

---

Subject: Re: Download files from the web

Posted by on Tue, 17 Dec 2013 09:39:42 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Den måndagen den 16:e december 2013 kl. 22:27:02 UTC+1 skrev Mike Galloy:

> On 12/16/13, 7:46 AM, Mats Ljfdahl wrote:

>> Den måndagen den 16:e december 2013 kl. 15:14:10 UTC+1 skrev Mats Ljfdahl:

>>>

>>> Thanks. But it seems it has the same problem as the wget  
>>> function, in that it can't tell the difference between a proper  
>>> download and a 404 error web page.

>>>

>>> I simplified your code a bit (because I don't need the progress  
>>> bar) and came up with this:

>>>

>>> function downloadurl, url, file

>>>

>>> url\_scheme = (strsplit(url, ':',/extract))[0]

>>>

>>> url\_hostname = strjoin((strsplit(url, '/',/extract))[1:\*,'/'])

>>>

>>> oUrl = OBJ\_NEW('IDLnetUrl', URL\_SCHEME = url\_scheme, URL\_HOSTNAME =  
>>> url\_hostname)

>>>

>>> retrievedFilePath = oUrl->Get(FILENAME=file)

>>>

>>> oUrl->GetProperty, RESPONSE\_CODE=RespCode ; 200 = OK

>>>

>>> oUrl->CloseConnections

>>>

>>> OBJ\_DESTROY, oUrl

>>>

>>> return, RespCode eq 200 ; True if OK

>>>

>>> end

>>>

>>> I tried it both with a url pointing to an existing web page and to  
>>> a non-existing page. In both cases I get RespCode eq 200. With the  
>>> non-existing page I again had downloaded a 404 error page.

>>>

>>> I got the value 200 for OK from the list at

>>> [http://www.exelisvis.com/docs/IDLnetURL.html#objects\\_network\\_1009015\\_1417867](http://www.exelisvis.com/docs/IDLnetURL.html#objects_network_1009015_1417867)

>

>> Thought it might work better to use spawn and wget and read its exit  
>> status. But that seems to have the same problem: 404 error page  
>> downloaded in case the remote file doesn't exist, but exit status 0  
>> (=OK) regardless.

```

>
>> So this does not seem to be an IDL problem. It is just hard to get
>> the information I want from the download process.
>
>> The web server obviously knows the requested file does not exist but
>> isn't there a way to make it tell the downloading process this in a
>> more condensed way than constructing a web page with a 404 error?
>
> I use IDLnetURL in my routine and it is able to tell if there is a 404
> error:
>
> IDL> c = mg_get_url_content('http://michaelgalloy.com/nothing', $
> IDL>          error_message=em, $
> IDL>          response_code=rc, response_header=rh)
> IDL> help, em
> EM          STRING    = 'IDLNETURL::GET: CCurlException: Error:
> Http Get Request Fai'...
> IDL> help, rc
> RC          LONG      =      404

```

Aha. This is an important clue! It seems to be a property of the web server and not of the way we try to download. With your url, <http://michaelgalloy.com/nothing>, I also get response code = 404 with the code I wrote but with, e.g, <http://www.exelisvis.com/docs/nothing>, I get 200.

I guess I should really try it on the server I will be downloading from. Problem is just that it is down right now and it is remote enough that nobody is there to turn it on until Christmas. I also have some influence over that server so I should be able to request that it is set up so that it returns the proper code in case the file does not exist.

Hm. But how to test what happens when the file exists on the server but does not get downloaded properly? Would be great if one could get the file size or even better a checksum from the server and then compare that with the same property of the downloaded file. But I don't see any of those properties listed here [http://www.exelisvis.com/docs/IDLnetURL\\_Properties.html](http://www.exelisvis.com/docs/IDLnetURL_Properties.html).