

Reference Links:

https://www.diffbot.com/dev/docs/

https://www.import.io/post/how-to-crawl-a-website-the-right-way/

http://blog.diffbot.com/crawling-news-sites-for-new-articles-and-extracting-clean-text/