

How can I follow a trail of 302 redirects with MSXML2.XMLHTTP?

Source:

<http://www.tech-archive.net/Archive/Scripting/microsoft.public.scripting.vbscript/2006-06/msg00129.html>

- *From:* doug@xxxxxxxxxx
 - *Date:* 6 Jun 2006 13:38:36 -0700
-

Hi all ...

I have an interesting one here. I have some code that I've been using for years to pull down HTTP content as needed. However, I've just run into a situation that I've never seen before, and can't quite figure a way around ...

Background: If a URL fed to MSXML2.XMLHTTP via VBScript returns a 302 "redirect" when loaded – VBScript will crap out on the page load due to security settings in IE, if the redirected URL is in another domain. You won't get any content back as a part of the response, and the HTTP status code actually comes back as "0". However ... if you enable the "Access data sources across domains" option in the security settings for IE – the MSXML2.XMLHTTP control will follow the 302 path and eventually get the final result page and return the contents to you.

(/RANT-ON ... I shouldn't have to change settings in IE to control how my VBScript program will work! Ugh! /RANT-OFF)

However, I would like to detect this 302 activity and track it from the starting page to the ending page. For instance, here's WGET tracking a two-layer redirect from a Feedburner link that demonstrates the situation:

1. WGET <http://feeds.geffen.com/artist/robzombie/content?m=145>
HTTP request sent, awaiting response... 302 Moved Temporarily
Location:
<http://feeds.feedburner.com/~r/artist/robzombie/content/~0/145>
[following]

2. WGET <http://feeds.feedburner.com/~r/artist/robzombie/content/~0/145>
HTTP request sent, awaiting response... 302 Moved Temporarily
Location:
http://www.geffen.com/artist/news/default.aspx/nid/8031/aid/414/?utm_source=rss&utm_campaign=rss&utm_medium=News&utm_content=nid_8031
[following]

How can I follow a trail of 302 redirects with MSXML2.XMLHTTP?

3. WGET

```
http://www.geffen.com/artist/news/default.aspx/nid/8031/aid/414/?u  
tm_source=rss&utm_campaign=rss&utm_medium=News&utm_content=nid_8031  
Resolving www.geffen.com... 192.251.67.164  
Connecting to www.geffen.com|192.251.67.164|:80... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 8,294 (8.1K) [text/html]
```

Here's some code I put together, to try to loop through each redirect until a final non-302 result is returned ... and state what the final link was. But I can't seem to find any way to get this concept to work based on what MSXML2.XMLHTTP provides me, given its interaction with Internet Explorer. Basically, if I enable the "Access data sources across domains" the control will simply follow the 302's and not tell me about it, and return the final page but with a link that (technically) doesn't match. If I disable the option, the GET performed by XMLHTTP craps out, returns a 0 result code.

Any thoughts on a way to accomplish this? Is it even possible, or am I trying to do something that is completely outside of the control provided by MS?

Thanks in advance ...

```
Const READYSTATE_UNINITIALIZED = 0, _  
READYSTATE_INITIALIZED = 1, _  
READYSTATE_LOADED = 2, _  
READYSTATE_INTERACTIVE = 3, _  
READYSTATE_COMPLETE = 4  
  
MaxLoadTime = 20  
MaxLoops = 10  
url = "http://feeds.geffen.com/artist/robzombie/content?m=145  
  
StatusResult = 302  
LoopCount = 0  
  
wscript.echo "START: " & url  
  
Do while loopcount < maxloops and statusresult = 302  
  
loopcount = loopcount + 1  
  
CheckPageLoadTime = 0  
PageLoadStartTime = now()  
  
Set objHTTP = CreateObject("MSXML2.XMLHTTP.3.0")  
Call objHTTP.Open("GET", url, TRUE)  
objHTTP.Send  
  
Do Until objHTTP.ReadyState = READYSTATE_COMPLETE or CheckPageLoadTime
```

How can I follow a trail of 302 redirects with MSXML2.XMLHTTP?

How can I follow a trail of 302 redirects with MSXML2.XMLHTTP?

MaxLoadTime

wscript.sleep 100

PageLoadEndTime = now()

CheckPageLoadTime = DateDiff("s", PageLoadStartTime, PageLoadEndTime)

Loop

HTTPGET = objHTTP.ResponseText

StatusResult = objHTTP.Status

HTTPRedirectLocation = objHTTP.GetResponseHeader("Location")

If CheckPageLoadTime > MaxLoadTime then

wscript.echo "TIMEOUT"

end if

IF StatusResult = 302 then

wscript.echo "REDIRECT: " & HTTPRedirectLocation

url = HTTPRedirectionLocation

end if

loop

wscript.echo "END: " & url

wscript.echo

wscript.echo "HTTP Status: " & StatusResult

wscript.echo "Redirects: " & loopcount

Here's my results with the security option set to disabled:

START: <http://feeds.geffen.com/artist/robzombie/content?m=145>

END: <http://feeds.geffen.com/artist/robzombie/content?m=145>

HTTP Status: 0

Redirects: 10

And here's my result with the security option set to enabled:

START: <http://feeds.geffen.com/artist/robzombie/content?m=145>

END: <http://feeds.geffen.com/artist/robzombie/content?m=145>

HTTP Status: 200

Redirects: 1

So, to me, it seems as though it is possible to trace the string of

How can I follow a trail of 302 redirects with MSXML2.XMLHTTP?

How can I follow a trail of 302 redirects with MSXML2.XMLHTTP?

302's through the MSXML2.XMLHTTP control??

Thanks in advance....

•