

## Automatic News Scraping with Python, Newspaper and Feed parser

The problem we are trying to solve here is to extract relevant information from news articles, such as the title, author, publish date, and the main content of the article. This information can then be used for various purposes such as creating a personal news feed, analyzing trends in the news, or even creating a dataset for natural language processing tasks. In the news, or even creating a dataset for natural language processing tasks. In this article, we will look at how we can use the Python programming language, along with the Newspaper and Feedparser modules, to scrape and parse news articles from various sources.

### Automatic news scraping with Python

To solve this problem, we can use the Python programming language, along with the Newspaper and Feedparser modules. The Newspaper module is a powerful tool for extracting and parsing news articles from various sources, while the Feedparser module is useful for parsing RSS feeds. RSS (Really Simple Syndication) is a web feed that allows users and applications to access updates to websites in a standardized, computer-readable format. These updates can include blog entries, news articles, audio, video, and any other content that can be provided in a feed.

### Required Module

```
!pip install newspaper3k
```

```
!pip install feedparser
```

Some of the Important Methods are:

The Newspaper and Feedparser modules have several useful methods for extracting and parsing news articles:

- `newspaper.build()`: This method is used to build a newspaper object from a given URL.
- `newspaper.download()`: This method is used to download the HTML of a given URL.
- `newspaper.parse()`: This method is used to parse the HTML of a given URL and extract relevant information such as the title, author, publish date, and main content of the article.
- `feedparser.parse()`: This method is used to parse an RSS feed and extract relevant information such as the title, author, publish date, and link of the article.

Now that we have an understanding of the modules and methods we will be using, let's look at how we can use them to scrape and parse news articles from various sources.

### Code Implementation

First, we import the required modules `newspaper`, and `feedparser`. Next, we define a function called `scrape_news_from_feed()` which takes a feed URL as input. Inside the function, we first parse the RSS feed using the `feedparser.parse()` method. This returns a dictionary containing various information about the feed and its entries.

Create a newspaper article object using the `newspaper.Article()` constructor and passing it the link of the article. Then download and parse the article using the `article.download()` and `article.parse()` methods. Extract relevant information such as the title, author, publish date, and main content of the article. Append this information to a list of articles. Finally, the function returns the list of articles.