XYZ website for Web Scraping

In [1]:
```
from bs4 import BeautifulSoup    # Importing Beautiful Soup
import csv                       # Importing CSV
import requests                  # Importing requests  --
# The import requests statement in Python allows you to make HTTP requests and interact with web servers, making it easier t
import pandas as pd              # Pandas for data manipulation
```

In [8]:
```
def get_url(search_term):
    '''Generate a url from search term'''
    template='https://www.xxyz.com/s?k={}'
    search_term=search_term.replace(' ','+')

    # add term query to url
    url = template.format(search_term)

    # add page query placeholder
    url+='&page{}'
    return template.format(search_term)

def extract_record(item):
    '''Extract and return data from a single record'''

    # description and url
    atag=item.h2.a
    description=atag.text.strip()
    url='http://www.xyz.com' + atag.get('href')

    try:
        # price
        price_parent=item.find('span','a-price a-text-price a-size-base')
        price=price_parent.find('span','a-offscreen').text
    except AttributeError:
        return

    try:
        # rating
        rating=item.i.text
    except AttributeError:
        rating = ""

    result =(description,price,rating)

    return result

def main(search_term):
    '''run main program routine'''
    # startup the webdriver

    driver = webdriver.Chrome()

    records=[]
    url=get_url(search_term)

    for page in range(1,5):
        driver.get(url.format(page))
        soup=BeautifulSoup(driver.page_source,'html.parser')
        results=soup.find_all('div',{'data-component-type':'s-search-result'})

        for item in results:
            record=extract_record(item)
            if record:
                records.append(record)

    driver.close()

    # save data to csv file
    with open('results.csv','w',newline='',encoding='utf-8') as f:
        writer =csv.writer(f)
        writer.writerow(['Description','price','rating','url'])
        writer.writerows(records)
```

In [9]:
```
main('Product Name')
```

Sample OUTPUT:

```
('LG 34WN650-W UltraWide Monitor 34" 21:9 FHD (2560 x 1080) IPS Display, VESA DisplayHDR 400, AMD FreeSync, 3-Side
Virtually Borderless Design - Silver',
```

'$349.99', '4.6 out of 5 stars', 'http://www.xyz.com/sspa/click?ie=UTF8&spc=MTo4NjQzNjgwMTg1MDU5Nzk4OjE2OTEwNjIyMDE6c3BfYXRmOjIwMDA2NTI3NTYwNTQ5ODo6MDo6&url=%2FLG-34WN650-W-34-Inch-UltraWide-DisplayHDR%2Fdp%2FB087JB656Q%2Fref%3Dsr_1_1_sspa%3Fkeywords%3Dultrawide%2Bmonitor%26qid%3D1691062201%26sr%3D8-1-spons%26sp_csd%3Dd2lkZ2V0TmFtZT1zcF9hdGY%26psc%3D1' (http://www.xyz.com/sspa/click?ie=UTF8&spc=MTo4NjQzNjgwMTg1MDU5Nzk4OjE2OTEwNjIyMDE6c3BfYXRmOjIwMDA2NTI3NTYwNTQ5ODo6MDo6&url=%2FLG-34WN650-W-34-Inch-UltraWide