

## Practical 5 WEB Crawler

```
type(webpage.content)
```

bytes

```
soup = BeautifulSoup(webpage.content, "html.parser")
```

soup

```
<!DOCTYPE html>
<html class="a-no-js" data-19ax5a9jf="dingo" lang="en-in"><!-- sp:feature:head-start -->
<head><script>var aPageStart = (new Date()).getTime();</script><meta charset="utf-8"/>
<!-- sp:end-feature:head-start -->
<!-- sp:feature:csm:head-open-part1 -->
<script type="text/javascript">var ue_t0=ue_t0||+new Date();</script>
<!-- sp:end-feature:csm:head-open-part1 -->
<!-- sp:feature:cs-optimization -->
<meta content="on" http-equiv="x-dns-prefetch-control"/>
<link href="https://images-eu.ssl-images-amazon.com" rel="dns-prefetch"/>
<link href="https://m.media-amazon.com" rel="dns-prefetch"/>
<link href="https://completion.amazon.com" rel="dns-prefetch"/>
<!-- sp:end-feature:cs-optimization -->
<!-- sp:feature:csm:head-open-part2 -->
<script type="text/javascript">
window.ue_ihb = (window.ue_ihb || window.ueinit || 0) + 1;
if (window.ue_ihb === 1) {

var ue_csm = window,
    ue_hob = +new Date();
(function(d){var e=d.ue||{};f=Date.now||function(){return+new Date();}e.d=function(b){return f()-(b?
0:d.ue_t0)};e.stub=function(b,a){if(!b[a]) {var c=[];b[a]=function()
{c.push([c.slice.call(arguments),e.d(),d.ue_id]);}b[a].replay=function(b){for(var
a;a=c.shift();)b(a[0],a[1],a[2]);}b[a].isStub=1}};e.exec=function(b,a){return function(){try{return
b.apply(this,arguments)}catch(c){ueLogError(c,{attribution:a||"undefined",logLevel:"WARN"})}}}})(ue_csm);

    var ue_err_chan = 'jserr';
(function(d,e){function h(f,b){if(!(a.ec>a.mxe)&&f){a.ter.push(f);b=b||{};var
c=f.logLevel||b.logLevel;c&&c!==k&&c!==m&&c!==n&&c!==p||a.ec++;c&&c!=k||a.ecf++;b.pageURL=""+(e.location?
e.location.href:"");b.logLevel=c;b.attribution=f.attribution||b.attribution;a.erl.push({ex:f,info:b})}}function l(a,b,c,e,g)
{d.ueLogError({m:a,f:b,l:c,"+":e,err:g,fromOnErrorHandler:1,args:arguments},g?{attribution:g.attribution,logLevel:g.logLevel}:void
0);return!1}var k="FATAL",n="ERROR",p="WARN",p="DOWNGRADED",a={ec:0,ecf:0,
pec:0,ts:0,erl:[],ter:[],buffer:[],mxe:50,startTimer:function(){a.ts++;setInterval(function()
{d.ue&&a.pec<a.ec&&d.uex("at");a.pec=a.ec},1E4)}},l.skipTrace=1;h.skipTrace=1;h.isStub=1;d.ueLogError=h;d.ue_err=a;e.onerror=
(ue_csm>window);

var ue_id = 'FZ8QGGQYVV2QDY9PPRDM',
    ue_url = '/rd/uedata',
    ue_navtiming = 1,
    ue_mid = 'A21TJRUUN4KGV',
    ue_sid = '520-1665529-0104821',
    ue_sn = 'www.amazon.in',
    ue_furl = 'fls-eu.amazon.in',
    ue_surl = 'https://unagi-na.amazon.com/1/events/com.amazon.csm.nexusclient.prod',
    ue_int = 0,
    ue_fcsm = 1,
    ue_urtt = 3,
    ue_rpl_ns = 'cel-rpl',
    ue_ddq = 1,
    ue_fpf = '//fls-eu.amazon.in/1/batch/1/OP/A21TJRUUN4KGV:520-1665529-0104821:FZ8QGGQYVV2QDY9PPRDM$uedata=s:',
    ue_sbuimp = 1,
    ue_ibft = 0,
    ue_sswmts = 0,
    ue_jsmtf = 0,
    ue_fnt = 0,
```

```
links = soup.find_all("a", attrs={'class': 'a-link-normal s-line-clamp-2 s-line-clamp-3-for-col-12 s-link-style a-text-normal'})
```

## links

[<a class="a-link-normal s-line-clamp-2 s-line-clamp-3-for-col-12 s-link-style a-text-normal" href="/HP-Pavilion-i5-1335U-Laptop-Touchscreen/dp/B0CJ2KWNQ/ref=sr\_1\_1?dib=eyJ2IjoiMSJ9.Idvdwt\_UXS2\_-E6IS9-yikQp2MGPp5Zy\_jVH9i3ztLfPoqmYNjW6qzc70A3dgNRB\_v0Ly\_WMPLGFOIoRg725b5IkMxoj1P\_yB59IpK-CnpsQMY5UhPQhwSg99vh4bV60J7ROSUJnLe0qS81aaXgu2EVDngcZi-MmxeA1ymqjqET4md8V25usp1TphTFZcKjq-Te612d8\_G2f7DksT9JXsL\_B71xIm62j8h6b1XUyDA.NFFF9wjpHVybsCyUSj7hd\_S2K4TBULiqUVi\_NAUhdIk&tag=se&keywords=laptop&1" target="\_blank"><h2 aria-label='HP Pavilion x360, 13th Gen Intel Core i5-1335U (16GB DDR4, 512GB SSD) Touchscreen, 14"/35.6cm, FHD, Win 11, Office 21, Silver, 1.51kg, ek1074tu/1148tu, Iris Xe, FPR, 5MP Camera, Backlit KB Laptop' class="a-size-medium a-spacing-none a-color-base a-text-normal"><span>HP Pavilion x360, 13th Gen Intel Core i5-1335U (16GB DDR4, 512GB SSD) Touchscreen, 14"/35.6cm, FHD, Win 11, Office 21, Silver, 1.51kg, ek1074tu/1148tu, Iris Xe, FPR, 5MP Camera, Backlit KB Laptop</span></h2></a>, <a class="a-link-normal s-line-clamp-2 s-line-clamp-3-for-col-12 s-link-style a-text-normal" href="/ASUS-5050-8GB-Windows-Eclipse-G614PH-RV033WS/dp/B0FB9G9XVG/ref=sr\_1\_2?dib=eyJ2IjoiMSJ9.Idvdwt\_UXS2\_-E6IS9-yikQp2MGPp5Zy\_jVH9i3ztLfPoqmYNjW6qzc70A3dgNRB\_v0Ly\_WMPLGFOIoRg725b5IkMxoj1P\_yB59IpK-CnpsQMY5UhPQhwSg99vh4bV60J7ROSUJnLe0qS81aaXgu2EVDngcZi-MmxeA1ymqjqET4md8V25usp1TphTFZcKjq-Te612d8\_G2f7DksT9JXsL\_B71xIm62j8h6b1XUyDA.NFFF9wjpHVybsCyUSj7hd\_S2K4TBULiqUVi\_NAUhdIk&tag=se&keywords=laptop&2" target="\_blank"><h2 aria-label='ASUS ROG Strix G16, AMD Ryzen 9 8940HX, Gaming Laptop(RTX 5050-8GB/115W TGP/16GB RAM/1TB SSD/FHD+/16"/165Hz/90Whr/Windows 11/M365 Basic (1Year)\*/Office Home 2024/Eclipse Gray/2.5 Kg) G614PH-RV033WS' class="a-size-medium a-spacing-none a-color-base a-text-normal"><span>ASUS ROG Strix G16, AMD Ryzen 9 8940HX, Gaming Laptop(RTX 5050-8GB/115W TGP/16GB RAM/1TB SSD/FHD+/16"/165Hz/90Whr/Windows 11/M365 Basic (1Year)\*/Office Home 2024/Eclipse Gray/2.5 Kg) G614PH-RV033WS</span></h2></a>, <a class="a-link-normal s-line-clamp-2 s-line-clamp-3-for-col-12 s-link-style a-text-normal" href="/ASUS-15-6-inch-GeForce- Graphite-FA506NCR-HN054W/dp/B0D5DFR78J/ref=sr\_1\_3?dib=eyJ2IjoiMSJ9.Idvdwt\_UXS2\_-E6IS9-yikQp2MGPp5Zy\_jVH9i3ztLfPoqmYNjW6qzc70A3dgNRB\_v0Ly\_WMPLGFOIoRg725b5IkMxoj1P\_yB59IpK-CnpsQMY5UhPQhwSg99vh4bV60J7ROSUJnLe0qS81aaXgu2EVDngcZi-MmxeA1ymqjqET4md8V25usp1TphTFZcKjq-Te612d8\_G2f7DksT9JXsL\_B71xIm62j8h6b1XUyDA.NFFF9wjpHVybsCyUSj7hd\_S2K4TBULiqUVi\_NAUhdIk&tag=se&keywords=laptop&3" target="\_blank"><h2 aria-label='ASUS TUF Gaming A15, AMD Ryzen 7 7435HS Gaming Laptop(NVIDIA RTX 3050-4GB/60W TGP/16GB RAM/512GB SSD/FHD/15.6"/144Hz/RGB KB/48Whr/Windows 11//Graphite Black/2.30 Kg) FA506NCR-HN054W</span></h2></a>, <a class="a-link-normal s-line-clamp-2 s-line-clamp-3-for-col-12 s-link-style a-text-normal" href="/Acer-Celeron-Processor\_LPDDR4X-A325-45/dp/B0D7PYTTGH/ref=sr\_1\_4?dib=eyJ2IjoiMSJ9.Idvdwt\_UXS2\_-E6IS9-yikQp2MGPp5Zy\_jVH9i3ztLfPoqmYNjW6qzc70A3dgNRB\_v0Ly\_WMPLGFOIoRg725b5IkMxoj1P\_yB59IpK-CnpsQMY5UhPQhwSg99vh4bV60J7ROSUJnLe0qS81aaXgu2EVDngcZi-MmxeA1ymqjqET4md8V25usp1TphTFZcKjq-Te612d8\_G2f7DksT9JXsL\_B71xIm62j8h6b1XUyDA.NFFF9wjpHVybsCyUSj7hd\_S2K4TBULiqUVi\_NAUhdIk&tag=se&keywords=laptop&4" target="\_blank"><h2 aria-label='Acer [SmartChoice Aspire 3 Laptop Intel Core Celeron N4500 Processor Laptop (8 GB LPDDR4X SDRAM/512 GB SSD/Win11 Home/38 WHR/HD Webcam) A325-45 with 39.63 cm (15.6") HD Display, Pure Silver, 1.5 KG' class="a-size-medium a-spacing-none a-color-base a-text-normal"><span>Acer [SmartChoice Aspire 3 Laptop Intel Core Celeron N4500 Processor Laptop (8 GB LPDDR4X SDRAM/512 GB SSD/Win11 Home/38 WHR/HD Webcam) A325-45 with 39.63 cm (15.6") HD Display, Pure Silver, 1.5 KG</span></h2></a>, <a class="a-link-normal s-line-clamp-2 s-line-clamp-3-for-col-12 s-link-style a-text-normal" href="/Acer-ALG-Intel-Core- 13th/dp/B0DXVGVNKN/ref=sr\_1\_5?dib=eyJ2IjoiMSJ9.Idvdwt\_UXS2\_-E6IS9-yikQp2MGPp5Zy\_jVH9i3ztLfPoqmYNjW6qzc70A3dgNRB\_v0Ly\_WMPLGFOIoRg725b5IkMxoj1P\_yB59IpK-CnpsQMY5UhPQhwSg99vh4bV60J7ROSUJnLe0qS81aaXgu2EVDngcZi-MmxeA1ymqjqET4md8V25usp1TphTFZcKjq-Te612d8\_G2f7DksT9JXsL\_B71xIm62j8h6b1XUyDA.NFFF9wjpHVybsCyUSj7hd\_S2K4TBULiqUVi\_NAUhdIk&tag=se&keywords=laptop&5" target="\_blank"><h2 aria-label='Acer ALG, Intel Core i5 13th Gen - 13420H, NVIDIA GeForce RTX 3050-6GB, 16GB RAM, 512GB SSD, FHD 15.6"/39.62cm, 144Hz, Windows 11 Home, Steel Gray, 1.99kg, AL15G-53 Gaming Laptop</span></h2></a>, <a class="a-link-normal s-line-clamp-2 s-line-clamp-3-for-col-12 s-link-style a-text-normal" href="/HP-Laptop-15-6-inch-Graphics-fc0154AU/dp/B0D3HG5CMG/ref=sr\_1\_6?dib=eyJ2IjoiMSJ9.Idvdwt\_UXS2\_-E6IS9-yikQp2MGPp5Zy\_jVH9i3ztLfPoqmYNjW6qzc70A3dgNRB\_v0Ly\_WMPLGFOIoRg725b5IkMxoj1P\_yB59IpK-CnpsQMY5UhPQhwSg99vh4bV60J7ROSUJnLe0qS81aaXgu2EVDngcZi-MmxeA1ymqjqET4md8V25usp1TphTFZcKjq-Te612d8\_G2f7DksT9JXsL\_B71xIm62j8h6b1XUyDA.NFFF9wjpHVybsCyUSj7hd\_S2K4TBULiqUVi\_NAUhdIk&tag=se&keywords=laptop&6" target="\_blank"><h2 aria-label='HP 15, AMD Ryzen 3 7320U (8GB LPDDR5, 512GB SSD) FHD, Anti-Glare, Micro-Edge, 15.6"/39.6cm, Win 11, Office 21, Silver, 1.59kg, fc0154AU, AMD Radeon Graphics, 1080p FHD Camera Laptop' class="a-size-medium a-spacing-none a-color-base a-text-normal"><span>HP 15, AMD Ryzen 3 7320U (8GB LPDDR5, 512GB SSD) FHD, Anti-Glare, Micro-Edge 15 6"/39 6cm Win 11 Office 21 Silver 1 59kg fc0154AU AMD Radeon Graphics 1080p FHD Camera Laptop</span>

`len(links)`

16

links[0]

[<h2 aria-label='HP Pavilion x360, 13th Gen Intel Core i5-1335U \(16GB DDR4, 512GB SSD\) Touchscreen, 14"/35.6cm, FHD, Win 11, Office 21, Silver, 1.51kg, ek1074tu/1148tu, Iris Xe, FPR, 5MP Camera, Backlit KB Laptop' class="a-size-medium a-spacing-none a-color-base a-text-normal"><span>HP Pavilion x360, 13th Gen Intel Core i5-1335U \(16GB DDR4, 512GB SSD\) Touchscreen, 14"/35.6cm, FHD, Win 11, Office 21, Silver, 1.51kg, ek1074tu/1148tu, Iris Xe, FPR, 5MP Camera, Backlit KB Laptop</span></h2></a>](/HP-Pavilion-i5-1335U-Laptop-Touchscreen/dp/B0CJ2KWNQ/ref=sr_1_1?dib=eyJ2Ijo1MSJ9.Idvwdt_UKS2_E61S9-yikQp2MGPp5Zy_jVH9i3ztLfpOqmYNjWqzcT0A3dgNRB_v0ly_WMPPLGFOIoTg725b5IKmXoj1P_yB591IpK-CnpsMY5UhpQhsWg99vh4bW6070RSUJNLe0qS8laAxgu2EVDngcZi-MnxeA1ymqjqET4md8V25usplTphTFZckjq-Te612d8_G2F7DksT9JXsl_B71xIm62j8h6b1XUyDA.NFF9wjpHvYbsCyUSj7hd_S2K4TBUL1iqUVi_NAUhdIk&dib_tag=se&keywords=laptop&q 1)

```
links[0].get('href')
```

'/HP-Pavilion-i5-1335U-Laptop-Touchscreen/dp/B0CJ2KWNNQ/ref=sr\_1\_1?dib=eyJ2IjoiMSJ9.Idvdt\_UXS2\_-E6IS9-yikQp2MGPp5Zy\_jVH9i3ztl\_fPoqmYNjw6qzc70A3dgNRB\_v0Ly\_WMPLGOfIOrg725b5IkMxoj1P\_yB59IpK-CnpnsQMY5UhPQhwSg99vh4bV60J7ROSUjnLe0qS81aaXgu2EVDngcZi-MmxmA1ymq\_jqET4md8V25usplTphTFZcKjq-Te6l2d8\_G2f7DksT9JXsl\_B71xIm6j2j8h6b1XUyDA.NFFF9wjPhVybcsCyUsj7hd\_S2K4TBULiqUVi\_NAUhdIk&dib\_tag=se&key

```
link = links[0].get('href')
```

```
product_list = "https://amazon.in" + link
```

```
url = product_list  
clean_url = url.strip("')  
print(clean_url)
```

[https://amazon.in/HP-Pavilion-i5-1335U-Laptop-Touchscreen/dp/B0CJ2KWNNO/ref=sr\\_1\\_1?dib=eyJ2IjoiMSJ9.Idvdwt\\_UXS2\\_E61S9-yikQp2M](https://amazon.in/HP-Pavilion-i5-1335U-Laptop-Touchscreen/dp/B0CJ2KWNNO/ref=sr_1_1?dib=eyJ2IjoiMSJ9.Idvdwt_UXS2_E61S9-yikQp2M)