# 1. Link Extraction using Requests module

August 11, 2018

Saumay Agrawal
16BCE1151

In [1]:
```
# Importing the requests module and re module for regular expressions
import requests
import re
```

In [2]:
```
# This method find the links in an HTML page using regular expressions and returns the
def getlink(code):
    links = re.findall('"((http|ftp)s?://.*?)"',  code)
    links = [i[0] for i in links]
    return links
```

In [3]:
```
# Separate out the 'img' tag from HTML code and then use getlink() to extract image li
def findimg(code):
    links = []
    start = 0
    while(start != -1):
        start = code.find('<img')
        end = code.find('>', start)
        tag  = code[start:end+1]
        link = getlink(tag)
        links.append(link)
        code = code[end+1:]
    return links
```

In [4]:
```
# Fetching a web page and storing its metadata
page = requests.get('https://towardsdatascience.com/getting-started-with-graph-analysis
print(page)
```

```
<Response [200]>
```

In [5]:
```
# Get all the links in webpage using getlink() and print them.
alllinks = getlink(page.text)
print('All the links are ' + str(len(alllinks)) + ' :')
for link in alllinks:
    print(link)
```

```
All the links are 148 :
http://creativecommons.org/ns#
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://cdn-images-1.medium.com/max/1200/1*iNZ76lGNlec2DuYOjrbjOw.png
https://cdn-images-1.medium.com/max/1200/1*iNZ76lGNlec2DuYOjrbjOw.png
https://plus.google.com/103654360130207659246
https://towardsdatascience.com/@FelixRvrt
https://www.facebook.com/towardsdatascience
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
http://schema.org
https://cdn-images-1.medium.com/max/2000/1*iNZ76lGNlec2DuYOjrbjOw.png
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://cdn-images-1.medium.com/max/2000/1*iNZ76lGNlec2DuYOjrbjOw.png
https://towardsdatascience.com/@FelixRvrt
https://towardsdatascience.com
https://cdn-images-1.medium.com/max/616/1*OMF3fSqH8t4xBJ9-6oZDZw.png
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://cdn-static-1.medium.com/_/fp/css/main-branding-base.Fc55unvcP5htkyH_Q-aGIA.css
https://www.google-analytics.com/analytics.js
https://cdn-static-1.medium.com/_/fp/js/shiv.RI2ePTZ5gFmMgLzG5bEVAA.js
https://cdn-static-1.medium.com/_/fp/icons/favicon-rebrand-medium.3Y6xpZ-0FSdWDnPM3hSBIA.ico
https://cdn-images-1.medium.com/fit/c/304/304/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
https://cdn-images-1.medium.com/fit/c/240/240/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
https://cdn-images-1.medium.com/fit/c/152/152/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
https://cdn-images-1.medium.com/fit/c/120/120/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
https://cdn-static-1.medium.com/_/fp/icons/monogram-mask.KPLCSFEZviQN0jQ7veN2RQ.svg
https://medium.com/
https://towardsdatascience.com?source=logo-lo_24wL5L7CKshG---7f60cf5620c9
https://medium.com/_/subscribe/collection/towards-data-science
https://twitter.com/TDataScience
https://medium.com/m/signin?redirect=https%3A%2F%2Ftowardsdatascience.com%2Fgetting-started-wi
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://medium.com/m/signin?redirect=https%3A%2F%2Ftowardsdatascience.com%2Fgetting-started-wi
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://towardsdatascience.com
https://towardsdatascience.com/data-science/home
https://towardsdatascience.com/machine-learning/home
https://towardsdatascience.com/programming/home
https://towardsdatascience.com/data-visualization/home
https://towardsdatascience.com/editors-picks/home
https://towardsdatascience.com/contribute/home
https://towardsdatascience.com/search
https://towardsdatascience.com/@FelixRvrt?source=post_header_lockup
https://cdn-images-1.medium.com/fit/c/120/120/1*y_mPGPA3yxnl6725B6HK8w.jpeg
https://towardsdatascience.com/@FelixRvrt?source=post_header_lockup
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://medium.com/_/subscribe/user/414435874b4b
```

https://safetyapp.shinyapps.io/GoWvis/
https://safetyapp.shinyapps.io/GoWvis/
https://cdn-images-1.medium.com/max/1600/1*iNZ76lGNlec2DuYOjrbjOw.png
https://cdn-images-1.medium.com/max/1600/1*CO3jSwe_xYKW-tucvMVxjw.png
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://cdn-images-1.medium.com/max/1600/1*VldI27-_FVle9aX5bN9Ufw.png
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://cdn-images-1.medium.com/max/1600/1*dcrYjJYK-xAxJ3GGVvZjRg.png
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://cdn-images-1.medium.com/max/1600/1*MeRNyRVeuy-iRiRen2JPCg.png
https://cdn-images-1.medium.com/max/1600/1*FkkaUv23_UGGA1jiV8YZOA.png
https://cdn-images-1.medium.com/max/1600/1*xjFo0qx927c9DYT-LdJM8g.png
https://github.com/FelixChop/MediumArticles/blob/master/Graph_analysis_Python.ipynb
https://github.com/FelixChop/MediumArticles/blob/master/Graph_analysis_Python.ipynb
https://towardsdatascience.com/tagged/data-science?source=post
https://towardsdatascience.com/tagged/python?source=post
https://towardsdatascience.com/tagged/graph-analysis?source=post
https://towardsdatascience.com/tagged/machine-learning?source=post
https://towardsdatascience.com/tagged/artificial-intelligence?source=post
https://medium.com/_/vote/p/5e2d2f82f18e
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://medium.com/_/subscribe/user/414435874b4b
https://towardsdatascience.com/@FelixRvrt?source=footer_card
https://cdn-images-1.medium.com/fit/c/120/120/1*y_mPGPA3yxnl6725B6HK8w.jpeg
https://towardsdatascience.com/@FelixRvrt
https://medium.com/_/subscribe/collection/towards-data-science
https://towardsdatascience.com?source=footer_card
https://cdn-images-1.medium.com/fit/c/120/120/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
https://towardsdatascience.com?source=footer_card
https://medium.com/_/vote/p/5e2d2f82f18e
https://medium.com/_/bookmark/p/5e2d2f82f18e
https://towardsdatascience.com
https://cdn-images-1.medium.com/fit/c/80/80/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
https://medium.com/@Medium/personalize-your-medium-experience-with-users-publications-tags-26a4
https://medium.com/_/subscribe/collection/towards-data-science
https://d1fcbxp97j4nb2.cloudfront.net
https://towardsdatascience.com
https://cdn-images-1.medium.com
https://cdn-static-1.medium.com/_/fp/gen-js/main-base.bundle.h0kHe3Kj69pB1_RTXtbVgw.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-common-async.bundle.-WAlEkrMFYcECllbJVLRDg.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-hightower.bundle.Km43xbQv03cHLu9Q0CrsYQ.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-home-screens.bundle.Tr7kDcKl8ZN53Cepp1t_IA.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-misc-screens.bundle.bXsSI0iNhNahzoFQs8UPJQ.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-notes.bundle.uxOEHNCNOfQw4AKlAJSayQ.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-payments.bundle.GC_d9-vvOXAz_Ah6q9Gsyw.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-posters.bundle.4BbOmQwbLaG1YSyuHlUJbg.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-power-readers.bundle.Zazez5NZXWT5SKYy_rlLgg.js

https://cdn-static-1.medium.com/_/fp/gen-js/main-pubs.bundle.PmzWPl3pBTh-tcm-rWQmRA.js
https://cdn-static-1.medium.com/_/fp/gen-js/main-stats.bundle.z9afrKKJ6Bh8ThgtXRjYqw.js
https://cdn-static-1.medium.com/_/fp/icons/monogram-mask.KPLCSFEZviQN0jQ7veN2RQ.svg
https://cdn-static-1.medium.com/_/fp/icons/favicon-dev-editor.YKKRxBO8EMvIqhyCwIiJeQ.ico
https://cdn-static-1.medium.com/_/fp/icons/favicon-hatch-editor.BuEyHIqlyh2s_XEk4Rl32Q.ico
https://cdn-static-1.medium.com/_/fp/icons/favicon-medium-editor.PiakrZWB7Yb80quUVQWM6g.ico
https://medium.com
https://MQ57UUUQZ2-dsn.algolia.net
https://itunes.apple.com/app/medium-everyones-stories/id828256236?pt=698524&mt=8
https://play.google.com/store/apps/details?id=com.medium.reader
https://accounts.google.com/o/oauth2/auth
https://cdn-static-1.medium.com/_/fp/js/shiv.RI2ePTZ5gFmMgLzG5bEVAA.js
https://cdn-static-1.medium.com/_/fp/icons/favicon-rebrand-medium.3Y6xpZ-0FSdWDnPM3hSBIA.ico
https://glyph.medium.com/css/e/sr/latin/e/ssr/latin/e/ssb/latin/m2.css
https://glyph.medium.com/css/m2.css
https://glyph.medium.com/css/mkt.css
https://glyph.medium.com/css/elv8.css
https://cdn-static-1.medium.com/_/fp/icons/favicon-rebrand-medium-editor.3Y6xpZ-0FSdWDnPM3hSBI
https://glyph.medium.com
https://goldfinch.medium.com
https://buggle.medium.com
https://cdn-videos-1.medium.com
https://cdn-audio-1.medium.com
https://cdn-static-1.medium.com/_/fp/gen-js/main-base.bundle.h0kHe3Kj69pB1_RTXtbVgw.js
https://towardsdatascience.com/data-science/home
https://towardsdatascience.com/machine-learning/home
https://towardsdatascience.com/programming/home
https://towardsdatascience.com/data-visualization/home
https://towardsdatascience.com/editors-picks/home
https://towardsdatascience.com/contribute/home
https://safetyapp.shinyapps.io/GoWvis/
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://i.embed.ly/1/image?url=https%3A%2F%2Favatars3.githubusercontent.com%2Fu%2F8631089%3Fs%3
https://github.com/FelixChop/MediumArticles/blob/master/Graph_analysis_Python.ipynb
https://github.com/FelixChop/MediumArticles/blob/master/Graph_analysis_Python.ipynb
https://safetyapp.shinyapps.io/GoWvis/
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://towardsdatascience.com/data-science/home
https://towardsdatascience.com/machine-learning/home
https://towardsdatascience.com/programming/home
https://towardsdatascience.com/data-visualization/home
https://towardsdatascience.com/editors-picks/home
https://towardsdatascience.com/contribute/home
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://towardsdatascience.com/getting-started-with-graph-analysis-in-python-with-pandas-and-ne
https://towardsdatascience.com/data-science/home

```
https://towardsdatascience.com/machine-learning/home
https://towardsdatascience.com/programming/home
https://towardsdatascience.com/data-visualization/home
https://towardsdatascience.com/editors-picks/home
https://towardsdatascience.com/contribute/home
```

```
In [6]: # Get all the links in webpage using findimg() and print them.
        imglinks = findimg(page.text)
        print('All the image links are ' + str(len(imglinks)) + ' :')
        for link in imglinks:
            if len(link):
                print(link[0])
```

```
All the image links are 12 :
https://cdn-images-1.medium.com/fit/c/120/120/1*y_mPGPA3yxnl6725B6HK8w.jpeg
https://cdn-images-1.medium.com/max/1600/1*iNZ76lGNlec2DuYOjrbjOw.png
https://cdn-images-1.medium.com/max/1600/1*CO3jSwe_xYKW-tucvMVxjw.png
https://cdn-images-1.medium.com/max/1600/1*VldI27-_FVle9aX5bN9Ufw.png
https://cdn-images-1.medium.com/max/1600/1*dcrYjJYK-xAxJ3GGVvZjRg.png
https://cdn-images-1.medium.com/max/1600/1*MeRNyRVeuy-iRiRen2JPCg.png
https://cdn-images-1.medium.com/max/1600/1*FkkaUv23_UGGA1jiV8YZOA.png
https://cdn-images-1.medium.com/max/1600/1*xjFo0qx927c9DYT-LdJM8g.png
https://cdn-images-1.medium.com/fit/c/120/120/1*y_mPGPA3yxnl6725B6HK8w.jpeg
https://cdn-images-1.medium.com/fit/c/120/120/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
https://cdn-images-1.medium.com/fit/c/80/80/1*F0LADxTtsKOgmPa-_7iUEQ.jpeg
```