ALLAN THRAEN | 4 years ago | PDF |

Information Retrieval Optimizely (Episerver) CMS

PUBLICWWW - SEARCHING FOR INTERESTING EPISERVER CMS USE PATTERNS



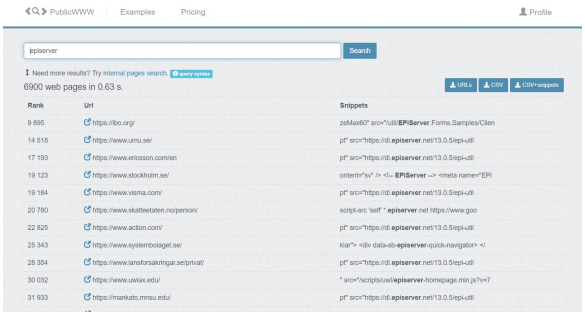
I recently discovered publicwww.com a cool service that lets you search for any text in the html/css/js of all it's 550 million (2019-05-09) indexed web pages, including the cookies sent out and the http header. In this post I put my Episerver goggles on and had some fun with this data.

First of all a bit of a disclaimer: I'm in no way associated with the publicwww.com service, but for a recent project I needed to produce a few lists of web sites using certain technologies and I found it very handy. I bought a single month of pro-subscription $49 (the free version is useful but will only give you partial results) and now that I had that, I figured I might as well have some fun with it and examine some Episerver usage online!

Identifying Episerver

Now, this is a fun one - what could I possibly search for to identify an Episerver site?
Publicwww comes with a bunch of great examples on how to identify a wide number of both CMS's and E-commerce solutions, as well as analytics and marketing technologies. But Episerver does not seem to be listed there. And for good reason. It's not really supposed to identify if a site is running Episerver. I'm guessing that's for security reasons - if, for example, a vulnerability is detected and not everybody patches it right away it wouldn't be a good idea to simply let the hackers look for a certain Generator tag to know which sites were vulnerable, right?!

So - even though Episerver sites might have certain url patterns or html elements that are commonly used, they can (and probably should) be customized in the implementations. This being said, it's still fun to experiment a little.



First off, I tried to simply search for "episerver" to see what kind of results I got. It returned 6900 sites, but looking at the results - although a high number of them do appear to be Episerver sites I'm definitely not getting all the Episerver based sites - and I also get sites that are not Episerver sites, but merely mentioning Episerver - like wordpress based developer blogs (you know who you are!).

Another approach I thought of was to use the fact that Episerver often likes to put assets in a url beneath /globalassets/. However that yielded 14858 results - including a lot of noise (apperently that path is popular in other systems as well).

Searching for a Generator meta-tag with Content="EPiServer" does return 2500+ results. Now - while I'm pretty certain that they have all been Episerver sites at one point in time I'm not quite sure how often the publicwww index is refreshed - because when I tested a few of them they had clearly changed their html significantly and no longer exposed that tag. So - keep that in mind for the remainder of this blog post - I'm writing about what's in the index - not the reality of today.

Technology through time

Sometimes it can be fun to do a bit of internet archaeology. Like - for example I started looking for some sites using older technology.

Webforms based sites are easy to spot - and using some tricks from the paragraph we can easily pull out 2700+ webforms based sites that are very likely based on Episerver.

Do you remember IIS 6.0? It was included with Windows Server 2003 (released in April 2003, just a few months after the Columbia space shuttle disaster). Well - I guess Microsoft made quality back then because it still seems to be in use on 52(!) Episervers sites in the index. (I spotted a few on IIS 5, as well, but a closer examination just reveals an outdated index and those sites are now resting comfortably on an IIS 8.5).

Back in the day - before the wonderful world of HTML5 - brave pioneers that wanted to do things right relied on the XHTML standard! And it's nice to see a few that has hung on. At least 2770 it would seem.

Feature Usage

There are many fun things to examine - things like how many have built responsive websites on Episerver, which technologies are they typically accompanying it with, which sites uses Episerver with Angular and who has taken the trouble ofimplementing the Dublic Core standard of meta-

tags (23).

But I have also found it to be quite interesting to have a closer look at some of the add-ons and features available in Episerver - especially for a few of the 'babies' that I feel a particular connection with :-)

I was for instance quite amazed at how many sites in the index are linking to the public Episerver Find javascript library. And its great to see all the sites building cool editorial forms using Episerver forms.

For some features, it can be pretty useful to look for cookies as well - like if we want to spot 500+ sites using visitor group personalization based on how many visits you have made to the site. Or maybe even check out a few of theawesome first movers implementing Episerver tracking.

When it goes wrong

Let me round of this post with a few fun ones (ok, I'll admit to having geeky humor - not everyone might find these ones funny).

Keeping track of licenses can be hard. And accidentally putting a development license into production can happen to the best of us. I wouldn't dream of pointing any fingers - but I chuckle a bit of at least one of the names on the list :-)

Episerver has the concept of 'internal links' and 'external / friendly' urls. Now, in an ideal world a search engine like publicwww that indexes publicly available sites would only catch external / friendly urls. But - you guessed it - that's not always the case. Like for these 70 sites - something seems to have gone wrong. And we can only guess if it's an editor that has gone crazy with copying and pasting - or maybe a developer that got confused when generating a link. I've seen both cases happen.

The last one has me baffled a bit - cause on 5 would seem that on a few sites, somehow the javascript for the editorial menu is actually being indexed (maybe even shown?) - maybe there are a few permissions that could use a tweak... (and no - I'm not giving you the link for that search).

Do you have any suggestions for fun searches to try? Share them in the comments below so we can all enjoy!

Information Retrieval | Optimizely (Episerver) | CMS

RECENT POSTS

CodeArt ApS
Teknikerbyen 5, 2830 Virum, Denmark
Email: info@codeart.dk
Phone: +45 26 13 66 96
CVR: 39680688

Copyright © 2024