**Read, Watch, Listen and Summarize Multi-modal Summarization for Asynchronous**

**ABSTRACT:**

Automatic text summarization is a fundamental natural language processing (NLP) application that aims to condense a source text into a shorter version. The rapid increase in multimedia data transmission over the Internet necessitates multi-modal summarization (MMS) from asynchronous collections of text, image, audio and video. In this work, we propose an extractive MMS method that unites the techniques of NLP, speech processing and computer vision to explore the rich information contained in multi-modal data and to improve the quality of multimedia news summarization. The key idea is to bridge the semantic gaps between multi-modal content. Audio and visual are main modalities in the video. For audio information, we design an approach to selectively use its transcription and to infer the salience of the transcription with audio signals. For visual information, we learn the joint representations of text and images using a neural network. Then, we capture the coverage of the generated summary for important visual information through text-image matching or multi-modal topic modeling. Finally, all the multi-modal aspects are considered to generate a textual summary by maximizing the salience, non-redundancy, readability and coverage through the budgeted optimization of submodular functions. We further introduce a publicly available MMS corpus in English and Chinese. The experimental results obtained on our dataset demonstrate that our methods based on image matching and image topic framework outperform other competitive baseline methods.

**MODULES:**

There are three modules can be divided here for this project they are listed as below

* Image to Text Classification
* Video to Text classification
* Text Categories
* Graphical analysis

From the above four modules, project is implemented. Bag of discriminative words are achieved

1. **Image to Text Classification:**

The data can be uploaded by admin without any particular scenario but with the details of Image data. The most importantly large amount of can be handled in order to do practically. The data that are handling throughout the project can be done in this module. Users have permission to view data but not edit the data in online they can request the user to get the data. Check the image in the OCR conversation used in image to text conversion

1. **Video to Text Classification:**

The data can be uploaded by admin without any particular scenario but with the details of Audio data. The most importantly large amount of can be handled in order to do practically. The data that are handling throughout the project can be done in this module. Users have permission to view data but not edit the data in online they can request the user to get the data. Check the Audio to text conversation after that classify the category.

3.**Text Categories:**

The centroids are fixed set of words that are actually makes the context of the content to be classified and clustered into folders. The general sets of documents are into their respective clusters based on the separation of centroids.

**4. Graphical analysis:**

The graph analysis is done by the values taken from the result analysis part and it can be analyzed by the graphical representations. Such as line chart here in this project. The data can be categorized by group by data based on the some scenarios. The data can be cluster with various factors in order to get data properly.

**REQUIREMENT ANALYSIS**

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well-ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

**REQUIREMENT SPECIFICATION**

**Functional Requirements**

* Graphical User interface with the User.

**Software Requirements**

For developing the application the following are the Software Requirements:

1. Python
2. Django

**Operating Systems supported**

1. Windows 7
2. Windows XP
3. Windows 8

**Technologies and Languages used to Develop**

1. Python

**Debugger and Emulator**

* Any Browser (Particularly Chrome)

**Hardware Requirements**

For developing the application the following are the Hardware Requirements:

* Processor: Pentium IV or higher
* RAM: 256 MB
* Space on Hard Disk: minimum 512MB

**CONCLUSION:**

* This paper addresses an asynchronous MMS task, namely, how to use related text, audio and video information to generate a textual summary. We formulate the MMS task as an optimization problem with a budgeted maximization of submodular functions. We address readability by selectively using the transcription of audio through guidance strategies. More specifically, we design a novel graph-based model to effectively calculate the salience score for each text unit, leading to more readable and informative summaries. We investigate various approaches to identify the relevance between the image and text, and find that the image match model and the image topic model perform well for the MMS task. The final experimental results obtained using our MMS corpus in both English and Chinese demonstrate that our system benefits from multi-modal information. Adding audio and video does not appear to dramatically improve the performance with respect to the text only model, which indicates that better models are needed to capture the interactions between text and other modalities, especially for visual information. We also plan to expand our MMS dataset, specifically to collect more videos.