.com');		
<pre>\$page->screenshot([</pre>	<pre>wath' => 'example.png']);</pre>		
<pre>\$browser->close();</pre>			
Example - evaluate I s	ipt in the context of the page:		
use Nesk\Puphpeteer	uppeteer;		
use Nesk\Rialto\Dat.	JsFunction;		
\$puppeteer = new Pu	eteer;		
\$browser = \$puppeter	->launch();		
\$page = \$browser->n	Page();		
<pre>\$page->goto('https:</pre>	'example.com');		
// Get the "viewpor	of the page, as reported by the page.		
\$dimensions = \$page	evaluate(JsFunction::createWithBody("		
return {			
width: docu	nt.documentElement.clientWidth,		

PHP Wrapper for Puppeteer https://github.com/nesk/puphpeteer

@Aleyda Mentioned Monitoring Tools



Aleyda Solis 🤣 @aleyda

If you want to automatically monitor some of the most fundamental SEO related configurations I highly recommend you check out @littlewardenapp, @contentking and @seomonitor \checkmark

Following

MyCool King @iPullRank

Ladies and gentlemen, I'd like to share with you a presentation called "Software Testing for SEO"

bit.ly/save-the-seo...

5:34 AM - 28 Sep 2018

ر*ه



Monitoring is **not the same** as a automated testing.





Follow

V

Replying to @aleyda @littlewardenapp @seomonitor Thanks @aleyda!

@iPullRank ContentKing's built to tackle most of the situation you've covered in your presentation. Using our API, you can also integrate it in your CI process

We'd be happy to set up a demo account for you to play around with! 🤓

5:41 AM - 28 Sep 2018

ContentKing Has an API

69

Agenda

- Diagnose SEO and page speed issues related to the incorrect nesting of HTML tags and scripts.
- Learn to use the Chrome JavaScript Debugger to track down serious SEO issues.
- Speed up JavaScript by removing unused code and implementing code splitting when appropriate.
- Leverage service workers and to edge workers for more powerful use cases.

(*)

How Misplaced HTML tags Hurt SEO

Let's see what happens to the canonical tag when we insert a <div> in the HTML HEAD.

1	
2	<pre></pre>
з	
-4	<html lang="en"></html>
5	<head></head>
6	<meta charset="utf-8"/>
7	
8	<title>This is a test</title>
9	<meta content="This is a test" name="description"/>
10	<meta content="Hamlet Batista" name="author"/>
11	
12	<pre><div>This is shouldn't be here </div></pre>
13	
14	<iink nrei="<u" rei="canonical">http://dev.ranksense.com/invalid_head_hard_coded.ntmi</iink>
15	hereit</td
30	
18	shadya
19	
20	<div>This is a test</div>
21	
22	
23	
24	
25	

(

How Misplaced HTML tags Hurt SEO

Б	Elements	Console	Sources	Network	Performance	Memory	Application	Security	Audits
docty<br <html 1<br="">*<head <me <ti <me <th><pre>rpe html> ang="en" rs_ rs_id="22"> ta charset=" tle rs_id="2 ta name="des ta name="aut d> rs id="27"</pre></th><td>id="21"> utf-8" rs_i 4">This is cription" c hor" conten data-or-c-s</td><td>d="23"> a test</td></me </ti ontent="Th it="Hamlet :-loaded="t</me </head </html>	<pre>rpe html> ang="en" rs_ rs_id="22"> ta charset=" tle rs_id="2 ta name="des ta name="aut d> rs id="27"</pre>	id="21"> utf-8" rs_i 4">This is cription" c hor" conten data-or-c-s	d="23"> a test	tle> is is a tes Batista" rs rue">	:t" rs_id="25"> i_id="26">				
<di <li </li </di 	v rs_id="28" nk rel="cano v rs_id="30" y>	>This is sh nical" href >This is a	ouldn't be =" <u>http://d</u> test	here <td>▶ e.com/invalid</td> <td>head hard</td> <td>coded.html"</td> <td>rs_id="29"></td> <td>• == \$8</td>	▶ e.com/invalid	head hard	coded.html"	rs_id="29">	• == \$8
html bo	dy link								

It ends up inside the HTML BODY.

Y
You never get an "Invalid Syntax" error on an HTML page. Browsers fix any invalid content and go on.

Take this HTML for example:

-			
<ntmt></ntmt>			
<mytag></mytag>			
<div></div>			
<0>			
Really lou	SV HTMI		
dias	by mme		
5/P2			
Really lou 	sy HTML		

I must have violated about a million rules ("mytag" is not a standard tag, wrong nesting of the "p" and "div" elements and more) but the browser still shows it correctly and doesn't complain. So a lot of the parser code is fixing the HTML author mistakes.

Error handling is quite consistent in browsers, but amazingly enough it hasn't been part of HTML specifications. Like bookmarking and back/forward buttons it's just something that developed in browsers over the years. There are known invalid HTML constructs repeated on many sites, and the browsers try to fix them in a way conformant with other browsers.

How Misplaced HTML tags Hurt SEO

This is the result of browsers error tolerance:

"the element being added is explicitly forbidden inside some outer tag. In this case we should close all tags up to the one which forbids the element, and add it afterwards."

https://www.html5rocks.com/en/ tutorials/internals/howbrowsersw ork/#Browsers_error_tolerance

Move SEO tags to the top of the HTML HEAD.

Check: The dangers of misplaced third-party scripts

https://searchengineland.com/t he-dangers-of-misplaced-third-pa rty-scripts-327329

Search Engine Land SEO SEM MOBILE LOCAL RETAIL GOOGLE BING SOCIAL MORE EVENTS

How to fix this

Fortunately, fixing this problem is actually very simple. We have two alternatives. A lazy one proper one.

The proper fix is to track down scripts that insert invalid HTML tags in the head and move t HTML body.

The lazy and quickest fix is to move all SEO tags (and other important tags) before any thir scripts. Preferably, right after the opening <HEAD> tag.

You can see how I do it here.

1	html
2	
3	<html lang="en"></html>
4	<head></head>
5	<meta charset="utf-8"/>
5	
7	<title>This is a test</title>
в	<link href="http://localhost" rel="canonical"/>
9.	<meta content="This is a test" name="description"/>
8.	<meta content="Hamlet Batista" name="author"/>
Ŀ	
2	<script></script>

Does this affect Googlebot? If it does, does the fix work too?

Let's see!

ر*ه

L in "http://dev.ranksense.com/"

- ? 🥊 🖩 🎨

/missing_body.html

	× Crawled page
Google, but has issues	HTML SCREENSHOT MORE INFO
in Google Search results (if not subject to a mar- vever, some issues prevent it from being eligible	<pre>1 <!DOCTYPE html> 2 <html lang="en"><head> 3 4</head></html></pre>
ED PAGE	<pre>5 <title>This is a test</title> 6 <meta content="This is a test" name="description"/> 7 <meta content="Hamlet</pre></td></tr><tr><td>Indexed, not submitted in</td><td>Batista" name="author"/> 8 9 <link href="http://dev.ranksense.com/missing_body .html" rel="canonical"/></pre>
Page is not mobile frienc	10 11 12 13 <body >
	<pre>15 <body><div>This is a test</div> 16 17</body></pre>

If the page is missing the BODY tag, Google adds it back. Good.

. in "http://dev.ranksense.com/" 📀 🤦

/invalid_head_hard_coded.html

	× Crawled page
	HTML SCREENSHOT MORE INFO
Dogle, but has issues Google Search results (if not subject to a mar rer, some issues prevent it from being eligible	<pre>1 <!DOCTYPE html> 2 <html lang="en"><head> 3 4 5 <title>This is a test</title></head></html></pre>
ED PAGE	<pre>6 <meta content="This a test" name="description"/> 7 <meta content="Hamlet Batista" name="author"/> 9</pre>
Indexed, not submitted in	<pre>9 <body><div>This is shouldn't be here </div> 10</body></pre>
	<pre>11 <link href="http://localhost/" rel="canonical"/> 12 13</pre>
Page is not mobile frienc	14 15 16

If we add a <DIV> manually to the HTML HEAD, Google pushes our canonical to the BODY. Same as in the browser.



If we add a <DIV> to the HTML HEAD using a script, the URL Inspection Tool gives an error and the page doesn't get indexed. The browser can handle the page.

L in "http://dev.ranksense.com/"

? 🦺 🏢

/invalid_head_lazy_fix.html

	× Tested page
	HTML SCREENSHOT MORE INFO
ailable to Google, but has issues xed and selected as canonical, it could appear in ne issues prevent it from being eligible for all en	<pre>1 <!DOCTYPE html> 2 <html lang="en"><head> 3 4 5 <title>This is a test</title> 6 <link <="" pre="" rel="canonical"/></head></html></pre>
) PAGE	<pre>href="http://dev.ranksense.com/invalid_hes _lazy_fix.html" /></pre>
URL can be indexed	Batista" /> 9 10 < script > 11 document.write(" <div>This script is</div>
xed only if <u>certain conditions</u> are met	<pre>misplaced"); //2 <body><div>This script : misplaced</div></body></pre>
	<pre>13 14 <div>This is shouldn't be here </div></pre>

If we move the SEO tags to the top of the HTML and leave the invalid <DIV>, the canonical remains in the HTML HEAD.

Q Inspect any URL in "http://dev.ranksense.com/

http://dev.ranksense.com/invalid_head_lazy_fix.html

URL will be indexed only if certain conditions are met Discovery Not checked in live tests Crawl Jan 26, 2020, 10:50:03 PM Time Googlebot smartphone Crawled as Crawl allowed? Yes Page fetch Successful Indexing allowed? Yes Indexing User-declared canonical http://dev.ranksense.com/invalid_head_lazy_fix.html Google-selected canonical ⑦ Only determined after indexing

?

RL in "http://dev.ranksense.com/

? 📍

III 🌉

m/proper_fix.html

	N/A	×	Crawled	page	
9	None detected		HTML	SCREENSHOT	MORE INFO
known from other so	urces that are currently not re	6	<link r<br=""/> href="htt tml">	el ="canonical" p://dev.ranksense.com	۹ (
	Jan 21, 2020, 6:44:30 PN Googlebot smartphone	7	<meta n<br=""/> a test"> <meta n<br=""/> Batista">	<pre>ame="description" con ame="author" content=</pre>	tent ="This "Hamlet
?	Yes Successful	9 10 11			
ed?	Yes	12 13			
		14 15	<pre><body></body></pre>		
canonical	http://dev.ranksense.cor	16 17	<div></div>	This is a test	
ed canonical ⊘	Inspected URL	18 19	<pre><script d<="" documen="" placed<="" pre=""></script></pre>		

The optimal solution is to move the invalid scripts and tags to the HML BODY.

- Misplaced HTML tags in the HEAD can push SEO tags to the BODY
- The issue is visible in the browser and Search Console URL Inspection Tool
- Moving SEO tags to the top of the HTML HEAD helps

Share these #SMXInsights on your social channels!

(*)

```
t <idoctype html>
  <html lang="en">
   <head>
    <meta charset="utf-8">
    <title>This is a test</title>
    <meta name="description" content="This is a test">
    <meta name="author" content="Hamlet Batista">
10
    <link rel="canonical" href="http://dev.ranksense.com/override canonical.html">
11
12
13
  </head>
14
  <body>
15
16
      <div>This is a test</div>
17
18
      <script src="canonical_script.js"></script>
19
20
21 </body>
22 </html>
```

Let's learn to use the Chrome Debugger to identify obscure scripts that override SEO tags. This test page has a canonical and the script linked overrides it.



Let's use the JavaScript Debugger to track down scripts that override SEO tags.

R	6]	Elements	Console	Sources	Network	Performance	Memory	Application	Security	Audits
	Add a Edit a Delete	ttribute s HTML e element		_id="23"> a test <td>tle></td> <td></td> <td></td> <td></td> <td></td> <td></td>	tle>					
	Сору		Þ	content="Th nt="Hamlet	is is a te: Batista" r	st" rs_id="25": s_id="26">	*			
	Hide of Force	element state	•	f=" <u>http://d</u>	ev.ranksen:	se.com/new_can	onical.html	" rs_id="27"	> == \$0	
	Break	on	×	subtree r	nodificatio	ns				
	Expar Collar	d recursivel se children	у	node rem	modificatio noval	ons t>				
	Scroll Focus	into view								
	Store	as global va	riable							
	Speed	:h	►							

First, we set up a DOM breakpoint to stop JavaScript execution when the attributes of the canonical tag are modified. Next, we hit refresh.



```
RankSense
   Network Performance Memory
                                      Application Security
                                                            Audits
unused_script.js ×
                unused_code.html
 var link = !!document.guerySelector("link[rel='canonical']") ? document.guerySelector("link[rel='canoni
 link.setAttribute('rel', 'canonical');
 link.setAttribute('href', 'http://dev.ranksense.com/new_canonical.html');
 document.head.appendChild(link);
 function add_debug_div() {
   document.write("<div>Debug DIV</div>");
 debug_flag = false;
Line 1, Column 1
```



(*)



We can track unused code using the Code Coverage tool.

R Elements Console Sources Network Performance	ce Memory	Application	Security Audits RankSer	nse	÷×
Page Filesystem >> ; I unused_code.html × unus	ed_script.js			E	u 🔿 🗄 🕆 🗤 💋
unused_code.html {} 2 lines, 36 characters selecter	əd				▶ Watch
Console What's New Coverage ×					×
🥥 C 🛇 🛨 URL filter 🛛 All 🔹 🖂 Conter	nt scripts				
URL	Туре	Total Bytes	Unused Bytes		
http://dev.ranksense.com/unused_script.js	JS	609	109 17.9 %		
http://dev.ranksense.com/unused_code.html	JS	43	16 37,2 %		
527 B of 652 B (81%) used so far. 125 B unused.					

 \bullet

T

RÓ] Ele	ments Co	onsole	Sources	Network	Performance	Memory	Application	Security	Audits	RankSens
Page	Filesyst	tem »		I un	used_code.h	tml unused_	script.js ×				
* 🗆 te * 🔾	op dev.rank unuser unuser	sense.com d_code.html d_script.js		10 11 12 13 14 15 16 17 18 19 20 21	<pre>link.setAt document.f function e document } debug_flag</pre>	ttribute('href head.appendChi add_debug_div(t.write(" <div> g = false;</div>	', 'http:// ld(link);) { Debug DIV </td <td>/dev.ranksense /div>");</td> <td>.com/new_ca</td> <td>nonical.</td> <td>html');</td>	/dev.ranksense /div>");	.com/new_ca	nonical.	html');
				{} Li	ne 1, Column						
: c	onsole	What's New	v Cov	erage ×							
• 0	0 0	🛨 URL	filter	4	NI ¥	Content s	ripts				
URL						Tj	pe	Total Bytes	Unused By	tes	
http://de	ev.rankse	nse.com/unu	used_scrip	Ljs		J	3	60	9	10	9 17.9 % 0
http://de	ev.rankse	nse.com/unu	ised_code	.html		J.	3	4	3	1	8 37.2 %

The red vertical bars highlight the unused code.

Y

TL;DR

- By default, CSS is treated as a render blocking resource.
- Media types and media queries allow us to mark some CSS resources as non-render blocking.
- The browser downloads all CSS resources, regardless of blocking or non-blocking behavior.





🕞 🔂 Elements Console	Sources	Network Perform	ance Merno	ry Application	Security Aud	lits Ra	inkSense							×			
Page Filesystem >> :	II un	used_CSS.html ×						E	ш 7				7/0	ø			
≠ 🗖 top	1 5	!doctype html>							Watc Watc	h							
✓	2 3	<pre>html lang="en"> </pre>															
unused_CSS.html	4	Required me</td <td>ta tags></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td colspan="6">Not paused</td>	ta tags>						Not paused								
▹	6 7	<pre><meta thatset="utl=s"/> <meta content="width=device-width, initial-scale=1, shrink-to-fit=no" name="viewport"/></pre>															
	8	Bootstrap (<br <link btn-grou<="" href="</td><td>https://stackpat</td><td>th.bootstrapcdn.</td><td>com/boot</td><td>tstrap/4.3.1/css/bootstrap.min.</td><td>less</td><td></td><td></td><td colspan=5>Not paused</td></tr><tr><td></td><td>10</td><td></td><td>-141-2/4/4141-05</td><td></td><td></td><td></td><td></td><td></td><td>v Break</td><td colspan=4>Breakpoints</td><td></td></tr><tr><td></td><td>11</td><td></head></td><td>rug:</title></td><td></td><td></td><td></td><td></td><td></td><td></td><td>No</td><td>o breał</td><td>kpoints</td><td colspan=4></td></tr><tr><td></td><td>13
14</td><td><body>
<h1>Hello, world</td><td>!!</h1></td><td></td><td></td><td></td><td></td><td></td><td>> XHR/</td><td>fetch I</td><td>Breakp</td><td>points</td><td></td><td></td></tr><tr><td></td><td>15</td><td>div class=" rel="style</td><td>SS>
sheet" td=""/> <td>o" cole="orou</td> <td>p" aria-label="</td> <td>asic example"></td> <td></td> <td></td> <td></td> <td>T DOM</td> <td>Break</td> <td>kpoints</td> <td>s</td> <td></td> <td></td>	o" cole="orou	p" aria-label="	asic example">				T DOM	Break	kpoints	s					
	17				Left					10Ne	o break	kpoints					
	{} u	ne 1, Column 1							+ Globa	al Liste	eners						
Console What's New Cov	erage ×													×			
🥚 C 🛇 🗄 🛓 URL filter	10	All 🔻 🗍 🔲 Cor	tent scripts														
URL			Туре	Total Bytes	Unused Bytes												
https://stackpath.bootstrapcdn.com/bo	otstrap/4.3	.1/css/bootstrap.min.c	s CSS	155 758	152	895 98.2	2 96										
https://stackpath.bootstrapcdn.com/bo	otstrap/4.3	.1/js/bootstrap.min.js	JS (coarse)	58 072	57	247 98.6	3 %										
https://code.jquery.com/jquery-3.3.1.slim.min.js J				69 917	50	864 72.7	7 %										
https://cdnjs.cloudflare.com/ajax/libs/p	opper.js/1.1	4.7/umd/popper.min.js	JS (coarse)	21 004	20	787 99.0) %										
http://dev.ranksense.com/unused_scrip	t.js		JS	609		109 17.9	9 % L										
22.9 KB of 298 KB (8%) used so far. 2	75 KB unus	sed.															

 \bullet

T

- The JavaScript Debugger can help track down the scripts that override SEO tags and cause performance issues
- The Code Coverage tool helps identify JavaScript and CSS code that is never used so that we can remove it

Share these #SMXInsights on your social channels!

、・カ

1	<idoctype html=""></idoctype>
2	
3	<html lang="en"></html>
- 4	<head></head>
5	<meta charset="utf-8"/>
6 7	<title>This is a test</title>
8	<pre><meta content="This is a test" name="description"/></pre>
9	<meta content="Hamlet Batista" name="author"/>
10	
11	k rel="canonical" href="http://dev.ranksense.com/override_canonical.html";
12	10.0
13	
15	<body></body>
16	
17	<div>This is a test</div>
18	Corrint arc="canonical scrint is">
20	Souther and Conversion of Strate
21	
22	

Splitting Code to Increase Performance

Let's learn to use the Chrome Debugger to identify obscure scripts that override SEO tags.

This test page has a canonical and the script linked overrides it.



Wrapping up.

Who am I and where am I from?



I'M #ZORASDAD

First and foremost.



MY NAME IS MIKE KING

Razorfish, Publicis Modem alum Full Stack Developer Full Stack Marketer Moz Associate



















Content Strategy SEO

Paid Media

Machine Learning

Marketing Automation Measurement & Optimization

D

1

Machine Learning for Marketers Guide



PIPULLRANK

Machine Learning for Marketers

A COMPREHENSIVE GUIDE TO MACHINE LEARNING



http://ipullrank.com/machine-learning-guide/







Q&A / Thank You!

Twitter:@ipullrankagencyEmail:hello@ipullrank.comWebsite:https://ipullrank.com

Get in touch if you need some help!



Mike King Managing Director at iPullRank Hamlet Batista CEO at RankSense

 \checkmark