In [1]:

```
!pip install requests bs4 html5lib
```

```
Requirement already satisfied: requests in /srv/conda/envs/notebook/lib/pyth
on3.6/site-packages (2.25.1)
Requirement already satisfied: bs4 in /srv/conda/envs/notebook/lib/python3.
6/site-packages (0.0.1)
Requirement already satisfied: html5lib in /srv/conda/envs/notebook/lib/pyth
on3.6/site-packages (1.1)
Requirement already satisfied: beautifulsoup4 in /srv/conda/envs/notebook/li
b/python3.6/site-packages (from bs4) (4.9.3)
Requirement already satisfied: webencodings in /srv/conda/envs/notebook/lib/
python3.6/site-packages (from html5lib) (0.5.1)
Requirement already satisfied: six>=1.9 in /srv/conda/envs/notebook/lib/pyth
on3.6/site-packages (from html5lib) (1.15.0)
Requirement already satisfied: idna<3,>=2.5 in /srv/conda/envs/notebook/lib/
python3.6/site-packages (from requests) (2.10)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /srv/conda/envs/note
book/lib/python3.6/site-packages (from requests) (1.26.3)
Requirement already satisfied: certifi>=2017.4.17 in /srv/conda/envs/noteboo
k/lib/python3.6/site-packages (from requests) (2020.12.5)
Requirement already satisfied: chardet<5,>=3.0.2 in /srv/conda/envs/noteboo
k/lib/python3.6/site-packages (from requests) (4.0.0)
Requirement already satisfied: soupsieve>1.2 in /srv/conda/envs/notebook/li
b/python3.6/site-packages (from beautifulsoup4->bs4) (2.2.1)
```

In [2]:

```
import requests
URL = "https://www.thestar.com.my/search/?q=HIV&qsort=oldest&qrec=10&qstockcode=&pgno=1"
r = requests.get(URL)
print(r.content)
```

ss","ontimeout"],y=[];e.exports=u;var x=window.XMLHttpRequest=function(t) {var e=new p(t);try{u.emit("new-xhr",[e],e),e.addEventListener(v,i,!1)}cat ch(n){try{u.emit("internal-error",[n])}catch(r){}}return e};if(c(p,x),x.pr ototype=p.prototype,d.inPlace(x.prototype,["open","send"],"-xhr-",s),u.on ("send-xhr-start",function(t,e){r(t,e),o(e)}),u.on("open-xhr-start",r),h) {var b=m&&m.resolve();if(!w&&!m){var E=1,R=document.createTextNode(E);new h(a).observe(R,{characterData:!0})}}else f.on("fn-end",function(t){t[0]&&t $[0].type===v|a()),{}],11:[function(t,e,n){function r(t){if(!s(t))return}}$ null;var e=window.NREUM;if(!e.loader config)return null;var n=(e.loader co nfig.accountID||"").toString()||null,r=(e.loader config.agentID||"").toStr ing()||null,f=(e.loader_config.trustKey||"").toString()||null;if(!n||!r)re turn null;var h=p.generateSpanId(),m=p.generateTraceId(),w=Date.now(),v={s panId:h,traceId:m,timestamp:w};return(t.sameOrigin||c(t)&&l())&&(v.traceCo ntextParentHeader=i(h,m),v.traceContextStateHeader=o(h,w,n,r,f)),(t.sameOr igin&&!u()||!t.sameOrigin&&c(t)&&d())&&(v.newrelicHeader=a(h,m,w,n,r,f)),v}function i(t,e){return"00-"+e+"-"+t+"-01"}function o(t,e,n,r,i){var o=0, a="",s=1,c="",f="";return i+"@nr="+o+"-"+s+"-"+n+"-"+r+"-"+t+"-"+a+"-"+c +"-"+f+"-"+e}function a(t,e,n,r,i,o){var a="btoa"in window&&"function"==ty peof window.btoa;if(!a)return null;var s={v:[0,1],d:{ty:"Browser",ac:r,ap: i,id:t,tr:e,ti:n}};return o&&r!==o&&(s.d.tk=o),btoa(JSON.stringify(s))}fun

In [4]:

```
from bs4 import BeautifulSoup
```

In [5]:

```
soup = BeautifulSoup(r.content, 'html51ib') # If this line causes an error, run 'pip instal
print(soup.prettify())
o.replaceState&&(a=o),i.inPlace(a,["pushState","replaceState"],"-")},{}],
8:[function(t,e,n){var r=t("ee").get("raf"),i=t("wrap-function")(r),o="equ
estAnimationFrame";e.exports=r,i.inPlace(window,["r"+o,"mozR"+o,"webkitR"+
o, "msR"+o], "raf-"), r.on("raf-start", function(t){t[0]=i(t[0], "fn-")})},{}],
9:[function(t,e,n){function r(t,e,n){t[0]=a(t[0],"fn-",null,n)}function i
(t,e,n){this.method=n,this.timerDuration=isNaN(t[1])?0:+t[1],t[0]=a(t
[0], "fn-", this, n)}var o=t("ee").get("timer"), a=t("wrap-function")(o), s="se
tTimeout",c="setInterval",f="clearTimeout",u="-start",d="-";e.exports=o,a.
inPlace(window,[s,"setImmediate"],s+d),a.inPlace(window,[c],c+d),a.inPlace
(window,[f,"clearImmediate"],f+d),o.on(c+u,r),o.on(s+u,i)},{}],10:[functio
n(t,e,n){function r(t,e){d.inPlace(e,["onreadystatechange"],"fn-",s)}funct
ion i(){var t=this,e=u.context(t);t.readyState>3&&!e.resolved&&(e.resolved
=!0,u.emit("xhr-resolved",[],t)),d.inPlace(t,g,"fn-",s)}function o(t){y.pu
sh(t),h\&(b?b.then(a):w?w(a):(E=-E,R.data=E)) function a()\{for(var\ t=0;t<
y.length;t++)r([],y[t]);y.length&&(y=[])function s(t,e){return e}function
c(t,e){for(var n in t)e[n]=t[n];return e}t(5);var f=t("ee"),u=f.get("xh
r"),d=t("wrap-function")(u),l=NREUM.o,p=1.XHR,h=1.MO,m=1.PR,w=1.SI,v="read
ystatechange",g=["onload","onerror","onabort","onloadstart","onloadend","o
```

In [7]:

```
time = []
name = []
for item in soup.find_all('div' , attrs={'class' : 'row list-listing'}):
    time.append(item.find('span' , attrs={'class' : 'timestamp'}).text)
    name.append(item.find('h2' , attrs={'class' : 'f18'}).text)
```

nprogress","ontimeout"],y=[];e.exports=u;var x=window.XMLHttpRequest=funct
ion(t){var e=new n(t):trv{u.emit("new-xhr".[e].e).e.addEventListener(v.i.l)

In [9]:

```
import pandas as pd
```

In [10]:

```
data = pd.DataFrame({'name':name , 'timestamp':time})
```

```
In [11]:
```

```
time, name
Out[11]:
(['08 Jan 2003 | 12:00 AM',
  '08 Jan 2003 | 12:00 AM',
  '17 Jan 2003 | 12:00 AM',
  '21 Jan 2003 | 12:00 AM',
  '26 Jan 2003 | 12:00 AM'
  '26 Jan 2003 | 12:00 AM'
  '12 Feb 2003 | 12:00 AM',
  '12 Feb 2003 | 12:00 AM',
  '13 Feb 2003 | 12:00 AM'
  '14 Feb 2003 | 12:00 AM'],
 ['\n
                              Heart in the right place\n
\n
  'Africa can learn from the experience of East Asian nations',
  'Suhakam: Prison overcrowded',
  "'Hubby injected 'me with HIV'",
  'Furore over teen sex education',
  'Some diseases manifest even after screening',
  'Supermax confident of doubling net profit by year-end',
  'Matching horoscopes with condoms',
  'Medical tests highly inaccurate',
  '\n
                              Enter the Devil\n
                                                                        \n
'])
```

In [12]:

```
data.head()
```

Out[12]:

	name	timestamp
0	\n Heart in the right p	08 Jan 2003 12:00 AM
1	Africa can learn from the experience of East A	08 Jan 2003 12:00 AM
2	Suhakam: Prison overcrowded	17 Jan 2003 12:00 AM
3	'Hubby injected 'me with HIV'	21 Jan 2003 12:00 AM
4	Furore over teen sex education	26 Jan 2003 12:00 AM

In [15]:

```
data.to_csv("data.csv")
```

In [16]:

```
d=pd.read_csv("data.csv")
```

In [19]:

d.head()

Out[19]:

	Unnamed: 0	name	timestamp
0	0	\n Heart in the right p	08 Jan 2003 12:00 AM
1	1	Africa can learn from the experience of East A	08 Jan 2003 12:00 AM
2	2	Suhakam: Prison overcrowded	17 Jan 2003 12:00 AM
3	3	'Hubby injected 'me with HIV'	21 Jan 2003 12:00 AM
4	4	Furore over teen sex education	26 Jan 2003 12:00 AM