
Subject: Download files from the web

Posted by on Mon, 16 Dec 2013 12:37:11 GMT

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

I need to make an idl program download a couple of text files from the web. I found the webget() function from astrolib, now also distributed with idl, see <http://www.exelisvis.com/docs/webget.html>

My problem with that is that I don't know if the download succeeded. If I give a non-existing URL, I get the 404 error page downloaded and everything looks fine. If I set the COPYFILE keyword to some local file name, the file gets stored there and instead of the file contents, the return value is a scalar long. The web page does not say how this number should be interpreted but it seems to be unity whether or not the URL was really valid.

I guess I could search the downloaded file for some variation of "404 - Page not found" but I don't know how much that string varies from web server to web server. And it seems a hassle anyway.

I had a look at <http://www.exelisvis.com/docs/socket.html>, too. Promisingly it has an ERROR keyword but I don't understand how to tell it which file I want. Seems possible only to specify the host but not which file on the host. Near the bottom of the socket.html page there are two links to pages that promise to tell me how to read web pages and access ftp servers through socket, but when I click on them I get "Article does not exist or Permission Denied" errors.

So, how does one make idl download files from the web - and tell you if it worked?

Subject: Re: Download files from the web

Posted by [David Fanning](#) on Mon, 16 Dec 2013 12:46:00 GMT

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

Mats Löfdahl writes:

> I need to make an idl program download a couple of text files from the web. I found the webget() function from astrolib, now also distributed with idl, see <http://www.exelisvis.com/docs/webget.html>

>

> My problem with that is that I don't know if the download succeeded. If I give a non-existing URL, I get the 404 error page downloaded and everything looks fine. If I set the COPYFILE keyword to some local file name, the file gets stored there and instead of the file contents, the return value is a scalar long. The web page does not say how this number should be interpreted but it seems to be unity whether or not the URL was really valid.

>

> I guess I could search the downloaded file for some variation of "404 - Page not found" but I don't know how much that string varies from web server to web server. And it seems a hassle anyway.

>

> I had a look at <http://www.exelisvis.com/docs/socket.html>, too. Promisingly it has an ERROR keyword but I don't understand how to tell it which file I want. Seems possible only to specify the host but not which file on the host. Near the bottom of the socket.html page there are two links to

pages that promise to tell me how to read web pages and access ftp servers through socket, but when I click on them I get "Article does not exist or Permission Denied" errors.

>

> So, how does one make idl download files from the web - and tell you if it worked?

It seems like FILE_TEST might be helpful.

Cheers,

David

--

David Fanning, Ph.D.

Fanning Software Consulting, Inc.

Coyote's Guide to IDL Programming: <http://www.idlcoyote.com/>

Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: Download files from the web

Posted by [Helder Marchetto](#) on Mon, 16 Dec 2013 12:47:14 GMT

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

On Monday, December 16, 2013 1:37:11 PM UTC+1, Mats Löfdahl wrote:

> I need to make an idl program download a couple of text files from the web. I found the webget() function from astrolib, now also distributed with idl, see

<http://www.exelisvis.com/docs/webget.html>

>

>

>

> My problem with that is that I don't know if the download succeeded. If I give a non-existing URL, I get the 404 error page downloaded and everything looks fine. If I set the COPYFILE keyword to some local file name, the file gets stored there and instead of the file contents, the return value is a scalar long. The web page does not say how this number should be interpreted but it seems to be unity whether or not the URL was really valid.

>

>

>

> I guess I could search the downloaded file for some variation of "404 - Page not found" but I don't know how much that string varies from web server to web server. And it seems a hassle anyway.

>

>

>

> I had a look at <http://www.exelisvis.com/docs/socket.html>, too. Promisingly it has an ERROR keyword but I don't understand how to tell it which file I want. Seems possible only to specify the host but not which file on the host. Near the bottom of the socket.html page there are two links to

pages that promise to tell me how to read web pages and access ftp servers through socket, but when I click on them I get "Article does not exist or Permission Denied" errors.

>

>

>

> So, how does one make idl download files from the web - and tell you if it worked?

Hi Mats,

if it helps, i used the IDLnetUrl object and then use the getProperty method to get the Response_code value. Not sure if it helps, but I have been downloading files successfully with https. I also use the callback_function to make a progress bar.

Not sure if it helps, but it might be a place to start...

Cheers,

H

Subject: Re: Download files from the web

Posted by on Mon, 16 Dec 2013 13:24:59 GMT

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

Den måndagen den 16:e december 2013 kl. 13:46:00 UTC+1 skrev David Fanning:

> Mats Löfdahl writes:

>

>> I need to make an idl program download a couple of text files from the web. I found the webget() function from astrolib, now also distributed with idl, see <http://www.exelisvis.com/docs/webget.html>

>

>> My problem with that is that I don't know if the download succeeded. If I give a non-existing URL, I get the 404 error page downloaded and everything looks fine. If I set the COPYFILE keyword to some local file name, the file gets stored there and instead of the file contents, the return value is a scalar long. The web page does not say how this number should be interpreted but it seems to be unity whether or not the URL was really valid.

>

>> I guess I could search the downloaded file for some variation of "404 - Page not found" but I don't know how much that string varies from web server to web server. And it seems a hassle anyway.

>

>> I had a look at <http://www.exelisvis.com/docs/socket.html>, too. Promisingly it has an ERROR keyword but I don't understand how to tell it which file I want. Seems possible only to specify the host but not which file on the host. Near the bottom of the socket.html page there are two links to pages that promise to tell me how to read web pages and access ftp servers through socket, but when I click on them I get "Article does not exist or Permission Denied" errors.

>

>> So, how does one make idl download files from the web - and tell you if it worked?

>

> It seems like FILE_TEST might be helpful.

I'm not sure if you mean I should use it locally after downloading to see if the file was downloaded or remotely to see if the file is there before I try to download it.

If the locally, it would find *a* file also if that file is only partially downloaded or if it is a 404 error html file.

If the remotely, I guess it is nice to know that the file exists on the server but that does not guarantee that I will be able to download it. And, although I see there is a SOCKET keyword that I didn't know about, I still don't know how to specify the remote file with the socket command.

Subject: Re: Download files from the web

Posted by [David Fanning](#) on Mon, 16 Dec 2013 13:29:38 GMT

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

Mats Löfdahl writes:

> I'm not sure if you mean I should use it locally after downloading to see if the file was downloaded or remotely to see if the file is there before I try to download it.

>

> If the locally, it would find *a* file also if that file is only partially downloaded or if it is a 404 error html file.

>

> If the remotely, I guess it is nice to know that the file exists on the server but that does not guarantee that I will be able to download it. And, although I see there is a SOCKET keyword that I didn't know about, I still don't know how to specify the remote file with the socket command.

I meant after the file was downloaded to tell if it was there, was the right size, etc. But, I think Helder's suggestion to use the functionality of the IDLnetURL object is likely to be more useful.

Cheers,

David

--

David Fanning, Ph.D.

Fanning Software Consulting, Inc.

Coyote's Guide to IDL Programming: <http://www.idlcoyote.com/>

Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: Download files from the web

Posted by on Mon, 16 Dec 2013 13:32:36 GMT

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

Den måndagen den 16:e december 2013 kl. 13:47:14 UTC+1 skrev Helder:

> On Monday, December 16, 2013 1:37:11 PM UTC+1, Mats Löfdahl wrote:

>

>> I need to make an idl program download a couple of text files from the web. I found the webget() function from astrolib, now also distributed with idl, see <http://www.exelisvis.com/docs/webget.html>

>

>>

>

>>

>

>>

>

>> My problem with that is that I don't know if the download succeeded. If I give a non-existing URL, I get the 404 error page downloaded and everything looks fine. If I set the COPYFILE keyword to some local file name, the file gets stored there and instead of the file contents, the return value is a scalar long. The web page does not say how this number should be interpreted but it seems to be unity whether or not the URL was really valid.

>

>>

>

>>

>

>>

>

>> I guess I could search the downloaded file for some variation of "404 - Page not found" but I don't know how much that string varies from web server to web server. And it seems a hassle anyway.

>

>>

>

>>

>

>>

>

>> I had a look at <http://www.exelisvis.com/docs/socket.html>, too. Promisingly it has an ERROR keyword but I don't understand how to tell it which file I want. Seems possible only to specify the host but not which file on the host. Near the bottom of the socket.html page there are two links to pages that promise to tell me how to read web pages and access ftp servers through socket, but when I click on them I get "Article does not exist or Permission Denied" errors.

>

>>

>

>>

>

>>

>

>> So, how does one make idl download files from the web - and tell you if it worked?

>
>
>
> Hi Mats,
>
> if it helps, i used the IDLnetUrl object and then use the getProperty method to get the
Response_code value. Not sure if it helps, but I have been downloading files successfully with
https. I also use the callback_function to make a progress bar.
>
> Not sure if it helps, but it might be a place to start...

Looks promising.

So I would create an instance of the class, initialize it, download with the get method, get the
RESPONSE_CODE (anything but zero is a fail?) through the getproperty method, and then
destroy the class instance?

Subject: Re: Download files from the web
Posted by [Helder Marchetto](#) on Mon, 16 Dec 2013 13:41:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Monday, December 16, 2013 2:32:36 PM UTC+1, Mats Löfdahl wrote:

> Den måndagen den 16:e december 2013 kl. 13:47:14 UTC+1 skrev Helder:

>

>> On Monday, December 16, 2013 1:37:11 PM UTC+1, Mats Löfdahl wrote:

>

>>

>

>>> I need to make an idl program download a couple of text files from the web. I found the
webget() function from astrolib, now also distributed with idl, see
<http://www.exelisvis.com/docs/webget.html>

>

>>

>

>>>

>

>>

>

>>>

>

>>

>

>>>

>

>>

>

>>> My problem with that is that I don't know if the download succeeded. If I give a non-existing

URL, I get the 404 error page downloaded and everything looks fine. If I set the COPYFILE keyword to some local file name, the file gets stored there and instead of the file contents, the return value is a scalar long. The web page does not say how this number should be interpreted but it seems to be unity whether or not the URL was really valid.

```
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> I guess I could search the downloaded file for some variation of "404 - Page not found" but I
don't know how much that string varies from web server to web server. And it seems a hassle
anyway.
```

```
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>>
>
>>
>
>>> I had a look at http://www.exelisvis.com/docs/socket.html, too. Promisingly it has an ERROR
keyword but I don't understand how to tell it which file I want. Seems possible only to specify the
host but not which file on the host. Near the bottom of the socket.html page there are two links to
pages that promise to tell me how to read web pages and access ftp servers through socket, but
when I click on them I get "Article does not exist or Permission Denied" errors.
```

```

>
>>>
>
>>
>
>>>
>
>>
>
>>> So, how does one make idl download files from the web - and tell you if it worked?
>
>>
>
>>
>
>>
>
>> Hi Mats,
>
>>
>
>> if it helps, i used the IDLnetUrl object and then use the getProperty method to get the
Response_code value. Not sure if it helps, but I have been downloading files successfully with
https. I also use the callback_function to make a progress bar.
>
>>
>
>> Not sure if it helps, but it might be a place to start...
>
>
>
> Looks promising.
>
>
>
> So I would create an instance of the class, initialize it, download with the get method, get the
RESPONSE_CODE (anything but zero is a fail?) through the getproperty method, and then
destroy the class instance?

```

Hi Mats,
here is what I do. Not sure if it solves the problem of a missing file.

```

FUNCTION UrlBigFileGetCallbackStatus, status, progress, oProgressbar
  IF progress[0] THEN oProgressbar->Update, 100.0*progress[2]/progress[1]
  print, 'Check for update: '+status
  return, 1
END

```



```
PRO TEST_Download
CurrDir = 'E:\MySoftware\'
oProgressbar = Obj_New('cgprogressbar', TITLE='Downloading FileName.sav...')
oUrl = OBJ_NEW('IDLnetUrl', URL_SCHEME = 'http', URL_HOSTNAME =
'www.#####/FileName.sav', URL_USERNAME = 'user', URL_PASSWORD = 'pwd',
CALLBACK_FUNCTION = 'UrlBigFileGetCallbackStatus', CALLBACK_DATA=oProgressbar)
VersionFileName = CurrDir+'FileName.sav'
oProgressbar->Start
retrievedFilePath = oUrl->Get(FILENAME=VersionFileName)
oUrl->GetProperty, RESPONSE_CODE=RespCode
oUrl->CloseConnections
OBJ_DESTROY, oUrl
oProgressbar->Destroy
END
```

The output I get is the following:

Check for update: Sending Http Get Request:

Check for update: http://www.#####/FileName.sav

Check for update: Http: get: received (483), total expected (483)

Check for update: Http: get: received (483), total expected (483)

Check for update: Http: get: received (1208), total expected (15501460)

Check for update: Http: get: received (834656), total expected (15501460)

Check for update: Http: get: received (2173400), total expected (15501460)

Check for update: Http: get: received (3384368), total expected (15501460)

Check for update: Http: get: received (4622924), total expected (15501460)

Check for update: Http: get: received (5867288), total expected (15501460)

Check for update: Http: get: received (7105844), total expected (15501460)

Check for update: Http: get: received (8354564), total expected (15501460)

Check for update: Http: get: received (9594572), total expected (15501460)

Check for update: Http: get: received (10831676), total expected (15501460)

Check for update: Http: get: received (12023768), total expected (15501460)

Check for update: Http: get: received (13295720), total expected (15501460)

Check for update: Http: get: received (14551700), total expected (15501460)

Check for update: Http: get: received (15501460), total expected (15501460)

Check for update: Http Get response written to:

Check for update: E:\MySoftware\FileName.sav

Check for update: Http Get completed.

When I download a big file the progress bar does not update after a few seconds. Any idea why?
(Actually any widget stops updating...)

Cheers,
Helder

Subject: Re: Download files from the web

Posted by [David Fanning](#) on Mon, 16 Dec 2013 14:10:33 GMT

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

Helder writes:

> When I download a big file the progress bar does not update after a few seconds. Any idea why? (Actually any widget stops updating...)

You might try adding a EMPTY command after updating to flush the graphics buffer.

Cheers,

David

--

David Fanning, Ph.D.

Fanning Software Consulting, Inc.

Coyote's Guide to IDL Programming: <http://www.idlcoyote.com/>

Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: Download files from the web

Posted by on Mon, 16 Dec 2013 14:14:10 GMT

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

Den måndagen den 16:e december 2013 kl. 14:41:08 UTC+1 skrev Helder:

>
>>> if it helps, i used the IDLnetUrl object and then use the getProperty method to get the Response_code value. Not sure if it helps, but I have been downloading files successfully with https. I also use the callback_function to make a progress bar.
>
>>> Not sure if it helps, but it might be a place to start...

Thanks. But it seems it has the same problem as the webget 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

Subject: Re: Download files from the web

Posted by on Mon, 16 Dec 2013 14:46:05 GMT

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

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

> Den måndagen den 16:e december 2013 kl. 14:41:08 UTC+1 skrev Helder:

>
>>>> if it helps, i used the IDLnetUrl object and then use the getProperty method to get the Response_code value. Not sure if it helps, but I have been downloading files successfully with https. I also use the callback_function to make a progress bar.

>
>>>> Not sure if it helps, but it might be a place to start...

>
> 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
```

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?

Subject: Re: Download files from the web
Posted by [Michael Galloy](#) on Mon, 16 Dec 2013 21:27:02 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

```
> Den måndagen den 16:e december 2013 kl. 15:14:10 UTC+1 skrev Mats
> Löfdahl:
>> Den måndagen den 16:e december 2013 kl. 14:41:08 UTC+1 skrev
>> Helder:
>>
>>>> > if it helps, i used the IDLnetUrl object and then use the
>>>> > getProperty method to get the Response_code value. Not sure
>>>> > if it helps, but I have been downloading files successfully
>>>> > with https. I also use the callback_function to make a
>>>> > progress bar.
>>
>>>> > Not sure if it helps, but it might be a place to start...
>>
>> 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
>
>>
> 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: CUrlException: Error:
Http Get Request Fai'...
IDL> help, rc
RC      LONG    =      404
IDL> help, rh
RH      STRING  = 'HTTP/1.1 404 Not Found
Date: Mon, 16 Dec 2013 21:24:34 GMT
'...
```

It's available on my library Github page: github.com/mgalloy/mglib.

Mike

--

Michael Galloy

www.michaelgalloy.com

Modern IDL: A Guide to IDL Programming (<http://modernidl.idldev.com>)

Research Mathematician

Tech-X Corporation

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 webget

>>> 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
>
>> 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.

Subject: Re: Download files from the web

Posted by [Helder Marchetto](#) on Tue, 17 Dec 2013 09:51:07 GMT

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

On Tuesday, December 17, 2013 10:39:42 AM UTC+1, Mats Löfdahl wrote:

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

>

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

>

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

>

>>>>

>

>>>> Thanks. But it seems it has the same problem as the webget

>

>>>> 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
>
>>
>
>>> 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.

Hi Mats,

the files size is given in my example. Look at the output that I print out. For example:

Check for update: Http: get: received (9594572), total expected (15501460)

A checksum would be great, but at the moment (as far as I know), the size will have to do...

Regards,

Helder

Subject: Re: Download files from the web

Posted by [Helder Marchetto](#) on Tue, 17 Dec 2013 09:53:06 GMT

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

On Tuesday, December 17, 2013 10:51:07 AM UTC+1, Helder wrote:

> On Tuesday, December 17, 2013 10:39:42 AM UTC+1, Mats Löfdahl wrote:

>

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

>

>>

>

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

>

>>

>

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

>

>>

>

>>>> >

```

>
>>
>
>>>> > 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
>
>>
>
>>>>
>
>>
>
>>>> 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.
>
>
>
> Hi Mats,
>
> the files size is given in my example. Look at the output that I print out. For example:
>
> Check for update: Http: get: received (9594572), total expected (15501460)
>
>
>
> A checksum would be great, but at the moment (as far as I know), the size will have to do...
>
>
>
> Regards,
>
> Helder

Hi Mats,

Check this out for the Checksum (SHA1):

<https://groups.google.com/d/msg/comp.lang.idl-pvwave/J6rE26rO8ko/PXRr8RJAHSYJ>

Cheers,

Helder

Subject: Re: Download files from the web

Posted by on Tue, 17 Dec 2013 09:59:06 GMT

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

Den tisdagen den 17:e december 2013 kl. 10:51:07 UTC+1 skrev Helder:

> On Tuesday, December 17, 2013 10:39:42 AM UTC+1, Mats Löfdahl wrote:

>> 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.

>

> Hi Mats,

>

> the files size is given in my example. Look at the output that I print out. For example:

>

> Check for update: Http: get: received (9594572), total expected (15501460)

OK, so that is echoed during the transfer? Can you access the size information through some keyword?

Subject: Re: Download files from the web

Posted by [Helder Marchetto](#) on Tue, 17 Dec 2013 10:24:58 GMT

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

On Tuesday, December 17, 2013 10:59:06 AM UTC+1, Mats Löfdahl wrote:

> Den tisdagen den 17:e december 2013 kl. 10:51:07 UTC+1 skrev Helder:

>

>> On Tuesday, December 17, 2013 10:39:42 AM UTC+1, Mats Löfdahl wrote:

>

>>> 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.

>

>>

>

>> Hi Mats,

>

>>

>

>> the files size is given in my example. Look at the output that I print out. For example:

>

>>

>

>> Check for update: Http: get: received (9594572), total expected (15501460)

>

>

>
> OK, so that is echoed during the transfer? Can you access the size information through some keyword?

Not that I know of (and trust me, I don't know much). However, it is included in the callback function. In my example from before, I used this info to make the progress bar. The size of the file is in progress[1] and the downloaded size is in progress[2].

If you only want to get the size, you could exit the callback function by returning a zero. However, this method far from being elegant. Therefore experience teaches: there must be a better way.

What you want to do is to check the directory content (size). I **think** that using some "url-get" method you will not get anywhere, because what you get is the html page.

Another idea would be to parse the html for "href" content. To do this, the web page must contain a listing of the available files.

Let me know if you come up with better ideas.

cheers,
h

Subject: Re: Download files from the web
Posted by on Fri, 10 Jan 2014 10:22:16 GMT
[View Forum Message](#) <> [Reply to Message](#)

Den tisdagen den 17:e december 2013 kl. 10:39:42 UTC+1 skrev Mats Löfdahl:
> Den måndagen den 16:e december 2013 kl. 22:27:02 UTC+1 skrev Mike Galloy:
>
>> On 12/16/13, 7:46 AM, Mats Lï½fdahl wrote:
>
>>> Den mï½ndagen den 16:e december 2013 kl. 15:14:10 UTC+1 skrev Mats Lï½fdahl:
>
>>>>
>
>>>> 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
>
>>
>
>>> 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.

```

I thought I had this working but now I suddenly get into another kind of problem when the file I'm trying to download does not exist on the server. The get method of IDLnetUrl seems to crash rather than just returning a status that is NE 200.

What I do now is:

```

urlComponents = parse_url(url)
oUrl = OBJ_NEW('IDLnetUrl' $
    , URL_SCHEME = urlComponents.scheme $
    , URL_HOSTNAME = urlComponents.host $
    , URL_PATH = urlComponents.path $
    , URL_PORT = urlComponents.port $
)
tmpfile = String('tmp_', Bin_Date(SysTime()), format='(A, I4, 5I2.2)')
retrievedFilePath = oUrl -> Get(FILENAME=tmpfile)
oUrl -> GetProperty, RESPONSE_CODE=RespCode

```

The urlComponents look fine, tmpfile is a string as it should (e.g., 'tmp_2014011011025') but I get the following error message:

```
% IDLNETURL::GET: CCurlException: Error: Http Get Request Failed. Error =  
    http: Client Error. Remote Host(www.royac.iac.es), Http  
    ErrCode(404), Http Err(Not Found) Http ErrMsg(No HTML found).  
% Execution halted at: RED_GETURL      159
```

where red_geturl is the function where I put the code above (with line 159 being the one where the get method is called). retrievedFilePath is undefined after this of course.

I can execute the last line above after this and get a response code of 404 but at that point the program has stopped. I suppose I could add some error handling code but isn't the IDLnetUrl object supposed to take care of this?

Subject: Re: Download files from the web
Posted by [Helder Marchetto](#) on Fri, 10 Jan 2014 10:41:20 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Friday, January 10, 2014 11:22:16 AM UTC+1, Mats Löfdahl wrote:
> Den tisdagen den 17:e december 2013 kl. 10:39:42 UTC+1 skrev Mats Löfdahl:
>
>> Den måndagen den 16:e december 2013 kl. 22:27:02 UTC+1 skrev Mike Galloy:
>
>>
>
>>> On 12/16/13, 7:46 AM, Mats Lï½fdahl wrote:
>
>>
>
>>>> Den mï½ndagen den 16:e december 2013 kl. 15:14:10 UTC+1 skrev Mats
Lï½fdahl:
>
>>
>
>>>> >
>
>>
>
>>>> > Thanks. But it seems it has the same problem as the webget
>
>>
>
>>>> > 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
>
>>
>
>>>>
>
>>
>
>>>> 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.
>
>
>
> I thought I had this working but now I suddenly get into another kind of problem when the file
I'm trying to download does not exist on the server. The get method of IDLnetUrl seems to crash
rather than just returning a status that is NE 200.
>
>
>
> What I do now is:
>
>
>
>
>
> urlComponents = parse_url(url)
>
> oUrl = OBJ_NEW('IDLnetUrl' $
>
>         , URL_SCHEME = urlComponents.scheme $
>
>         , URL_HOSTNAME = urlComponents.host $
>
>         , URL_PATH = urlComponents.path $
>
>         , URL_PORT = urlComponents.port $
>
>         )
>
> tmpfile = String('tmp_', Bin_Date(SysTime()), format='(A, I4, 5I2.2)')
>
> retrievedFilePath = oUrl -> Get(FILENAME=tmpfile)
>
> oUrl -> GetProperty, RESPONSE_CODE=RespCode
>
>
>

```

> The urlComponents look fine, tmpfile is a string as it should (e.g., 'tmp_2014011011025') but I get the following error message:

>

>

>

> % IDLNETURL::GET: CCurlException: Error: Http Get Request Failed. Error =

>

> http: Client Error. Remote Host(www.royac.iac.es), Http

>

> ErrCode(404), Http Err(Not Found) Http ErrMsg(No HTML found).

>

> % Execution halted at: RED_GETURL 159

>

>

>

> where red_geturl is the function where I put the code above (with line 159 being the one where the get method is called). retrievedFilePath is undefined after this of course.

>

>

>

> I can execute the last line above after this and get a response code of 404 but at that point the program has stopped. I suppose I could add some error handling code but isn't the IDLnetUrl object supposed to take care of this?

Hi Mats,

this will not really help, but at least you feel my pain :-)

I'm working around this problem with a call to catch and then check the error code (404). You can then either report that the file is missing if the error is 404, otherwise simply give out the error code.

I found that this is the only way if you're using http. With ftp you can of course check the list of available files with the method GetFtpDirList.

But I guess you already know all of this.

If there's a better way... happy to hear about it.

Regards,
Helder

Subject: Re: Download files from the web
Posted by [Helder Marchetto](#) on Fri, 10 Jan 2014 11:08:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Friday, January 10, 2014 11:41:20 AM UTC+1, Helder wrote:

> On Friday, January 10, 2014 11:22:16 AM UTC+1, Mats Löfdahl wrote:
>

```

>> Den tisdagen den 17:e december 2013 kl. 10:39:42 UTC+1 skrev Mats Löfdahl:
>
>>
>
>>> Den måndagen den 16:e december 2013 kl. 22:27:02 UTC+1 skrev Mike Galloy:
>
>>
>
>>>
>
>>
>
>>>> On 12/16/13, 7:46 AM, Mats Löfdahl wrote:
>
>>
>
>>>
>
>>
>
>>>> > Den måndagen den 16:e december 2013 kl. 15:14:10 UTC+1 skrev Mats
Löfdahl:
>
>>
>
>>>
>
>>
>
>>>> >>
>
>>
>
>>>
>
>>
>
>>>> >> Thanks. But it seems it has the same problem as the webget
>
>>
>
>>>
>
>>
>
>>>> >> 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
>
>>
>
>>>
>
>>
>
>>>>
>
>>
>
>>>
>
>>
>
>>>> > 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: CUrlException: 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.
>
>>
>
>>
>
>>
>
>> I thought I had this working but now I suddenly get into another kind of problem when the file
I'm trying to download does not exist on the server. The get method of IDLnetUrl seems to crash
rather than just returning a status that is NE 200.
>
>>
>
>>
>
>>
>
>> What I do now is:
>
>>
>
>>
>
>>
>
>>
```

```

>
>>
>
>> urlComponents = parse_url(url)
>
>>
>
>> oUrl = OBJ_NEW('IDLnetUrl' $
>
>>
>
>>         , URL_SCHEME = urlComponents.scheme $
>
>>
>
>>         , URL_HOSTNAME = urlComponents.host $
>
>>
>
>>         , URL_PATH = urlComponents.path $
>
>>
>
>>         , URL_PORT = urlComponents.port $
>
>>
>
>>         )
>
>>
>
>> tmpfile = String('tmp_', Bin_Date(SysTime()), format='(A, I4, 5I2.2)')
>
>>
>
>> retrievedFilePath = oUrl -> Get(FILENAME=tmpfile)
>
>>
>
>> oUrl -> GetProperty, RESPONSE_CODE=RespCode
>
>>
>
>>
>
>>
>
>> The urlComponents look fine, tmpfile is a string as it should (e.g., 'tmp_2014011011025') but I

```

get the following error message:

```
>
>>
>
>>
>
>>
>
>> % IDLNETURL::GET: CCurlException: Error: Http Get Request Failed. Error =
>
>>
>
>> http: Client Error. Remote Host(www.royac.iac.es), Http
>
>>
>
>> ErrCode(404), Http Err(Not Found) Http ErrMsg(No HTML found).
>
>>
>
>> % Execution halted at: RED_GETURL      159
>
>>
>
>>
>
>>
>
>> where red_geturl is the function where I put the code above (with line 159 being the one
where the get method is called). retrievedFilePath is undefined after this of course.
>
>>
>
>>
>
>>
>
>> I can execute the last line above after this and get a response code of 404 but at that point the
program has stopped. I suppose I could add some error handling code but isn't the IDLnetUrl
object supposed to take care of this?
>
>
>
> Hi Mats,
>
> this will not really help, but at least you feel my pain :-)
```

>

> I'm working around this problem with a call to catch and then check the error code (404). You

can then either report that the file is missing if the error is 404, otherwise simply give out the error code.

>
>
>
> I found that this is the only way if you're using http. With ftp you can of course check the list of available files with the method GetFtpDirList.
>
>
>
> But I guess you already know all of this.
>
>
>
> If there's a better way... happy to hear about it.
>
>
>
> Regards,
>
> Helder

Hi Mats,
one more dirty trick. It is a one liner that uses windows powershell. Dunno about Linux, but you also have on linux the wGet command and it looks like it uses the same syntax. Otherwise curl should also work (--head option).

SPAWN, 'powershell -WindowStyle Hidden "wget --server-response --spider -o OutputFile.txt http://yourserver.com/MissingFile.txt"', wgetResult, /NOSHELL

wgetResult will unfortunately be empty.
The resulting file (OutputFile.txt) will contain a lot of stuff, among which lots of 404 mentions if the file does not exist. You can then search for that or check if the last line of the file contains "Remote file exists."

Does this help?

I'm always on the look for better/cleaner solutions.

Regards,
Helder

Subject: Re: Download files from the web
Posted by on Fri, 10 Jan 2014 14:48:34 GMT
[View Forum Message](#) <> [Reply to Message](#)

Den fredagen den 10:e januari 2014 kl. 11:41:20 UTC+1 skrev Helder:

> On Friday, January 10, 2014 11:22:16 AM UTC+1, Mats Löfdahl wrote:
>
>> [...] I can execute the last line above after this and get a response code of 404 but at that point the program has stopped. I suppose I could add some error handling code but isn't the IDLnetUrl object supposed to take care of this?
>
> Hi Mats,
>
> this will not really help, but at least you feel my pain :-)
>
> I'm working around this problem with a call to catch and then check the error code (404). You can then either report that the file is missing if the error is 404, otherwise simply give out the error code.

I did this now, as well. Seems to work. Thanks.

> I found that this is the only way if you're using http. With ftp you can of course check the list of available files with the method GetFtpDirList.
>
> But I guess you already know all of this.

No, I'm new to downloads from within IDL. I haven't tried ftp yet.

> If there's a better way... happy to hear about it.

Me too. :o)
