CNN All Links Retriever

Author : Nathan Ostrowski, @nostrowski

Description :

The following notebook, when evaluated, will find the link of each and every CNN Article matching a given search parameter. The notebook then will export those links to "articleLinks.mx".

To do so, the notebook will open a web session in your default browser, open a search page, and add up to 100 links at a time before closing that browser window. The process will repeat for as many hundreds of links you are attempting to find.

To run this notebook, simply change the search parameter, specify the number of articles you wish to find, then go to "Evaluation > Evaluate Notebook"

Executable Code

```
searchParameter = "YOUR_SEARCH"
totalArticles = NaN

baseUrl = StringReplace[
   "https://www.cnn.com/search?q=SEARCHTERM&size=100&sort=newest&type=article&from
     =0&page=1", "SEARCHTERM" → searchParameter];

searchLinks = {}

For[i = 1, i < (totalArticles / 100), i++,
 AppendTo[searchLinks, StringReplace[
   baseUrl, "from=3500" → ToString[("from=" <> ToString[(100 + (i * 100))])]
  ]
 ]
]

articleLinks = {}
```

```
For[i = 1, i < Length[searchLinks] + 1, i++,
  session = StartWebSession[] ;
  WebExecute["OpenPage" → searchLinks[[i]]] ;
  Pause[2] ;
  rawLinks = WebExecute["PageHyperlinks"] ;
  cleanLinks = DeleteDuplicates[Drop[Drop[rawLinks, 122], -120]] ;
  AppendTo[articleLinks, cleanLinks] ;
  DeleteObject[session]
 ];

articleLinks = Flatten[articleLinks];

Export[articleLinks, "articleLinks.mx"]
```