

# CNN Article Retriever

Author : Nathan Ostrowski, @nostrowski

---

## Description :

The following notebook, when evaluated, will read the List object "articleLinks" from the provided file "articleLinks.mx" and subsequently loop through each provided URL, importing the raw CNN Article text and adding it to an "articleTexts" object. That object will be exported into "articleTexts.mx", which can then be combined with other similarly exported files of the same name.

To run this notebook, simply go to "Evaluation > Evaluate Notebook"

\*\*\*Before running, you must define first article to download and the end article to download. Those variables are startArticle and endArticle respectively. Try running as many as you can! I've included a print statement for each successful import so that you can be aware of your progress at any given time.

---

## Executable Code

```
articleLinks = Import["articleLinksV2.mx"];  
  
articleTexts = {}  
  
startArticle = NaN  
endArticle = NaN  
  
For[i = startArticle, i < endArticle+1, i++,  
  AppendTo[articleTexts, Import[articleLinks[[i]]]];  
  Print[i]  
]  
  
Export[articleTexts, "articleTexts.mx"]
```