

DYNAMIC NEWS PORTAL - A ONE-STOP DESTINATION FOR INFORMATION UPDATES ON CURRENT EVENTS

FAIROOSA K.K
MES21MCA-2010

ABSTRACT

The online news portal aggregates and displays news articles from various sources. It provides a platform for users to access the latest news and updates on different topics, including politics, sports, entertainment, and technology. This provides a one-stop destination for users to stay informed and updated on current events and happenings. It is typically updated on a real-time basis, so it plays an essential role in keeping the public informed and up-to-date on the events that shape our world.

This application only collects trending and relevant data from different sources and constantly streams. So, users get real-time news very quickly, unlike other applications. In this proposed system, the reader can directly log into the system and users can choose the news domain they are looking for and read it. However, buying the newspaper or searching for other sources is optional. They can read it online. Also, every update/news will come to them directly. The updates will appear in their feeds automatically. This has many advantages over traditional news sources, such as the ability to access news and information from anywhere in the world, 24/7 availability, and a more range of perspectives and viewpoints

This application typically operates by gathering news from various sources using data scraping. Data scraping is the process of importing information from a website, this data is dumped into a centralized repository that allows you to store all your structured and unstructured data. This data is then processed using apache spark an open-source, distributed data processing framework computing system that provides a fast and easy-to-use platform for large-scale data processing and analysis, for ranking and modifying the content and title, then this data is loaded into the elastic search database fetches data from this database(Elasticsearch is a distributed, open-source search and analytics engine designed for handling large amounts of data and give back the answer in milliseconds).

This proposed system is based on pipeline technology. A pipeline consists of a chain of processing elements arranged so that the output of one component is the input for the next.

Python is used for data scrapping its importing information from websites. This data is dumped into a data lake in its local machine storage. PySpark is a distributed processing and interface for Apache Spark Python is used for processing the data, and the workflow orchestration is managed by using Airflow in python. The proposed web application is developed by using the Django framework and javascript.

This proposed system is an attempt to develop a News or information-based website. Online news portals have become increasingly popular due to their convenience and accessibility. They offer a wide range of news and information, including sports, entertainment, finance, etc. The implementation of this project helps the people and makes them aware so they can know any news or update their knowledge.