

In crafting the Python script, RSS FEED.py, to address the News Article Classification Assignment, I embraced a methodical approach to ensure effectiveness and simplicity. The script begins by systematically extracting pertinent details from diverse RSS feeds while intelligently managing duplicate articles. The organized storage of this information within a well-structured relational database, facilitated by SQLAlchemy, ensures efficiency and systematic data handling. For seamless article processing, I incorporated Celery, allowing for the asynchronous management of news articles and ensuring scalability. The Celery worker plays a pivotal role by executing essential tasks, such as category classification using spaCy. I opted for a clear-cut keyword-based categorization approach to enhance understanding. The script also features robust logging mechanisms, fostering transparency, and graceful error handling, thereby elevating overall reliability. My library choices, including Feedparser, SQLAlchemy, Celery, and spaCy, were made due to their widespread acceptance and appropriateness for the assignment's demands. The comprehensive documentation accompanying the script, coupled with the resulting CSV file named classified_articles.csv, aims to provide a lucid understanding of the script's logic and design choices.