

1. Feed Parser and Data Extraction[Done]:

- Utilized feedparser and beautiful soup for extraction of title, content, publication date, and source URL for each RSS feed

2. Database Storage[Done]:

- Utilized sqlite3 for data storage and schema design

3. Task Queue and News Processing[Done]:

- Utilized dramatiq for building queue which asynchronously process all articles and assigns categories through it.
- Utilized spaCy to classify each article into the provided categories.
- Classification is performed based on keyword matching.

4. Logging and Error Handling:

- Error handling has been implemented at various parts of the code using try except blocks.

Futher Improvements:

- When performing news processing, it faces a database locked error leading to some articles not being assigned a category.
- Code formatting needs to be worked on for scalability and easier understanding.
- Utilized a keyword matching approach instead could utilize a ml based approach as well for news classification.

Link to deliverables containing code, document and extracted_Data:

https://github.com/DarrenDsouza47/News_classifier