

```

'''
Introduction: A PYTHON Program to implement two level crawler for all >>
http or https << links within a particular URL specified by the user upto
2 levels.
This crawler can be extended as according to requirements. This is made
for crawling all unique http or https links because they can be
referred/re-opened for future/multilevel crawling. This crawler will also
keep on posting links to you in case it remains unable to open any link.
'''

#Steps to run the program on any linux based system (like Ubuntu, RedHat
etc.):
'''
1) Open 'Terminal'.
2) Write the following command at terminal window:
    python crawl.py
3) Specify/Give URL you want to crawl to the application program. And
wait a few seconds for Output.
'''

#A single level http links crawler implemented live at
http://www.startlingapps.appspot.com/
#One more crawler which crawls all unique links for its user's queries at
www.bing.com is live at http://www.qbingapp.appspot.com/
'''Starting Code for a two level http or https crawler.'''
#importing modules for URL related operations and parsing the
corresponding html page
import urllib2
from BeautifulSoup import BeautifulSoup
#Variables declaration section
unique=[]
third=[]
#defining a function for crawling a given URL
def crawler(url):
    try:
        var=urllib2.urlopen(url)           #opening URL
        state=var.getcode()                 #getting URL response code
        if state==200:
            html=var.read()                 #getting html
            soup=BeautifulSoup(html)        #considering html page for
parsing
            j=soup.findAll('a',href=True)   #parsing for <a> tag
in the html page
            first=[]                        #declaring a new list
            for i in j:
                i=i['href']                 #changing to 'href' for the page
                ht=i.split(':')             #splitting for checking
whether there is 'http' like link or not
                if ht[0]=="http" or ht[0]=="https":
                    first.append(i)         #appending all http links
to first list
                for i in first:             #checking for do we have
duplicate links in first list
                    if i not in unique:
                        unique.append(i)
    except:

```

```

        print "\033[1;31m Error: Url '"+ url + "' cannot be
opened.\033[0m"

def main():
    #defining main()
    print
    """
    *****
    """
    url=raw_input("Specify URL (For Example:
http://www.startlingapps.appspot.com/):\n")    #getting user input
    crawler(url)                                #crawling the url for the first time
    length=int(len(unique))                      #finding the length of unique
list
    third=unique
    print
    """
    *****
    """
    print ">>>>>>Varoius http or https Links at level-1 only:"
    print
    """
    *****
    """
    for i in third:
        print i
        #print length
    print
    """
    *****
    """
    print ">>>>>>Varoius http or https Links at level-1 and level-2:"
    print
    """
    *****
    """
    while(length>0):
        #for crawling at level 2
        length=length-1
        '''print length
        print "\033[1;32m" +third[length]
        print "\033[0m'''
        crawler(third[length])
    for i in unique:
        print i
        print "\nTotal unique links found: ",len(unique)
if __name__ == '__main__':
    #calling main()
    main()

```