WEEK-2

Date:5-06-2021

OBJECTIVE

Identify the Data Source (URLs) and Building a program to extract Title,Published Date etc. from the RSS feed

ACTIVITY

- *Task worked on and completed :*

  - Identified Data Sources i.e website and blogs which often update on the Project requirements with the help of KDNuggets.

  - Built a program to surf for feed URLs for the websites or blogs.

  - Built a program to extract the necessary from the features from Feeds.

- *New Challenges:*

  - Understanding RSS feed , this where Youtube Tutorials helped a lot.

  - Then find different kind of Feed (RSS,Atom) which we can use the same way.

  - Got stuck in getting Feed URLs , then had to use Source Code Inspection from browser and Using Open Source library .

  - Attribute extraction , using library and self built code. Find the difference in them.

  - Finding the time complexity for each run and comparing & analysing the extracted value from both the approach.

- *Planning for next week:*

  - Using the best approach for Feed extraction and Building Different Model for Key-Phrase Extraction.

  - Evaluating these different model using suitable metrics.

  - Planning the use DataBases for backend analysis (Mostly PostGreSQL, but MYSQL and SQLite3 are still in consideration.

Please go through the code (ipynb) file upload and provide FeedBack.