ABSTRACT
With so many events occurring on daily basis, it is important to stay up to date with current news events. In the age of Social media where Millennials have a very short attention span and want to get things done within a click or less, expecting them to read entire news articles is out of the question. Our project aims to exploit the social media trend by synthesizing videos of news articles which will provide relevant and concise information thus giving a quick overview of the news events

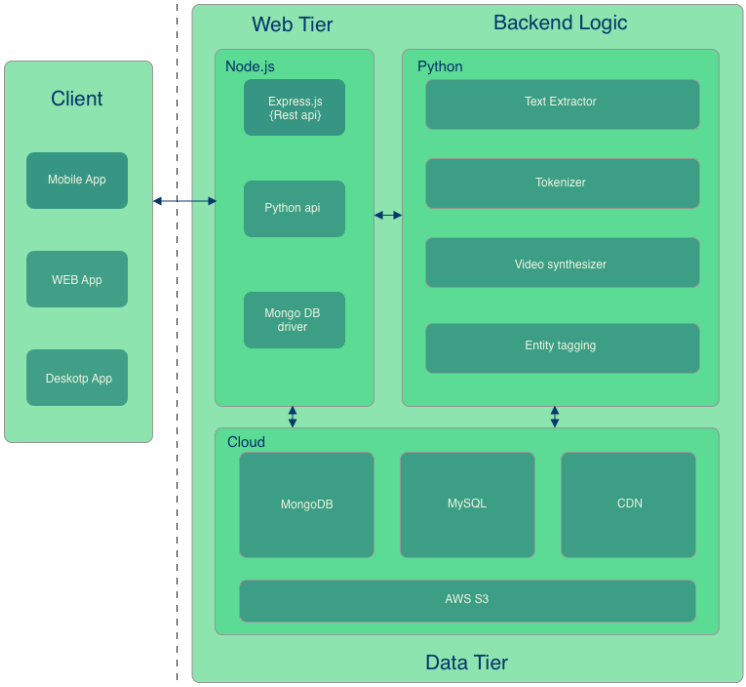
KEYWORDS

INTRODUCTION
This work is concerned with the issues of generating effective news videos for social media channels. The project is aimed to achieve the following:
•To implement an API that receives an article in request and returns keywords.
•To implement a Rule Engine that converts unstructured text from news articles to structured text .
•To generate a video stream of the input text using the Entity tagged Data.

EXISTING WORK

NECESSITY
The motivation is derived from the need of structured news data for the analysis and research and lack of videos that provide relevant information about news event. Online news videos are generated according to technological convenience, like the technology and platforms available rather than by consumer demand. Website users spend very little time watching online video news, only around 2.5% of average visit time. 97.5% of the average visit time is spent reading news articles. The main reason behind this is lack of relevant news video available due to the manual effort required to do so. This project is aimed at reducing the manual effort required to generate news videos. This will help provide creditable coverage of news through videos. It will certainly revolutionize the way news is delivered to the end user.Our project also helps researchers and analysts to get the news articles data in a structured format which in turn reduces manual effort and errors in converting it to a database

METHODOLOGY



News Scraping

Videos will be generated for every news article published online. News URLs will be constantly scraped real time. To reduce complexity, we will consider news articles from particular news sites at the beginning.

For example:

http://www.sify.com/movies/boney-kapoor-says-ajith-s-60th-film-is-an-original-script-and-not-remake-news-tamil-tb5jOgiihdcci.html

https://www.cricbuzz.com/cricket-news/106438/india-cricket-new-zealand-bangladesh-2019-world-cup-warm-up

https://timesofindia.indiatimes.com/sports/cricket/news/india-well-behaved-team-icc-ceos-response-to-queries-on-hardik-pandya-sexist-row/articleshow/67775753.cms

Text Extraction

The term text analytics describes a set of linguistic, statistical, and machine learning techniques that model and structure the information content of textual sources for business intelligence, exploratory data analysis, research, or investigation[5]. Text mining, also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interest. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling (i.e., learning relations between named entities) [6].

json output format:


```
{
  "id": 1,
  "title": "Business Standard",
  "date": "2031-01-19",
  "authors": [
   "Business Standard"
  ],
  "keywords": [
   "disabled",
   "unless",
   "business",
   "standard",
   "parts",
   "device",
   "site",
   "described",
   "settings",
   "policy",
   "function",
   "cookie"
  ],
  "summary": "By continuing to use this site you consent to the use of cookies on your device as described in our Cookie Policy unless you have disabled them.",
  "image_url": "https://bsmedia.business-standard.com/_media/bs/img/common/no_preview.jpg",
  "video_url": [],
  "time": "00:00:00"
 }
```

Key Words

Image Search

From the information and keywords extracted from the text, relevant images will be found from the search engine. Top 5 images will be used to in the video for a duration of 45 seconds overall. Images will be selected based on the scores given by the Convolutional Neural Network. Algorithms like Linear Regression and Logistic Regression can also be used for this purpose.

Video Generation

Information extracted from the news articles will be displayed against a backdrop of the images selected. Text summarization modules and NLP will be used to gain a grammatically accurate summary of the news article. The news title date, location will be mentioned for every news along with other details

Dataset

Videos generated along with the images and extracted information will be stored in an NoSQL database. The dataset will thus be continually growing. Entity tagging will be performed on the data stored. These data points (news articles) will be connected via knowledge graphs.This will help establish relation between related data points.

CHALLENGES

1. Making sure no duplicate news video is generated.
2. Ensuring accuracy and grammatical precision of summary being displayed.
3. Finding relevant images incase of a rare news issue.
4. Connecting the correct data points to make a knowledge graph
5. Increase the speed of the entire system to generate videos and keep up with the news influx.

CONCLUSION AND FUTURE WORK

1] Johanna Fulda, Matthew Brehmel, Tamara Munzner, 'TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text' IEEE Transactions on Visualization and Computer Graphics, October 20, 2004, NBC Nieuwegein, NL pp 206-213.
[2] Jinjun Wang, Engsiong Chng, Hanqinq Lu, Changsheng Xu and Qi Tian: 'Generation ofPersonalized Music Sports Video Using Multimodal Cues,'IEEE TRANSACTIONS ON MULTIMEDIA,VOL. 9, NO. 3, APRIL 2007.

[3] Li Xuemei, Li Yan and Li Jincheng,'Application of AI Algorithm in Video Indexing and Retrieval,'Third International Symposium on Intelligent Information Technology Applicationvol. 6, no. 2 pp. 170-179, April 2009.[4] Marco Rospocher, Marieke van Erpb, Piek Vossenb, Antske Fokkens, Itziar Aldabe,German Rigauc, Aitor Soroa, Thomas Ploeger, Tessel Bogaardd, Teh Ying Wah and David Ngo 'Building event-centric knowledge graphs from news,'Web Semantics: Science,Services and Agents on the World Wide Web21 October 2015

[4] https://pdfs.semanticscholar.org/28cd/34876dea8fcfac7fc6293090ba758aa62cd7.pdf
A Study of Information Extraction Tools for Online English Newspapers (PDF): Comparative Analysis
[5] Archived November 29, 2009, at the Wayback Machine

[6] https://en.wikipedia.org/wiki/Text_mining