**Week 1:**

31/05/2021

* **Tasks worked on and completed this week:**
* Using BeautifulSoup worked on parsing of data of Rss feed
* Using NLTK extracted keywords from title of the rss feeds
* **New things learned:**
* LEARNING BEAUTIFULSOUP LIBRARY AND LEARNING TO

EXTRACT LINES FROM XML FILE

<https://linuxhint.com/parse_xml_python_beautifulsoup/>

We can install both libraries using the ****pip**** installation tool,through the command : pip install bs4 lxml

Find\_all function is used to find the particular tags content

* Extracting Keyphrases from Text: RAKE in Python is a [modified](http://sujitpal.blogspot.co.nz/2013/03/implementing-rake-algorithm-with-nltk.html)

[version](http://sujitpal.blogspot.co.nz/2013/03/implementing-rake-algorithm-with-nltk.html) that uses the natural language processing toolkit NLTK for some of the calculations.

<https://pypi.org/project/rake-nltk/>

* **plan for the next week**

Planning on learning how to store this data into csv file and learning concepts how to analyze date published and update