**Scrapping the articles from the link** [**https://www.thebetterindia.com/archives/**](https://www.thebetterindia.com/archives/)

This repository is devoted to scrapping the contents of all the articles available in the archives of news website thebetterindia.com . The work is a subpart of a larger project based on Natural Language Processing. The project aims to build an engine which picks the articles which contain a certain kind of news from a pool of thousands of articles with almost cent percent accuracy.

**Approach**

I have divided the job in three subparts -1. Crawling the links of all the articles present in archives of one particular year. 2. Applying the same code for archives of each year with the help of a ‘for’ loop. 3.Getting the required content i.e. text and headlines from the link of a given page.

I have attached two jupyter notebooks. The first one contains steps 1 and 2 and the second one contains step 3.