

CNN Article Cleaner

Author : Nathan Ostrowski, @nostrowski

Description :

The following notebook, when evaluated, will read through the raw CNN articles imported in an "articleTexts.mx" file and clean each one before exporting a list of cleaned CNN articles in a "cleanedArticleTexts.mx" file.

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

Executable Code

```
In[263]:= articleTexts = Import["articleTexts.mx"]

CleanArticleText[article] :=
(
  If[StringContainsQ[article, "(CNN)"],
    cleanedArticle = StringSplit[StringSplit[article, "(CNN)"], "Search
US"];
  ];
  If[StringContainsQ[article, "(CNN "],
    cleanedArticle = StringSplit[StringSplit[article, "(CNN ", "
Content by"];
  ];
  Return [cleanedArticle[[2]][[1]]]
)

In[272]:= cleanedArticleTexts = {};

In[273]:= For[i = 0, i < Length[articleTexts], i++,
  AppendTo[cleanedArticleTexts, CleanArticleText[articleTexts[[i]]]]
]

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